

Human-AI Interaction and Visualization Perspectives on ADR



Kostiantyn Kucher, Magnus Bång, and Jonas Lundberg

Abstract Recent advances in artificial intelligence (AI), data, and robotics (ADR) have pushed the boundaries of the benchmark performance of the respective methods and have already started to change the landscape in various application domains. Some of these domains are mission critical, with control activity that must match running processes. For those domains, a number of questions and challenges related to safety, robustness, and trustworthiness of AI and ADR methods and models still remain open, especially in the scenarios involving human operators. In this chapter, we provide an overview of human-centered perspectives on ADR with an emphasis on human-AI interaction, interactive visualization, and visual analytics. We explain the relationship of these fields to the related disciplines and fields, including human factors and human-computer interaction. We introduce the readers to basic concepts from these fields and discuss how the prior work fits with ADR principles, focusing on examples in visualization for explainable AI, cognitive systems engineering for joint human-AI control, and evaluation approaches for human-AI decision support systems. We argue that the techniques and frameworks proposed in these human-centered fields can and should be integrated with ADR methods.

Keywords Human-AI interaction · Information visualization · Visual analytics · Vis4ML · Joint human-AI control

K. Kucher (✉) · J. Lundberg
Department of Science and Technology, Linköping University, Norrköping, Sweden
e-mail: kostiantyn.kucher@liu.se; jonas.lundberg@liu.se

M. Bång
Department of Computer and Information Science, Linköping University, Linköping, Sweden
e-mail: magnus.bang@liu.se

1 Introduction

The remarkable progress achieved by researchers and practitioners in artificial intelligence (AI) and robotics over the past decade has benefitted from access to larger datasets, better hardware and infrastructure, and more ingenious algorithms. New workflows and paradigms have emerged with an emphasis on making use of existing models that are often very powerful but also very costly to train and highly complex with respect to their structure and behavior. In order to facilitate further progress in AI, data, and robotics (ADR)—and to enable successful applications of the respective advances across various domains and scenarios, including mission-critical ones—a number of human-centered challenges must be addressed. While it may seem at first glance that focusing on humans is not so relevant for the disciplines of AI and robotics, counter-arguments are quite straightforward. Stakeholders at various stages of the underlying theoretical research, implementation, deployment, and regulation for AI and robotics are humans: researchers, developers, decision-makers, technical support staff, operators, end users, general public, authorities, etc. (as illustrated in a simplified way in Fig. 1). These stakeholders typically require a certain level of understanding and insight (ranging from high-level familiarity to in-depth expertise) in order to approve, use, and trust the corresponding ADR tools or robots. Regulations existing in specific fields and disciplines under specific jurisdictions can impose explicit requirements for such ADR systems. For instance, the recent EU AI Act lists a number of requirements for human oversight of high-risk AI systems in Article 14,¹ including the requirements for supporting monitoring, interpretation, and intervention in the operations of such systems [1].

Thus, it becomes clear that the concerns related to various forms of interaction and teaming across humans and ADR systems are highly relevant for most (if not all) stages of ADR design, development, and deployment. Such concerns have been studied across several disciplines and fields, from applied psychology and human factors to autonomous systems and data science. More specifically, researchers and developers may benefit from the techniques that provide overview and support interactive exploration of the data and models in detail, especially for machine learning (ML) scenarios, including deep learning (DL) and large language models (LLMs); system operators and decision-makers require carefully designed user interfaces that facilitate rather than hinder their activities; and further stakeholders as well as the general public may benefit from high-level summaries that convey the main points about the respective ADR systems without excessive detail and clutter. To explore the possibilities for addressing these challenges, we focus on the concepts and methods from human-computer interaction, information visualization, and visual analytics relevant to ADR.

This chapter relates to several concerns among cross-sectorial AI, data, and robotics technology enablers of the AI, Data, and Robotics Partnership [2],

¹<https://artificialintelligenceact.eu/article/14/>; last accessed on June 1, 2025.

