

June 2021

Towards an AI Assistant for Power Grid Operators

Antoine MAROT^a, Alexandre ROZIER^a, Matthieu DUSSARTRE^a,
Laure CROCHEPIERRE^a and Benjamin DONNOT^a

^a RTE, R&D and AI Lab, France

Abstract. Power grids are becoming more complex to operate in the digital age given the current energy transition to cope with climate change. As a result, real-time decision-making is getting more challenging as the human operator has to deal with more information, more uncertainty, more applications, and more coordination. While supervision has been primarily used to help them make decisions over the last decades, it cannot reasonably scale up anymore. There is a great need for rethinking the human-machine interface under more unified and interactive frameworks. Taking advantage of the latest developments in Human-Machine Interface and Artificial Intelligence, we expose our vision of a new assistant framework relying on an hypervision interface and greater bidirectional interaction. We review the known principles of decision-making driving our assistant design alongside with its supporting assistance functions. We finally share some guidelines to make progress towards the development of such an assistant.

Keywords. Artificial Intelligence, Assistant, Power Grid, Operation, Decision-making, Human-Machine Interface, Hypervision

1. Introduction

From the beginning, power grids have been complex artificial systems. As of today, complexity keeps rising at a time of energy transition given the advent of intermittent renewable energies on the production side and of prosumers on the demand side, coupled with the globalization of energy markets over a more and more interconnected European grid. Operators are facing aging grids, with slower grid asset developments due to decreasing social acceptability. Operators hence need to operate a system closer to its limits while dealing with greater uncertainty and increasing grid automation [1] inducing complex cyber-physical dynamics [2]. While there could eventually be a temptation to develop a fully autonomous grid to cope with that complexity, it falls short for such large critical system operations. Indeed, coordination, responsibility, accountability, and explainability are a must when operating such a system and can only be reasonably achieved by humans today: human operators remain key players [3].

Historically, control rooms have tackled grid evolution dynamics by gradually adding more applications and screens in control rooms. While incrementally growing a monitoring alarm-driven management system has been relatively effective until now, we are today struggling to maintain effective decision-making at a new scale of complexity, essentially in terms of manageable cognitive load for human operators, who should

remain at the center of decisions [1]. In that regard, Human-Machine Interfaces (HMI) have been identified as a risk factor for human error [4], and should now be considered more closely. In addition, several decades of research in psychology [5] and neuroscience [6] have shed light on the human decision-making process and its limits, which could in turn lead to improving it.

Lately, several works have proposed situation awareness frameworks [7,8,9] to help augment the operator's comprehension of safety-critical situations. In addition to better information processing, it is also urgent to rethink the operator's human-machine interface [7] and interaction to assist the operator's regular real-time decision-making. Rather than having operators adapt to the machine through a technology-centered system engineering design, machines and operators could co-adapt [10] following a more human centered-design [11,12] approach, possibly rooted in the older concept of man-computer symbiosis [13]. In terms of interfaces, we have seen tremendous innovations in other domains, especially in consumer products such as smartphones, connected homes, social networks, search engines, and recommendation systems. They have been well-adopted for ergonomically providing the most relevant information to the user, mostly on single screens, while dealing in the background with vast amounts of information. In the future, the webstrates [14] concept might enable very adaptive, evolving, personal and consistent interfaces by sharing media ubiquitously through devices and applications.

The development of today's interfaces has also been made possible thanks to the latest developments of Artificial Intelligence (AI), and particularly in the field of Machine Learning (ML). These advances enabled deeper and more practical large-scale real-time information processing, such as in computer vision [15], image understanding [16], natural language processing [17], and recommendations [18]. This shows a shift towards even more advanced HMI, through the concept of assistants. Assistants were found useful to both improve single-user performance and group collaboration on a common task [19], help chess players develop their playing skills [20], or help programmers write code [21]. In power grids, the notion of an AI assistant was used lately in [22,23] and listed as an opportunity to tackle climate change [24]. Nevertheless, eventual assistant pitfalls should be closely investigated, analogous to the very influential "ironies of automation", which notably used power grids as an introductory example [25].

In this paper, we present the future design of operator AI-infused assistant given the latest developments in HMI, AI, and decision-making science. Our contributions lie in:

- establishing links between these different domains and our application scenarios to push towards appropriate solution design while being aware of pitfalls;
- broadening the perspective to other research fields related to human behavior and cognition beyond the usual situation awareness focus in power systems [3,8];
- adapting existing AI assistant framework and guidelines to power grid specificity;
- proposing hypervision interface as an effective hybrid decision-making enabler.

We first describe in Section 2 the industrial use case we focus on. Then in Section 3, we review the effectiveness and limits of humans and AI in decision-making, highlighting their complementarity. We further define an assistant in Section 4. We then discuss in Section 5 the specificity of an assistant for grid operators. Finally, in Section 6, we devise guidelines for how researchers can develop and test such an assistant.

June 2022

2. Use case description

We are here considering the field of evolving power grid operations for the Energy Transition. Operations have to adapt as the ongoing decarbonization to cope with Climate Change modifies the generation, electrifies the demand, and digitizes system devices.

2.1. Today's operations

Today grid operators operate in real-time from control rooms to optimize the power flows on electrical lines, handle maintenance with planned outages or new equipment integration on the grid, and most importantly, avoid blackouts because of congestion. More details about their role and tasks can be found in [3]. They are highly trained engineers as their job requires studies, planning, and adaptable decision-making rather than simply reproducing pre-established event management scenarios. They operate based on simulation tools, real-time and forecasted data, but yet with little decision-making support tools such as assistants. When they feel they need to solve a problem, they mostly manually explore solutions and validate their decision in their simulation tool. They can modify the line connectivity on the grid to reroute power flows, but also modify some production, limit consumption by a few percent, or even use battery storage today to change the power flows on the grid. This is a large set of possible flexibilities among which they have to identify the effective ones in a given context. Day-ahead planning services give insights on the upcoming trends and possible actions to start considering if some problem occurs, in addition to outage planning. There then exists an intra-day established workflow with 5-minute times step forecast resolution over a few hours' horizons. But this requires a lot of supervision, a lot of manual entries and manual simulations. They operate mainly with experience to determine relevant remedial actions.

As the variability on the grid is increasing a lot with shifting dynamics and behaviors even within a year now given the energy transition, usual solutions might not work anymore in all or new contexts. Operators will have to adapt more quickly, leveraging human flexibility if given a proper work environment [26]. As it can take months today for an operator to gain the necessary knowledge on a different system through extensive training through manual studies, new assistance will be needed, as much as improved performance evaluation [27]. Also, while some flexibilities and actions can be leveraged curatively once a problem occurs, sometimes actions have to be taken preventively before it is too late to implement them given operational constraints, in particular when considering using batteries or redispatching. Several operators also act on the same interconnected grid at the same time, and this requires some coordination that is not always easily achieved despite common grid representations.