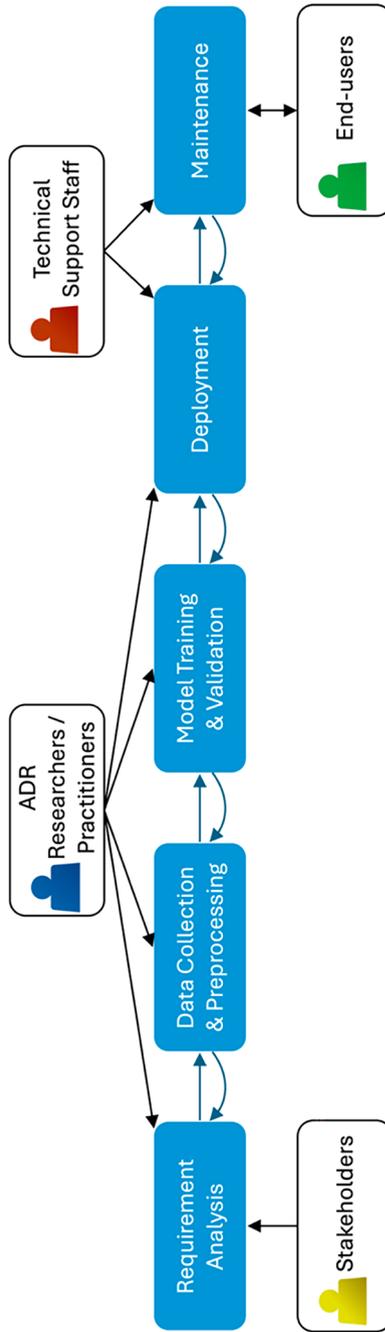


Fig. 1 A simplified model of ADR solution construction, deployment, and maintenance, including several key actors. While the focus of ADR research and applications may lie mainly on technical aspects, the involvement of humans in various roles alongside the pipeline (especially considering end-users) necessitates the involvement of human-centered methods and perspectives

including action and interaction, but also systems, methodologies, hardware, and tools. The rest of this chapter is organized as follows: in the next section, we start by introducing the basic concepts from the fields of human factors and human-computer interaction and then information visualization and visual analytics. Afterward, we discuss three topics that illustrate the potential of human-computer interaction and visualization applications for and interactions with ADR. Finally, we outline several potential directions for future work and conclude this chapter.

2 Background

In this section, we introduce the fields of human-computer interaction and information visualization (as well as several directly related fields) that is necessary for a further discussion of human-centered techniques and findings relevant to ADR in the next section.

2.1 *Human Factors and Human-Computer Interaction*

The discipline of *human factors* (or *ergonomics*) focuses on interactions within systems comprising humans and their environments, typically in the context of work (professional duties) performed by humans [3], but also services received by humans [4]. The environments include physical or virtual artefacts, organizational structures related to activities, and social structures related to other people. The focus of human factors research is thus to jointly improve the performance of the complete system, typically by fitting the environment to the human. Two major system outcomes can be achieved in this way through systems design: performance (productivity, quality, effectiveness, efficiency, etc.) and well-being (health and safety, satisfaction, learning, etc.) [4].

While the discipline of human factors has its roots in earlier research from the nineteenth century, it is generally considered that the contemporary discipline emerged in the late 1940s [3]. While the main concerns of human factors were originally related to *physical* ergonomics, with time, focus was also put on *cognitive* and *organizational* ergonomics. Cognitive aspects (especially from not only the reactive, summative perspective but also more proactive and potentially prescriptive one) have also been the subject of the field of cognitive systems engineering that is defined by Militello et al. as “an approach to the design of technology, training, and processes intended to manage cognitive complexity in sociotechnical systems” [5]. Relationship between these and further disciplines and fields is complicated by their interdisciplinary nature and existence of a rich lexicon of related terms and topics [6]. While the traditional human factors research and applications focusing on physical ergonomics had their roots in medical and biomechanical fields, the focus on

cognitive aspects required methods from applied psychology, systems engineering, and design.

Eventually, the discipline (or field) of *human-computer interaction* emerged on the intersection of human factors and computer science. Ebert et al. describe the goals and scope of human-computer interaction as “to improve the interaction between users and computing devices in the sense that this interaction should become more user-friendly and better adapted to the needs and capabilities of the users and the capabilities of the device” [7]. Dix mentions a number of disciplines and fields as overlapping with and helping define human computer interaction, but specifically, computer science, psychology, and ergonomics, afterward joined by social sciences due to the interest in sociotechnical design challenges [8]. One of the important topics in human-computer interaction is the design and evaluation of user interfaces. The methods from cognitive systems engineering can be applicable and helpful in this regard when considering support for *decision-making* in complex sociotechnical systems, in particular, the *ecological interface design* [9, 10]. This methodology relies on *work domain analysis* of the respective domain problem with a distinction between lower-level physical information and higher-level, more abstract functional information, which is relevant to the notion of *abstraction hierarchy*. Our previous work elaborated on the applications of this methodology for the air traffic control domain. Our initial interactive interface contribution for supporting “what-if” and “what-else” probes [11] was followed up by a study applying work domain analysis. The latter included design and evaluation of two ecological user interfaces for dealing with conflicts in air traffic control while aiming for safety, performance, and efficiency as functional purposes [12].

The development and adoption of technology has been naturally affecting human factors and human-computer interaction, initially with respect to physical but eventually also cognitive and organizational ergonomics. Bainbridge discussed a number of “ironies of automation”—challenges arising, among others, from the assumptions that replacing human operators with fully automated systems would only lead to benefits in the short and long term [13]. The issues that could ensue, however, include the loss of vigilance while monitoring a long-running process; losing the skills over a longer period of time; and reducing (or even eliminating) the opportunities to train new staff members. With recent advances in artificial intelligence, new ironies have emerged [14]. In addition to these concerns, the issue of *trust* in automation should also be considered [15], as both over- and under-trust can lead to suboptimal and even catastrophic results in mission-critical scenarios. While the notion of trust has been debated across numerous studies and fields, the definition by Lee and See has the benefit of being operationalizable for both research and application purposes: according to them, trust is “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” [15]. Further concerns relevant to human factors in human-automation collaboration scenarios include *cognitive load* [16] and *situation awareness* [17, 18]. The latter can be described as a dynamic process that comprises the ongoing and continuously updated comprehension of the situation at large (*frames*), more

specific details affecting the situation (*implications*), elements of the environment (*objects*), and the *event horizon* of actual and potential developments [18].

In order to evaluate/validate particular designs and techniques, human factors and specifically human-computer interaction researchers and practitioners rely on a multitude of experimental methods that can be traced back to behavioral and social sciences [19]. Such experimental methods can be characterized by the dimensions of precision, generalizability, and realism: for instance, a laboratory experiment can provide a high degree of precision, but it will lead to a trade-off with realism and potentially generalizability. Both quantitative and qualitative methods have their use in human-computer interaction [20], while one of the most common target measures is *usability*. Usability can be defined and evaluated alongside three aspects [21]:

Effectiveness (Are the users able to complete their tasks? How many errors they make?)

Efficiency (How much time does it take for the users to complete their tasks?)

Satisfaction (What is the users' opinion/attitude toward the tool?) [21]

While effectiveness and efficiency are often measured as part of task-based user studies (typically laboratory or perhaps field experiments), satisfaction is usually gauged by interviews and questionnaires with Likert-scale items, with several prominent examples being the 10-item System Usability Scale [22] or the 2-item UMUX-LITE [23]. The knowledge about such methodologies may be beneficial when designing ADR experiments involving human participants or judges; and they are also often relevant to evaluation of visualization techniques and tools, as discussed next.

2.2 Information Visualization and Visual Analytics

The field of information visualization (InfoVis) has emerged in 1990s on the intersection of research and applications in scientific visualization, computer graphics, human-computer interaction (especially graphical user interfaces), cognitive science, data science, knowledge discovery in databases, cartography, graphic design, color science, and further fields. The historical roots of InfoVis can be traced back for several centuries, including examples of thematic and topological maps, mathematical diagrams and statistical charts, as well as infographics handcrafted for printed media with illustratory purposes. However, the identity and special focus of the InfoVis field are strongly associated with a general workflow for visually representing and interacting with abstract data (typically multivariate/multidimensional and often nonspatial in contrast to scientific visualization). The commonly accepted definition of the scope of InfoVis is “The use of computer-supported, interactive, visual representations of abstract data to amplify cognition” [24]. As Stuart Card put it, “The purpose of information visualization is to amplify cognitive performance, not just to create interesting pictures. Information visualizations should do for the mind what automobiles do for the feet” [25].