2.2. Current limitations today and assistant need

The current "supervise everything" approach on dozens of screens cannot scale anymore to take critical decisions as the complexity rises: operators become overwhelmed by information without many indications on what to do and any recommendations for it. The volume of data to consider only gets increasing by 2 to 3 orders of magnitude to keep the ability to predict and anticipate in a more stochastic and constrained system. There is hence a need to help operators identify and prioritize tasks while displaying relevant

information and recommendations only for those tasks, with ideally a single interface assistant. In particular, usual approaches from operational research [28,29] don't leverage very well available historical data, whereas current data-driven approaches could provide a welcome speed-up. They don't integrate operator experience and preferred thought process, hence limiting their acceptability. New research is however emerging to better integrate human decision-making [30]. Finally, current workflow only considers snapshot-based solutions and not sequential decisions over a time horizon, thus putting the burden of the anticipation and evaluation of the long-term consequences of the actions on the human operator alone.

As the grid gets pushed towards its limits, decisions become more numerous and a lot more interdependent. Solutions should not only be effective at one given time but over a larger time horizon. Solutions should also be implemented with the right anticipation given their effective activation time: switching line connectivity is quick but starting production can sometimes take a few hours. So decisions will need to indeed consider the full underlying planning problem of power grid operations (as illustrated in section 6.2). Better near real-time operation planning assistance could also help anticipate the workload and level it up across operators in a given control room.

An assistant could hence help augment the operator's decision-making capability to address [1]: (a) More rapidly evolving and changing system environment; (b) More numerous, complex, and coordinated decisions to make; (c) More uncertainty to consider and more anticipation needed; (d) Overcrowded and fragmented work environment with multi-screen applications and data patchwork; (e) Human operator cognitive load saturation. It would eventually help limit the grid operational cost to a few percent increase rather than a 2-fold one. The ambition is to use at least twice as much flexibility as of today, i.e. to use them more frequently and with more diversity [31]. It would also facilitate workload [32] management across teams for improved coordination [33] while avoiding cognitive load saturation, and deal with information overload. The interactions should not be only limited to isolated Operator-Assistant pairs, but involve teams of interrelated activities through a shared information and task management system. It should however not come at the cost of decreasing responsibility and accountability, making clear who is in charge and providing necessary interpretability for operators to make explanations.

3. Human-AI dynamics in decision-making

3.1. Human decision-making

Human decision-making is a matter of attention and executive control [6]. Taking proper decisions first involves paying attention to the right information in the environment and making sense of it. It further implies selecting relevant actions while inhibiting inappropriate ones and eventually executing one in a timely manner. Following dual-process theory in the psychology of human reasoning [5], already successfully applied to domains such as medical decision-making [34], we can describe two underlying imaginary operating and cooperative agents, called System 1 (S_1) and System 2 (S_2). S_1 is the fast intuitive and heuristic agent while S_2 is the slow and reasoning agent. S_2 is the one responsible for decisions assisted by S_1 which continuously and automatically provides him predictions for action. Most of the time, S_2 just lazily accepts S_1 proposal in usual

situations without much more thinking, resulting in successful, quick and cognitively effortless decisions. When confronted with unusual situations, however, S_2 can develop more explicit conscious thinking, beyond S_1 predictions, to deliberate and come up with novel and acceptable decisions while cognitively costly.

Young operators will rely more heavily on S_2 and can hence struggle to take any good decision on time for several situations that still appear complex and unusual to them. As they are very focused on trying to make sense of it, they have narrow attention and might miss important new information. As they learn overtime through appropriate training and feedback become expert, climbing through the expertise ladder [35], their intuitive S_1 grows for that field of expertise. This enables them to make good and quick decisions even more often with ease. For an expert operator, it has become a lot easier to operate a system intuitively, being able to make more decisions as well as decisions in more difficult situations [36]. However, the downside can be overconfidence, overlooking unusual information that would require more deliberation from S_2 .

3.2. Human biases and desirable assistance

Because it relies on fast heuristics and mostly jumps to conclusions, System 1 indeed introduces several potential biases which can lead to human errors, hence limiting the effectiveness of human decision-making. Cognitive biases are summarized in the cognitive bias codex [37] and classified through 4 problems they are trying to circumvent: a limited memory, the need to act fast, the information overload and a lack of meaning. Among possibly damaging biases, we can more specifically list:

- anchoring bias: be over-reliant on the first piece of information we see.
- confirmation bias: paying more attention to information confirming our beliefs.
- overconfidence bias: too confident about one's abilities leads to greater risk-taking.
- information bias: tendency to seek information when it does not affect action (more information is not always better).
- availability heuristic: overestimate the importance of information that is available.
- ostrich effect: ignore dangerous or negative information.
- outcome bias: judge a decision based on the outcome rather than how it was made.

An assistant should hence help the operator avoid biases through assistance in:

- augmenting his memory, knowledge retrieval and keeping track of latest events.
- better information filtering or highlighting, enhancing attention focus.
- contextualizing a situation and giving feedback.
- making recommendations, possibly handling some tasks or alerting on undesirable expected consequences.

Let's now consider what AI in its latest developments could bring in that regard.

3.3. AI potential for assistance

The recent deep learning revolution demonstrated some impressive practical abilities of AI by being able to digest a lot of information, memorize large historical datasets, and learn by imitation to infer quickly effective actions in context. Turing-award Yoshua Bengio recently described current deep-learning AI as a S_1 kind of intelligence [38],

while missing S_2 reasoning. This type of AI is presented as advanced pattern matching and recognition machines like S_1 [5], being coined as artificial intuition [39]. It however lacks the ability to reason about causality [40], hence lacking understanding and common sense. Yet Human and AI can be seen as complementary heterogeneous intelligence that could achieve a superior outcome when co-adapting [10] and developing hybrid intelligence [41] or human-centered AI [12]. This is best exemplified by Kasparov's "Centaur chess" concept [42], where humans play alongside machines to reach superior performances. When powering assistance systems, AI seems capable of overcoming some previously mentioned human limitations by enhancing S_1 operator's ability, whose S_2 remains in charge of final decisions. Some initial assistance for S_2 can nonetheless be considered through explanations or counterfactual reasoning if a simulator exists.

3.4. Pitfalls of Human-Machine hybridization

As we have seen in the previous section, human-AI partnerships shows many promises. But such systems must be carefully designed, as numerous shortcomings might arise. We can make the analogy to "ironies or myth of automation" [25,43] which warns that operators should ultimately be more skilled than less skilled, and less loaded than more loaded to deal with the most difficult new and complex situations. Amongst many shortcomings mentioned in [44], over-reliance and deskilling are one of the most damaging for control rooms. As of today, operators have a deep knowledge of their grid area, both in terms of infrastructure and electrical phenomena. Adding an automatic system with persuasive recommendations could lead the operating staff to rely too heavily on the system, progressively losing their expertise, with potentially catastrophic consequences when it fails. Overconfidence should not be a trade-off with over-reliance, and operators should have the opportunity to keep developing their skills and cognitive strategies through regular training and manual problem-solving [25].