The overall design of InfoVis techniques and tools typically follows the InfoVis Reference Model [24] or similar frameworks: the tool allows the user to load their data, and then the data is preprocessed and mapped to certain *visual representations* and rendered in one or multiple *views*. The user is usually provided with a number of *interaction* options to investigate the data using the corresponding views, adjust them, or even change the representation. The options and concerns associated with each of the individual steps of the design process according to such a framework, their interactions/relationships, series of such interactions, and the overall context of use constitute the space of research and application challenges for the InfoVis field. For instance, the question of choosing the optimal (or at least more suitable than baseline alternatives) visual representation for a particular data type and scale with further constraints (e.g., the conventions existing for graphical representations and notations in a specific discipline/field such as biology) is a rather typical research problem in InfoVis, addressed through a combination of theoretical and empirical methods.

The complete process of designing, implementing, and validating InfoVis techniques and tools should be carried out in a human-centered way while being based on the existing theoretical models and empirical evidence accumulated in this field. One of the methodologies commonly adopted in InfoVis is the nested blocks model [26] that defines several nested steps for the iterative design and validation process, from domain problem characterization to abstraction to design of individual visual representations and interactions. Another prominent example is the data-users-tasks design triangle model [27] that highlights the respective concerns and their interactions. Finally, Sedlmair et al. describe the design study methodology for InfoVis suitable for applied research and collaborations across and beyond academia [28].

Continuing this discussion in a top-down fashion, the next crucial step of the design process is to define the intended *user tasks* that should drive the resulting design, as high-level tasks of exploratory or confirmatory data analyses would typically require a different design than storytelling/narrative visualization. Considering low-level tasks, the classical InfoVis workflow describing a mix of user tasks and corresponding interactions is known as the Visual Information Seeking Mantra: “Overview first, zoom and filter, then details-on-demand” [29]. Besides these four tasks, Shneiderman’s taxonomy also includes further tasks such as “relate,” “history,” and “extract.” Another prominent taxonomy focuses on ten low-level analysis tasks [30]: “retrieve value,” “filter,” “compute derived value,” “find extremum,” “sort,” “determine range,” “characterize distribution,” “find anomalies,” “cluster,” and “correlate.”

Analysis and design of individual visual representations can be approached from the perspective of combining basic *marks* (graphical elements such as points or lines) with further visual *channels* (graphical properties such as position, color, shape, curvature, orientation/tilt, size, texture, or motion) [31, 32], typically for a given spatial substrate (such as a 2D Cartesian coordinate system). For example, switching line segments into individual dots would turn a line plot into a scatterplot, while varying the area of dots based on the underlying data would turn a scatterplot into a bubble plot. It should be mentioned that InfoVis techniques rely “by default”

on 2D rather than 3D representations when the latter is not justified [32], especially for standard 2D computer monitor screens. The motivation for this includes distortion, occlusion, and navigation issues as well as lack of performance benefits for many data types and tasks, based on prior studies. When properly motivated and used in right contexts (including immersive environments), 3D may have its benefits, of course.

Since the purpose of InfoVis is not only to provide static visual representations of the data (such as diagrams and plots created using spreadsheet or scientific computation software, for instance) but generally also to facilitate further engagement with such representations, the next step of the design process is to consider *interactions* that should be supported for the user to facilitate their tasks. High-level interaction strategies in InfoVis [33] include support for *navigation* through the data space, *direct manipulation* (allowing the user select and adjust data items via their visual representations, rather than requiring database queries or scripting/coding), and *human/agent interaction*, which has become an active area of research and applications over the past years due to the development of conversational agents and natural language interfaces [34, 35]. Considering more low-level interaction techniques, the taxonomy by Shneiderman described above [29] is complemented by the interaction taxonomy by Yi et al.: “select,” “explore,” “reconfigure,” “encode,” “abstract/elaborate,” “filter,” and “connect” [36]. As an example of “abstract/elaborate,” *geometric zooming* allows the user to simply adjust the scaling when exploring representations such as scatter plots, while *semantic zooming* may change the visual representation based on the level of detail, e.g., display additional details and labels when zooming in (this technique is thus also related to the “encode” category). Further, multiple *focus + context* techniques that apply distortions (such as the fisheye lens) provide a more salient representation of the selected data subset (“focus”) while still supporting the overview of further data (“context”) in the same view [37].

While the discussion above was mainly centered on users and tasks, the data-related concerns also play a crucial role in the choice of appropriate visual representations and interactions: “The expressiveness principle dictates that the visual encoding should express all of, and only, the information in the dataset attributes” [32]. In other words, the visual representation should not “invent” additional dimensions or relations (such as typical 3D bar charts with the depth attribute not based on the underlying data) or mislead the users by hinting at such non-existent relations and patterns due to poor visual encoding choices. At the same time, while it is possible to directly map one, two, or even three data attributes/dimensions to the respective 2D/3D spatial substrate, the choice of visual representation for more than three attributes at a time is generally not a trivial task—and in the general case, certain trade-offs must be made that compromise the “... express all of, and only, the information” part of the expressiveness principle. Suggestions for the choice of visual representations of low-dimensional data as well as some “special” data types (such as geospatial data) are provided by Shneiderman as well as Heer et al. [29, 38]. Further, in-depth discussions and online survey browsers for hundreds of existing techniques and tools have also been contributed over the years in the InfoVis field.

For instance, our work on text visualization [39] describes the design space of respective techniques and tools, as depicted in Fig. 2. The respective online survey, the TextVis Browser,² not only provides access and links to the respective publications but also hosts a collection of links to similar resources focusing on the topics such as visualization of trees, multivariate graphs/networks, etc. Another prominent collection to mention is the Visualization Resources Web site³ [40].

The strategies for representation of high-dimensional data [41] can be mapped back to the elements of the InfoVis Reference Model [24], for instance, *dimensionality reduction* [42] could be applied early on to lay out the source data entries as points in a 2D or 3D space, with the assumption that the entries similar in the original high-dimensional data space will be positioned close to each other in the resulting projection space. This strategy could be generally applied for any high-dimensional dataset—however, the quality of the resulting projection as well as the ability of users to actually interpret the resulting point cloud layouts depends on multiple factors (algorithm/model properties, applicability for specific data and tasks, etc.). Alternative strategies include the use of *multiple coordinated views* [43] that usually rely on *brushing + linking* interactions: as the user directly selects items in one view such as a scatterplot, e.g., using a rectangular or lasso selection (the “brushing” step), the other views are dynamically updated to highlight, filter, change pan or zoom level, etc. accordingly (the “linking” step). For example, this strategy could be applied alongside a *scatterplot matrix* as yet another general approach for any high-dimensional dataset. This is a powerful strategy; however, it requires more space to juxtapose several views and potentially imposes the *change blindness* risk for a larger number of views and more complex and larger datasets. Finally, the visual metaphor itself could be designed in a nontrivial way that facilitates one or several tasks with such high-dimensional data, including iconic *glyphs* [44] or representations of a single data entry as a polyline in *parallel coordinates* [45]; the rather popular *radar chart* representation (also known as a *Kiviat diagram*, *star chart*, *spider chart*, etc.) is a variation of the parallel coordinate plot with a radial axis layout, for instance.

Last but not least, the stage of *evaluation/validation* of InfoVis approaches is crucial for ensuring that novel ideas or combined applications of existing techniques reach the intended goals. InfoVis relies on the body of knowledge of evaluation in human-computer interaction described above, ranging from formal methods to usability studies to expert reviews [46]. Isenberg et al. provide a systematic review of the evaluation methods used in InfoVis studies [47], while the works by Lam et al. and Elmquist and Yi provide guidelines for designing and conducting evaluations according to typical scenarios and patterns [48, 49].

While this description provides an overview of the field of InfoVis, the reader may have noticed that the focus of the discussion so far has been on the careful mapping from user requirements and data characteristics to the design and evaluation of

²<https://textvis.lnu.se>; last accessed on June 1, 2025.