Additionally, an assistant could misunderstand user intent, and hinder his actions in an example of perverse instantiation [45]. These limitations arise from the fact that, while human Systems 1 and 2 are fully integrated into a single cognitive ensemble, AI and humans are separated entities, where communication is key. System 1 works on learned representations of reality, not directly on reality itself, meaning that situations can be interpreted differently. Hence, reference representations should at least be shared in some form between agents to properly communicate, as with language. Not an easy task knowing that, for instance, in language, common words can sometimes have different meanings between people, highlighting not always properly shared representations [46].

Finally, we should resist the temptation of developing an anthropomorphic assistant [47], i.e. not mechanically mimicking humans and human-human interactions. We aim at augmenting the operator, not replacing him, by recognizing the complementary differences between humans and computers [13], and leveraging them for humans. These motivate the need for design guidelines, which we will define in sections 5 and 6, including carefully-crafted human-machine interactions [48], or interpretable [49], explainable [50] and trustworthy [51] AI. We aim at offering operators high control level despite increasing grid automation without compromising reliability and safety [52].

Moving away from these maturing fields that should prove useful to create an assistant, let's define this concept more precisely for our power grid application domain.

4. Defining an artificial assistant for power grid operators

We formalize an assistant as an artificial agent providing help on a subset of user tasks within its daily activity, with the final goal of increasing task efficiency while still developing operator skills.

4.1. Assistant: balancing assistance, user control and automation

To make things clearer, we should explain how the concept of assistant articulates with assistance functions and is distinct from a completely autonomous system. Assistance functions help users with domain-specific tasks, for example, by alerting when relevant information arrives, or monitoring user context and warning him about unforeseen risks. Situation awareness [8,3,7] offers, in that sense, advanced assistance functions.

Other similar industrial sectors evoke those functions through different autonomy levels [53,54,55], that we can reflect on. We here focus on the Grades of Automation (GoA) definition from the International Association of Public Transport [56,54]:

- GoA0: Manual operation with no automatic protection
- GoA1: Manual operation with automatic protection
- GoA2: Semi-automatic operation
- GoA3: Driverless operation
- GoA4: Unattended operation

GoA1 and GoA2 offer assistance functions as discussed previously, but without much consideration for HMI. GoA3 and GoA4 are targeting autonomy through automation. It highlights that many fields such as autonomous driving [53] aim at fully autonomous systems. They diverge from our goal of augmenting operators through an assistant. While being conceptually closer to GoA2, we move away from the usual automation typical of system engineering, to be rather operator centric.

An assistant, as we illustrate in Fig.1a and later discuss, is yet another level whose goal is to offer the right balance between user control and autonomy [57] for enhanced decision-making. It certainly has assistance functions at its core, but most importantly also engages actively with the user. It offers a unified interface and allows for dynamic bidirectional interaction with the user to cooperate efficiently on task completion and with others. It is essential to keep in mind that, in this paradigm, the responsibility of system management still ultimately falls on the operators of critical systems. While some tasks could eventually be automated away, it is paramount to avoid pitfalls such as operator deskilling (see section 3.4), often due to the “out-of-the-loop” effect [3]. Assistants should rather let operators remain in control, which will require more than automation supervision [58], and help reinforce their expertise, as emphasized in section 4.3. Finally, unlike GoA framework, getting a teamwork perspective is important, by taking into consideration the interactions between control room operators, to foster proper coordination and to allow for observability, predictability and directability. [59].

We now focus on core building blocks in creating such assistant, namely hypervision and bidirectional interaction. Applied AI bricks examples would be found in section 6.4.

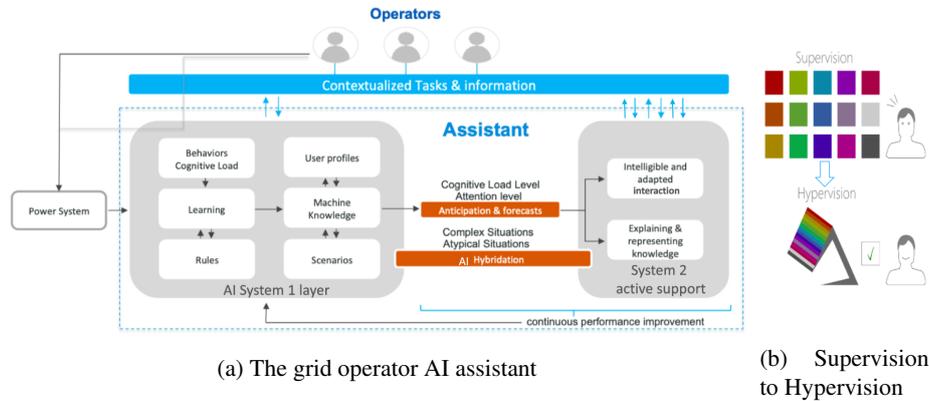


Figure 1. (a) The grid operator AI assistant relies on Hypervision interface represented as a blue bus, as well as underlying bidirectional interaction and AI components (see section 6.4) running altogether in a coordinated and modular fashion. Multiple Operator/Assistant pairs can also coordinate and collaborate through the Hypervision bus. Zooming in, Assistant System 1 type block helps in usual tasks or situations, exploiting core knowledge and intuition learnt. Right block is more dynamic and interactive, hence more bidirectional, to assist in situations that require more reasoning, focus, deliberation or exploration by operators in a more System 2 type fashion. Continuous revision is important to update shared representations. (b) Supervision to Hypervision in 1b allows moving away from alarm monitoring over many applications on screens (colored squares), to refocus the operator on task completion on a unified but contextualized single interface (mixed colors of merged information from background apps).

4.2. Hypervision: smart interface & information management

Today's supervision leaves to the user the cognitive load to prioritize, organize, and link every displayed information and alarms consistently before considering any decision. It can be regarded as a fragmented ecosystem from an operator's viewpoint. While it has been manageable for up to ten applications, it becomes impractical with always more information and uncoordinated applications to control under heterogeneous formats. Supervision gives access to the user to every information available without much more processing. However, it does not help deal with the information overload and "lack of meaning" problems (see 3.2) that need to be tackled for improved decision-making: it dilutes the operator's attention. Let's recall that humans can mainly focus on one task at a time, with a limited working memory space of few information chunks to manipulate [60].

To be effective at continuous decision-making, it is often important to focus on the highest priority task at a time, and present only the most relevant information to it. Yet, pausing on one task, dealing with another, and eventually completing the previous one should remain be possible. In that regard, we propose to rely on an "hypervision" framework to bring the right information at the right time to the right person [1] while keeping track of user progress for each task. It helps overcome multiple biases, such as both information bias and anchoring bias, by taking advantage of them rather than being influenced by them. Hypervision as presented in [1] relies on the definition of tasks created by processing and synthesizing the necessary information. Those tasks do not have to be only-real time. Indeed, they are still preferably the ones anticipated to be completed or configured ahead of time thanks to forecast, hence defining an expected trajectory that might be adapted along the way. Reaching this higher level of information enables the

June 2022

assistant to establish a simplified but relevant dialogue with the operator, eventually providing him with diagnostics or even recommendations on solutions. Hence, hypervision's goal is to help refocus the operator on task completion rather than alarm monitoring, as illustrated in figure 1b. While the assistant could instantiate tasks for anticipated risks by the system and help in prioritizing them, the operator should remain free to create and modify some himself, or manually change the order of priority. This manual editing is also important to correct overtime any selection bias that could have been introduced by automatic prioritization, among others. Finally, such tasks can be easily shared and completed between operators, allowing for easier team workload management [33]. These create the basis for more advanced and effective bidirectional interaction under shared representations of tasks [57] thanks to which users and assistants could work in tandem or in teams to achieve a common objective.

4.3. Bidirectional interaction

While the choice of the interaction modalities (visual, audio, haptic, etc.) matters for enhanced human-assistant partnership, we will leave it open here and focus on the importance of bidirectional interaction between an operator and its assistant.

Put in the spotlight in the 80's, expert systems developed with AI under predefined rules raised concerns about its practicality for human users in terms of Human-Machine interaction. Lucy Suchman [61] shed light on their ineffectiveness, mostly attributed to the lack of well-designed interaction and learning loops beyond knowledge retrieval. She noted that plans, similarly to predefined rules, are not prescriptive and not something to follow exactly, because everything eventually depends on circumstances and contingencies. Plans should rather be seen as heuristic and available resources for actions that help focus one's attention while abstracting the details. But they should get updated through interaction to take proper decisions. In the end, interfaces should not draw a dry delimitation with their user but re-configure themselves and conform with him.

Research [62] has shown an increased efficiency in Human-AI coupling when both agents were able to initiate and respond to interaction. These were historically mostly unidirectional, the assistant either asking a predefined set of questions to build its context representation or the user asking to perform some predefined tasks. Such badly designed assistant such as Clippy [63] could in the end disrupt the user, making it inefficient and frustrating. In a bidirectional relationship, the interaction is collaborative, with neither the system nor the user in control of the whole interaction [57]. The assistant is capable of interacting with the latter to refine its context representation (e.g. ask for a clarification when ambiguities arise), thus improving its efficiency when asked to perform a specific task. A good example of such bidirectional interaction is found in [48], where interaction between load-carrying robots and their human partners is learnt over time, resulting in better task completion through operator improvement and per-user adaptation of the robot's behavior. Further approaches let an assistant learn how and when to defer to an expert [64] when in doubt, or conversely let a user interactively learn about the weaknesses and strengths of the assistant under sources of uncertainties [65]. An assistant can also gently challenge or guide a user to reach novel solutions and build knowledge altogether [66,67]. Ultimately a user would like to grasp the underlying assistant model [68], to know what the model captured, what he can do with it and how to direct or correct it. Interpretability is at play more than post-hoc explanations [69]. In that regard, prospec-

tive design [12,70] through interaction and exploration should prove effective, removing the need for the assistant to always explain himself to the operator. An after-operations review process could prove useful for strengthening the relationship [71].

These advances show that creating true human-computer partnerships [72] based on the concepts of Discoverability, Appropriability and Expressivity become a reality as well-demonstrated in [73]. This calls for more academic and industrial collaboration like the “Cockpit and Bidirectional Assistant” project [74] for critical systems. In the next section, we review some design guidelines to initiate such partnerships in our use case.

5. Featured function for effective assistant support

After defining *what* an assistant should look like in the case of power systems, we now propose to specify *how* it should proceed when interacting with a user. To do so, we build upon the human-ai guidelines defined by Amershi et Al. [75] to feature import functions for such assistant. As these guidelines were initially defined from studying consumer products, we highlight some additional specificity when considering industrial systems.

Show contextually relevant information at the right time - Grid operators evolve in a time-constrained environment where having the right information at the right time is paramount. For instance, power lines reconnected after a routine outage operation for maintenance should be notified as soon as possible to increase grid robustness. An assistant should engage in interaction when the context allows it, taking into account when possible the operator’s mental availability and current task, and the expected impact of the interaction. Task prioritization is also a cornerstone of good grid management, and users should be provided with high granularity, curated task details for fast criticality assessment. All of the above-mentioned concerns call for efficient knowledge management and representation, as instantiated by our hypervision system defined in 4.2.

Scope services and inform the user when in doubt - Doubt can happen when the assistant is uncertain about the user’s goal, but also in our case because of assistant model limitations (imperfect grid simulator) or uncertainties in the system (poor weather forecast). Operators are often dealing with variability, be it when anticipating potential hazards due to volatile renewable energies or exploring the effects of preventive actions.

Let’s zoom on the “Anticipation & forecasts” slot of fig. 1a. When facing uncertainty, for example when provided with a highly volatile wind forecast, the operator can first simply indicate his intention of monitoring more stable, aggregated regional forecasts and let the assistant switch to a higher resolution as operations shift closer to real-time and more precise weather information arrives. This example also shows that uncertainties when providing assistance should be jointly presented with their probable causes (missing data, poor forecast...), to help decision-making and reinforce operator trust.

Support efficient invocation and dismissal - The number of actions an operator can do in a time window is limited. Interacting with an assistant has to be straightforward. Verbosity should be kept to a minimum, and function of the operator context. In tense situations (e.g. volatile renewable production), the assistant should initiate interactions more frugally, be more succinct, and support faster dismissal. Knowing how to adapt interactions is not straightforward, as it involves capturing a lot of implicit contextual hints, and getting operator feedback through dialog should play a major role.

Take into account and learn from user behavior and feedback - Grid operators are well-trained experts, capable of evaluating the assistant’s answers and providing feedback. Thus, to ensure continuous performance improvement (Fig. 1a), the latter must be

able to learn from users, for instance by understanding that additional context needs to be considered alongside a specific action, or remembering that a line is under maintenance during a user-provided period. Remembering recent interactions is paramount to user acceptance (nobody likes repeating requests) and capturing user context.

Moreover, each operator has a personal decision-making style, some relying on numerous power-flow simulations to assess a situation, others relying on their expertise. A good assistant should adjust to these user-specific profiles.

Convey the consequences of user actions - Not only should an assistant avoid operator deskilling for what was already done effectively, but also should it reinforce user expertise through interaction. In power systems, it is often deplored that the consequences of grid operations are poorly monitored, which in return deprives operators of valuable feedback. Assistants delivering a detailed report of how the grid evolved after a specific action would tremendously speed up the way operators acquire experience by mobilizing their deliberative thinking process (S2 in section 3), and yield better grid management. As what KPIs should be tracked is task-dependent, the operator is also implied in this bidirectional dialog. Logically, interpretability and explainability are inseparable from building this successful human-AI hybridization (fig. 1a).

We have seen that human-assistant partnership could be built upon hypervision (4.2), bidirectional interactions and a set of key features. We will now propose some guidelines on more concretely developing, implementing and testing such an assistant.

6. Guidelines for developing, implementing and testing an assistant

Designing an assistant in practice might still seem complex beyond the discussed framework and principles. We devise here some pragmatic guidelines to start simple on a common but modular ground, listing some already available building blocks as well.

6.1. Grounded Design Considerations

6.1.1. Tasks and Visualizations as shared representations

In smart grids, functions have been presented in [76,77,1] and tasks described at a high level in [3,26] or through a detailed example [36]. In other industrial sectors such as aeronautics [11], tasks in processes have been codified more precisely at a granular level with ontologies [78] or conceptual designs from human-computer design [11], which gives the operator a clearer framework to work and coordinate with, as for the assistant.

A task is first defined by the problem to solve specifically, such as a safety problem - a line overload, its priority and the residual time to complete it. It should contain relevant context to understand the root of the problem, what might be already known about it, recent related events or tasks, as well as the persons involved. Building on causal and counterfactual models [79] is desirable. A task should further come with suggestions about available actions, and their expected effectiveness. It should finally retain a decision for completion and meta-attributes about it. Structuring tasks this way would allow shared representations [57] between the operator and the assistant. Task categories and attributes should be more exhaustively drawn through future works. Eventually, comprehensive activity studies involving multiple tasks in time [11] should be run.

Also, unlike traditional approaches in power systems that mainly tries to focus on the most critical situations we ever have to solve, we suggest here to start studying tasks in regular situations and gradually increasing the number of needed bidirectional inter-

action. To operators, it should prove useful to start experimenting on the most routine but sometimes time-consuming tasks with often low added value [80]. That way, building trust in the first place should be easier while still helping ease their cognitive load.

Additionally, powerful and interactive visualizations are a common and much appreciated approach for operators to support shared representations and task completion. They are useful companions of an assistant. A recent survey in power grids advocates for such new developments [81]. Effective superimposed forecasted and current system state visualization [82], dynamic and temporal network exploration [83,84], and high-dimensional event-based visualizations [85] could prove beneficial in that regard.

6.2. Simple situational use case as a sandbox

We offer a simple interesting use case to highlight key difficulties in daily grid management through the interplay of preventive and corrective decisions under uncertainty. This makes us think about how the operator-agent interaction should take place. An operator starts monitoring a two-area grid at 7:00am. Forecasts show 2 issues:

- An incident in area 1 could happen around 9:00am and would lead to some overloads, with three available corrective actions after simulations.
- Another such incident in area 2 could happen around 8:30am with only one preventive action available. This leaves only a couple of minutes to execute it or not.

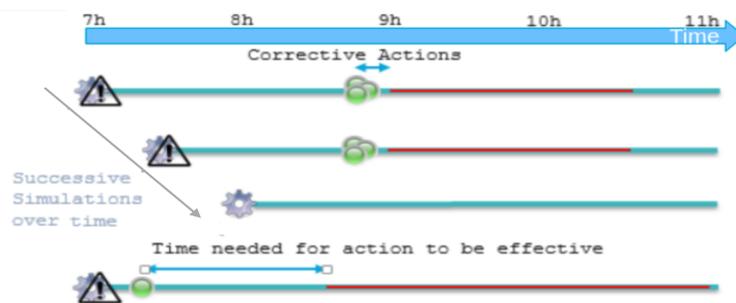


Figure 2. A simple scenario where incidents are forecasted in multiple parts of a grid, and several corrective actions with different setup duration are possible. Green dots each represent an available action, and the blue arrows the time it needs to be effective. For instance, starting a nuclear power plant takes longer than applying topological grid changes, and would have a longer blue arrow.

A couple of questions arise on the best strategy to follow:

- Which decisions have priority? It seems that a preventive action on area 2 should be urgently taken, but maybe the forecast isn't that reliable yet.
- When should we implement actions? Waiting for the last simulation at 8:00am in area 1 shows that the forecasted problem disappeared which is the best option.
- How does applying a preventive action on area 2 reflect on area 1? Would it lead to a less secure grid state? What coordination is required? Maybe there's a new outage operation in this area that isn't yet taken into account by the simulation.
- Which of the three corrective actions in area 1 should be taken? The operator has to mediate between economical, practical and safety arguments, each with a degree of uncertainty over an activation horizon.

June 2022

Our objective here is not to provide any viable solution, but rather to demonstrate that grid operators are confronted with complex decisions even on apparently simple cases, in which context-dependent trade-offs always have to be made. Future works could build a library of such canonical cases to be studied in the community.

6.3. Unified Interface & Data collection as an industrial stack

The hypervision framework relies on a generic single interface that should be able to integrate any kind of tasks, and apply to different industrial systems for instance. An example of an existing framework is the open-source Operator Fabric [86]. It could be used both by industrial and researchers as a unified interface for decision-making processes across teams. Such a framework is also a cornerstone to digitalize the decision-making process, centralize every necessary information, hence capitalize on them. This historical data collection is essential for continuous improvement, experiments, as well as for creating datasets from which AI can learn recommendations. Data should get labeled and its quality properly monitored. These developments should create a necessary technical stack or data platform for an assistant, to overcome challenges in deploying AI [87,88].

6.4. Power system AI modules for assistant functions

Recent surveys list interesting developments of AI for climate change [89] and for power systems more specifically [90], [91]. For an assistant, AI can today be used to make corrective action recommendations to an operator through adaptive interpretable expert system [23], imitation learning [22] or reinforcement learning on robustness, adaptability or trust dimensions [92,93] which are key in Human-centered AI [52]. It can learn from user behaviour and help convey the consequences of the operator's action by comparison. Exhaustive risk assessment [94] also helps in prioritizing tasks. Further, automatic hierarchical and contextual representations of the grid [95] enable scope services and give greater flexibility to convey the right context and interpret a situation. [96] also lets an AI learn interpretable and physically-consistent contextual indicators associated with a particular operator's task or help build knowledge graphs [97]. Finally, [98,99,100] let operators explore interactively and iteratively historical explainable factors across similar situations and decisions for augmenting and keeping up-to-date the system knowledge and proper labels. This should all be carefully developed within a trustworthy framework [101]. This is an illustrative sample of today's AI potential [102,103] to provide effective assistance functions and interactions which needs to be developed further.

6.5. Assistant evaluation & development of shared benchmarks

In order to assess the relevance of an assistant in a real-world scenario, and eventually compare multiple assistants, it is necessary to set up repeatable evaluation protocols and define common benchmark tasks.

As for now, there is not yet a standard testing protocol to evaluate AI assistants. However, we could draw insights from other domains such as interpretable machine-learning [49], explainable AI [104,105], or interactive visualization [106]. As done in [49], we could come to structured step-by-step experimental practices to evaluate candidate assistants on incrementally difficult tasks. Moreover, several Virtual-Assistant (VA) related studies have also tried to define custom evaluation criteria. In [107], authors

compare their VA against both a simpler interactive data-exploration scheme and a non-interactive solution-search. They assess the use of their assistant on three factors: *performance* with task-specific metrics; *human-learning* using questions and tests at the end of each task; *usability* for example using System Usability Scale [108]. During this evaluation biases, over-reliance and deskilling (from sections 3.2 and 3.4) can be addressed.

Because of confidentiality issues, it is often hard to share real-world data on decision-making problems. Thus, we should aim at developing synthetic but realistic environments from which to extract representative and relevant scenarios, for example by drawing inspiration from road safety decision-making assessment frameworks [109]. Moreover, synthetic and realistic environments for sequential decision-making have recently been developed for power grids [110,92,93] and could be interactively further studied with grid2viz [111] and grid2grame [112] study tools.

7. Conclusion

In this paper, we have presented the framework and principles for designing an AI assistant based on the concept of Hypervision and bidirectional interactions for power grid operators. It combined insights from various research fields, opening new research directions for augmented decision-making. We have provided initial guidelines of expected functions and practical needs, as well as already available materials in power grids, to continue exploring this promising and very much needed new field of human-machine partnership. Building on this proposed framework and existing L2RPN with Trust environment [93], future work should aim soon at extending testbed environments and instantiate first complete implementations of assistant to be benchmarked, in order to operate the grid with greater flexibility and coordination to support the ongoing energy transition.

References

- [1] A. Marot, A. Kelly, M. Naglic, V. Barbesant, J. Cremer, A. Stefanov, and J. Viebahn, "Perspectives on future power system control centers for energy transition," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 328–344, 2022.
- [2] F. Allgoewer, J. B. de Sousa, J. Kapinski, P. Mosterman, J. Oehlerking, P. Panciatici, M. Prandini, A. Rajhans, P. Tabuada, and P. Wenzelburger, "Position paper on the challenges posed by modern applications to cyber-physical systems theory," *Nonlinear Analysis: Hybrid Systems*, vol. 34, pp. 147–165, 2019.
- [3] A. M. Prostejovsky, C. Brosinsky, K. Heussen, D. Westermann, J. Kreusel, and M. Marinelli, "The future role of human operators in highly automated electric power systems," *Electric Power Systems Research*, vol. 175, p. 105883, 2019.
- [4] E. Flaspöler, A. Hauke, P. Pappachan, D. Reinert, T. Bleyer, N. Henke, and R. Beeck, "The human machine interface as an emerging risk," *EU-OSHA (European Agency for Safety and Health at Work). Luxemburgo*, 2009.
- [5] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.
- [6] L. Naccache, S. Dehaene, L. Cohen, M.-O. Habert, E. Guichart-Gomez, D. Galanaud, and J.-C. Willer, "Effortless control: executive attention and conscious feeling of mental effort are dissociable," *Neuropsychologia*, vol. 43, no. 9, pp. 1318–1328, 2005.
- [7] M. Naderpour, J. Lu, and G. Zhang, "An intelligent situation awareness support system for safety-critical environments," *Decision Support Systems*, vol. 59, pp. 325–340, 2014.
- [8] M. Panteli and D. S. Kirschen, "Situation awareness in power systems: Theory, challenges and applications," *Electric Power Systems Research*, vol. 122, pp. 140–151, 2015.

- [9] Y.-b. Liu, J.-y. Liu, G. Taylor, T.-j. Liu, J. Gou, and X. Zhang, "Situational awareness architecture for smart grids developed in accordance with dispatcher's thought process: a review," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 11, pp. 1107–1121, 2016.
- [10] W. Mackay, "Responding to cognitive overload: Co-adaptation between users and technology," *Intellectica*, vol. 30, no. 1, pp. 177–193, 2000.
- [11] G. A. Boy, "Human-centered design of complex systems: An experience-based approach," *Design Science*, vol. 3, 2017.
- [12] B. Shneiderman, *Human-Centered AI*. Oxford University Press, 2022.
- [13] J. C. Licklider, "Man-computer symbiosis," *IRE transactions on human factors in electronics*, no. 1, pp. 4–11, 1960.
- [14] C. N. Klokmoose, J. R. Eagan, S. Baader, W. Mackay, and M. Beaudouin-Lafon, "Webstrates: shareable dynamic media," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 280–290, 2015.
- [15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [16] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, and H. Laga, "A comprehensive survey of deep learning for image captioning," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1–36, 2019.
- [17] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.
- [18] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–38, 2019.
- [19] R. Winkler, M. Söllner, M. L. Neuweiler, F. Conti Rossini, and J. M. Leimeister, "Alexa, can you help us solve this problem? how conversations with smart personal assistant tutors increase task group outcomes," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019.
- [20] R. McIlroy-Young, S. Sen, J. Kleinberg, and A. Anderson, "Aligning superhuman ai with human behavior: Chess as a model system," in *Proceedings of the 25th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2020.
- [21] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. d. O. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, *et al.*, "Evaluating large language models trained on code," *arXiv preprint arXiv:2107.03374*, 2021.
- [22] B. Donnot, I. Guyon, M. Schoenauer, P. Panciatici, and A. Marot, "Introducing machine learning for power system operation support," *arXiv preprint arXiv:1709.09527*, 2017.
- [23] A. Marot, B. Donnot, S. Tazi, and P. Panciatici, "Expert system for topological remedial action discovery in smart grids," 2018.
- [24] P. Clutton-Brock, D. Rolnick, P. L. Donti, and L. Kaack, "Climate change and ai. recommendations for government action," tech. rep., GPAI, Climate Change AI, Centre for AI & Climate, 2021.
- [25] L. Bainbridge, "Ironies of automation," in *Analysis, design and evaluation of man-machine systems*, pp. 129–135, Elsevier, 1983.
- [26] S. Abram and A. Silvast, "Flexibility of real-time energy distribution: the changing practices of energy control rooms," 2021.
- [27] R. C. G. Teive, A. L. R. Rese, and J. P. Parreira, "An intelligent system for operators performance multicriteria evaluation of distribution networks," *Applied Artificial Intelligence*, pp. 1–24, 2022.
- [28] M. M. Bhaskar, M. Srinivas, and S. Maheswarapu, "Security constraint optimal power flow(scopf)- a comprehensive survey," *Global Journal of Technology and optimization*, vol. 2, no. 11, 2011.
- [29] M. Heidarifar, P. Andrianesis, P. Ruiz, M. C. Caramanis, and I. C. Paschalidis, "An optimal transmission line switching and bus splitting heuristic incorporating ac and n-1 contingency constraints," *International Journal of Electrical Power & Energy Systems*, vol. 133, p. 107278, 2021.
- [30] C. Möhrle, G. Giebel, R. J. Bessa, and N. Fleischhut, "How do humans decide under wind power forecast uncertainty—an iea wind task 36 probabilistic forecast games and experiments initiative," in *Journal of Physics: Conference Series*, vol. 2151, p. 012014, IOP Publishing, 2022.
- [31] RTE, "Energy pathways to 2050," tech. rep., Réseau de Transport d'électricité, 2021.
- [32] M. Ghalenoeei, S. B. Mortazavi, A. Mazloumi, and A. H. Pakpour, "Impact of workload on cognitive performance of control room operators," *Cognition, Technology & Work*, vol. 24, no. 1, pp. 195–207,

- 2022.
- [33] M. Harbers and M. A. Neerinx, "Value sensitive design of a virtual assistant for workload harmonization in teams," *Cognition, Technology & Work*, vol. 19, no. 2, pp. 329–343, 2017.
 - [34] B. Djulbegovic, I. Hozo, J. Beckstead, A. Tsalatsanis, and S. G. Pauker, "Dual processing model of medical decision-making," *BMC medical informatics and Decision Making*, vol. 12, no. 1, pp. 1–13, 2012.
 - [35] S. M. Stevens-Adams and F. Hannigan, "Defining expertise in the electric grid control room.," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2016.
 - [36] E. P. R. Institute, "Structured decision-making techniques in transmission control centers," 2021.
 - [37] "Cognitive bias codex." <https://busterbenson.com/piles/cognitive-biases/>.
 - [38] Y. Bengio, "From system 1 deep learning to system 2 deep learning," *Neural Information Processing Systems. December 11th*, 2019.
 - [39] C. E. Perez, *Artificial Intuition: The Improbable Deep Learning Revolution*. Carlos E. Perez, 2018.
 - [40] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," *arXiv preprint arXiv:1801.04016*, 2018.
 - [41] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister, "Hybrid intelligence," *Business & Information Systems Engineering*, vol. 61, no. 5, pp. 637–643, 2019.
 - [42] N. Case, "How to become a centaur," *Journal of Design and Science*, 2018.
 - [43] J. M. Bradshaw, R. R. Hoffman, D. D. Woods, and M. Johnson, "The seven deadly myths of' autonomous systems'," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 54–61, 2013.
 - [44] I. Seeber, E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Söllner, "Machines as teammates: A research agenda on ai in team collaboration," *Information Management*, vol. 57, no. 2, p. 103174, 2020.
 - [45] B. Goertzel, "Superintelligence: Fears, promises and potentials: Reflections on bostrom's superintelligence, yudkowsky's from ai to zombies, and weaver and veitas's "open-ended intelligence"," *Journal of Ethics and Emerging Technologies*, vol. 25, no. 2, pp. 55–87, 2015.
 - [46] L. Marti, S. T. Piantadosi, and C. Kidd, "Same words, same context, different meanings: People are unaware their own concepts are not always shared.," in *CogSci*, pp. 2296–2302, 2019.
 - [47] L. Mumford, *Technics and civilization*. University of Chicago Press, 2010.
 - [48] K. Blanchet, A. Bouzeghoub, S. Kchir, and O. Lebec, "How to guide humans towards skills improvement in physical human-robot collaboration using reinforcement learning?," in *SMC 2020: IEEE International Conference on Systems, Man, and Cybernetics*, (Toronto (online), Canada), pp. 4281–4287, IEEE Computer Society, Oct. 2020.
 - [49] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.
 - [50] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
 - [51] E. Commission, "White paper on artificial intelligence—a european approach to excellence and trust," 2020.
 - [52] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020.
 - [53] I. Barabas, A. Todoruț, N. Cordoș, and A. Molea, "Current challenges in autonomous driving," in *IOP conference series: materials science and engineering*, vol. 252, p. 012096, IOP Publishing, 2017.
 - [54] N. Brandenburger and A. Naumann, "On track: A series of research about the effects of increasing railway automation on the train driver," *IFAC-PapersOnLine*, vol. 52, no. 19, pp. 288–293, 2019. 14th IFAC Symposium on Analysis, Design, and Evaluation of Human Machine Systems HMS 2019.
 - [55] J. Frohm, V. Lindström, M. Winroth, and J. Stahre, "Levels of automation in manufacturing," *Ergonomics*, 2008.
 - [56] J. M. Cohen, A. S. Barron, R. J. Anderson, and D. J. Graham, "Impacts of unattended train operations on productivity and efficiency in metropolitan railways," *Transportation Research Record*, vol. 2534, no. 1, pp. 75–83, 2015.
 - [57] J. Heer, "Agency plus automation: Designing artificial intelligence into interactive systems," *Proceedings of the National Academy of Sciences*, vol. 116, no. 6, pp. 1844–1850, 2019.
 - [58] M. M. Bailey, "Training human operators to supervise automation,"
 - [59] M. Johnson, J. M. Bradshaw, and P. J. Feltovich, "Tomorrow's human–machine design tools: From levels of automation to interdependencies," *Journal of Cognitive Engineering and Decision Making*,

- vol. 12, no. 1, pp. 77–82, 2018.
- [60] N. Cowan, “The magical number 4 in short-term memory: A reconsideration of mental storage capacity,” *Behavioral and brain sciences*, vol. 24, no. 1, pp. 87–114, 2001.
- [61] L. A. Suchman, *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.
- [62] E. Horvitz, “Principles of mixed-initiative user interfaces,” in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 159–166, 1999.
- [63] N. Baym, L. Shifman, C. Persaud, and K. Wagman, “Intelligent failures: Clippy memes and the limits of digital assistants,” *AoIR Selected Papers of Internet Research*, 2019.
- [64] H. Mozannar and D. Sontag, “Consistent estimators for learning to defer to an expert,” *arXiv preprint arXiv:2006.01862*, 2020.
- [65] T. Sanchez, B. Caramiaux, P. Thiel, and W. E. Mackay, “Deep learning uncertainty in machine teaching,” in *27th International Conference on Intelligent User Interfaces*, pp. 173–190, 2022.
- [66] W. Liu, R. L. d’Oliveira, M. Beaudouin-Lafon, and O. Rioul, “Bignav: Bayesian information gain for guiding multiscale navigation,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5869–5880, 2017.
- [67] J. Malloch, C. F. Griggio, J. McGrenere, and W. E. Mackay, “Fieldward and pathward: Dynamic guides for defining your own gestures,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4266–4277, 2017.
- [68] S. Conversy, J. Garcia, G. Buisan, M. Cousy, M. Poirier, N. Saporito, D. Taurino, G. Frau, and J. Debatista, “Vizir: A domain-specific graphical language for authoring and operating airport automations,” in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 261–273, 2018.
- [69] C. Rudin, “Please stop explaining black box models for high stakes decisions,” *Stat.*, vol. 1050, p. 26, 2018.
- [70] R. Hull, B. Kumar, D. Lieuwen, P. F. Patel-Schneider, A. Sahuguet, S. Varadarajan, and A. Vyas, ““everything personal, not just business:” improving user experience through rule-based service customization,” in *International Conference on Service-Oriented Computing*, pp. 149–164, Springer, 2003.
- [71] J. Dodge, R. Khanna, J. Irvine, K.-h. Lam, T. Mai, Z. Lin, N. Kiddle, E. Newman, A. Anderson, S. Raja, et al., “After-action review for ai (aar/ai),” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 11, no. 3-4, pp. 1–35, 2021.
- [72] M. Beaudouin-Lafon, S. Bødker, and W. E. Mackay, “Generative theories of interaction,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 28, no. 6, pp. 1–54, 2021.
- [73] J. Koch, N. Taffin, M. Beaudouin-Lafon, M. Laine, A. Lucero, and W. E. Mackay, “Imagesense: an intelligent collaborative ideation tool to support diverse human-computer partnerships,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW1, pp. 1–27, 2020.
- [74] “Cockpit and bidirectional assistant project.” https://www.irt-systemx.fr/wp-content/uploads/2020/10/SystemX_201015_projetCAB_Final.pdf73.pdf.
- [75] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al., “Guidelines for human-ai interaction,” in *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–13, 2019.
- [76] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, “Smart transmission grid: Vision and framework,” *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 168–177, 2010.
- [77] A. Gopstein, C. Nguyen, C. O’Fallon, N. Hastings, D. Wollman, et al., *NIST framework and roadmap for smart grid interoperability standards, release 4.0*. Department of Commerce. National Institute of Standards and Technology, 2021.
- [78] M. Reiss, M. Moal, Y. Barnard, J.-P. Ramu, and A. Froger, “Using ontologies to conceptualize the aeronautic domain,” in *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics*. Cépaduès-Éditions, Toulouse, France, pp. 56–63, Citeseer, 2006.
- [79] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic books, 2018.
- [80] M. R. Endsley, “Level of automation forms a key aspect of autonomy design,” *Journal of Cognitive Engineering and Decision Making*, vol. 12, no. 1, pp. 29–34, 2018.
- [81] M. T. Fischer, “Towards a survey of visualization methods for power grids,” *arXiv preprint arXiv:2106.04661*, 2021.
- [82] A. Prouzeau, A. Bezerianos, and O. Chapuis, “Towards road traffic management with forecasting on wall displays,” in *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and*

- Spaces*, pp. 119–128, 2016.
- [83] B. Bach, E. Pietriga, and J.-D. Fekete, “Graphdiaries: Animated transitions and temporal navigation for dynamic networks,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 5, pp. 740–754, 2013.
- [84] B. Bach, N. H. Riche, R. Fernandez, E. Giannidakis, B. Lee, and J.-D. Fekete, “Networkcube: bringing dynamic network visualizations to domain scientists,” in *posters of the Conference on Information Visualization (InfoVis)*, 2015.
- [85] A. Fouse, N. Weibel, E. Hutchins, and J. D. Hollan, “Chronoviz: a system for supporting navigation of time-coded data,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, pp. 299–304, 2011.
- [86] “Operator fabric framework.” <https://opfab.github.io/>.
- [87] A. Paleyes, R.-G. Urma, and N. D. Lawrence, “Challenges in deploying machine learning: a survey of case studies,” *arXiv preprint arXiv:2011.09926*, 2020.
- [88] L. Baier, F. Jöhren, and S. Seebacher, “Challenges in the deployment and operation of machine learning in practice.,” in *ECIS*, 2019.
- [89] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, *et al.*, “Tackling climate change with machine learning,” *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–96, 2022.
- [90] M. Kezunovic, P. Pinson, Z. Obradovic, S. Grijalva, T. Hong, and R. Bessa, “Big data analytics for future electricity grids,” *Electric Power Systems Research*, vol. 189, p. 106788, 2020.
- [91] L. Duchesne, E. Karangelos, and L. Wehenkel, “Recent developments in machine learning for energy systems reliability management,” *Proceedings of the IEEE*, 2020.
- [92] A. Marot, B. Donnot, G. Dulac-Arnold, A. Kelly, A. O’Sullivan, J. Viebahn, M. Awad, I. Guyon, P. Panciatici, and C. Romero, “Learning to run a power network challenge: a retrospective analysis,” in *NeurIPS 2020 Competition and Demonstration Track*, pp. 112–132, PMLR, 2021.
- [93] A. Marot, B. Donnot, K. Chaouache, A. Kelly, Q. Huang, R. Hossain, and J. L. Cremer, “Learning to run a power network with trust,” vol. abs/2110.12908, 2021.
- [94] B. Donnot, I. Guyon, A. Marot, M. Schoenauer, and P. Panciatici, “Optimization of computational budget for power system risk assessment,” in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, IEEE, 2018.
- [95] A. Marot, S. Tazi, B. Donnot, and P. Panciatici, “Guided machine learning for power grid segmentation,” in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, IEEE, 2018.
- [96] L. Crochepierre, L. Boudjeloud-Assala, and V. Barbesant, “Interpretable dimensionally-consistent feature extraction from electrical network sensors,” in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML/PKDD’20*, 2020.
- [97] H. Huang, Z. Hong, H. Zhou, J. Wu, and N. Jin, “Knowledge graph construction and application of power grid equipment,” *Mathematical Problems in Engineering*, vol. 2020, 2020.
- [98] A. Marot, A. Rosin, L. Crochepierre, B. Donnot, P. Pinson, and L. Boudjeloud-Assala, “Interpreting atypical conditions in systems with deep conditional autoencoders: the case of electrical consumption,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 638–654, Springer, 2019.
- [99] L. Boudjeloud-Assala, P. Pinheiro, A. Blansch e, T. Tamisier, and B. Otjacques, “Interactive and iterative visual clustering,” *Information Visualization*, vol. 15, no. 3, pp. 181–197, 2016.
- [100] D. Gkorou, M. Larranaga, A. Ypma, F. Hasibi, and R. J. van Wijk, “Get a human-in-the-loop: Feature engineering via interactive visualizations,” 2020.
- [101] J. Stiasny, S. Chevalier, R. Nellikkath, B. S avarsson, and S. Chatzivasileiadis, “Closing the loop: A framework for trustworthy machine learning in power systems,” *arXiv preprint arXiv:2203.07505*, 2022.
- [102] F. Li and Y. Du, “From alphago to power system ai: What engineers can learn from solving the most complex board game,” *IEEE Power and Energy Magazine*, vol. 16, 2018.
- [103] Z. Shi, W. Yao, Z. Li, L. Zeng, Y. Zhao, R. Zhang, Y. Tang, and J. Wen, “Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions,” *Applied Energy*, vol. 278, 2020.
- [104] D. Pruthi, B. Dhingra, L. B. Soares, M. Collins, Z. C. Lipton, G. Neubig, and W. W. Cohen, “Evaluating explanations: How much do explanations from the teacher aid students?,” *arXiv preprint*

June 2022

arXiv:2012.00893, 2020.

- [105] A. Rosenfeld, “Better metrics for evaluating explainable artificial intelligence,” in *Proceedings of the 20th international conference on autonomous agents and multiagent systems*, pp. 45–50, 2021.
- [106] R. Borgo, L. Micalef, B. Bach, F. McGee, and B. Lee, “Information visualization evaluation using crowdsourcing,” *Computer Graphics Forum*, vol. 37, 2018.
- [107] A. V. i. Martin and D. Selva, “Daphne: A virtual assistant for designing earth observation distributed spacecraft missions,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 30–48, 2020.
- [108] J. Brooke, “Sus: A ‘quick and dirty’ usability scale,” 1996.
- [109] A. Benabbou, D. Lourdeaux, and D. Lenne, “Generation of obligation and prohibition dilemmas using knowledge models,” pp. 433–440, 11 2017.
- [110] A. Marot, B. Donnot, C. Romero, B. Donon, M. Lerousseau, L. Veyrin-Forrer, and I. Guyon, “Learning to run a power network challenge for training topology controllers,” *Electric Power Systems Research*, vol. 189, p. 106635, 2020.
- [111] “Grid2viz study tool.” <https://github.com/mjothy/grid2viz>.
- [112] “Grid2game graphical user interface.” <https://github.com/BDonnot/grid2game>.