³<https://sites.google.com/view/visres/>; last accessed on June 1, 2025.

interactive visualization techniques and tools with only cursory mentions of *computational* methods. Indeed, while the traditional InfoVis approaches relied on a rather direct mapping process from data (traditionally, from databases or static files) to visual representations, eventually the challenges related to processing larger, more complex, and heterogeneous datasets for broader and more demanding tasks gave way to the concept of *visual analytics*. Keim et al. provide the following definition for this field: “Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets” [50]. While the boundary between visual analytic and InfoVis fields and methods is often fuzzy, the following Visual Analytics Mantra by Keim et al. illustrates the respective focus: “Analyze First, Show the Important, Zoom, Filter and Analyze Further, Details on Demand” [50]. The need to accommodate analysts in the world with more complex and less predictable data and models was discussed by Amar and Stasko around the same time as a follow-up to their earlier work [30, 51], with the focus on supporting high-level analytic activities such as “complex decision-making, especially under uncertainty,” “learning a domain,” “identifying the nature of trends,” and “predicting the future.” The later Knowledge Generation Model by Sacha et al. describes the relationships across elements and activities that take place within the software (data, computational model, interactive visualization affordances) and beyond (the human analyst’s steps for exploring the data, formulating and verifying hypotheses, and eventually generating new knowledge about the domain problem) [52]. Visual analytics has become an important field over the past two decades, and its role especially in the context of ADR-related topics and concerns cannot be overstated. The recent discussion of grand challenges in visual analytic applications, for instance, included closer intertwining of visual analytics and AI, providing guidance (which requires mixed-initiative approaches relying on AI), and explainability in complex visual analytic systems [53], which opens up opportunities for mutual benefits for research and applications involving visual analytics and ADR.

Turning to industrial and societal uptake of InfoVis, novel tool prototypes as well as commercial products have been developed within or close to the visualization research community [54], such as Tableau, complemented by a variety of *business intelligence and analytics* tools that integrate some level of support for visual representation and interaction techniques with data preprocessing and computational models [55], such as Microsoft Power BI or IBM Cognos. Design, implementation, and evaluation of such tools aimed at professionals with a varying degree of expertise in visualization and computational disciplines pose a set of additional challenges. The existing work on this genre of *visualization dashboards* includes lessons learned [56], design spaces [57], and actionable design patterns [58] from visualization experts.

3 Human-Centered Perspectives on ADR

In this section, we discuss three topics based on the existing literature as well as our own work that highlight different constellations of human-computer interaction, visualization, and ADR concerns.

3.1 *Human-Computer Interaction and Visualization for Explainable AI*

One of the directions of research and applications in human-computer interaction, InfoVis, and visual analytics that is highly relevant to ADR is support for various stages of AI (especially ML/DL) model training and deployment pipeline. Some of the prominent examples of the work in visual analytics focusing on ML (rather than later DL) approaches include interactive support for construction, debugging, and analysis of application of decision trees [59], support vector machine classifiers for text retrieval [60], inter-active ensemble learning for video analytics [61], or visual interactive labeling [62].

Our own prior work on supporting the active learning-based multi-label data annotation and classifier training processes for research on sentiment and stance in text data led to designing and applying a similar approach with our collaborators, researchers in linguistics and computational linguistics [63]. ALVA is a visual analytic tool that provides user interfaces for annotation, visual analyses of the annotated data and annotation process itself, as well as monitoring of the classification model performance over the active learning training rounds. The Web-based tool supports several user roles with permissions to access and make annotations, control the active learning process, or access the visual analytic interfaces. The focus of the latter is on analyzing the progress of the annotation process as well as identifying patterns in the data annotated so far, as displayed in Fig. 3. By using a novel visual representation titled *CatCombos* (“category combinations”), our collaborators were able to identify interesting cases and patterns of interaction between several semantic categories of stance in the annotated text data, such as expressions of uncertainty, hypotheticals, and prediction. Similar approaches are sought after by researchers and practitioners in computational linguistics, as understanding and supporting the data annotation/labeling process contributes in a major way to the success of resulting models. Considerations for interdisciplinary collaborations in visual text analytics [64] can thus be extended for other areas of ADR in the future.

The rapid development of AI approaches across academia and industry over the past decade has also affected the fields of human-computer interaction and visualization, with new challenges and research directions emerging to support the research and applications of such approaches. The rising interest for *interpretable* (with the model internals understandable to humans) and *explainable* (with the model behavior and individual decisions being described and motivated for humans)

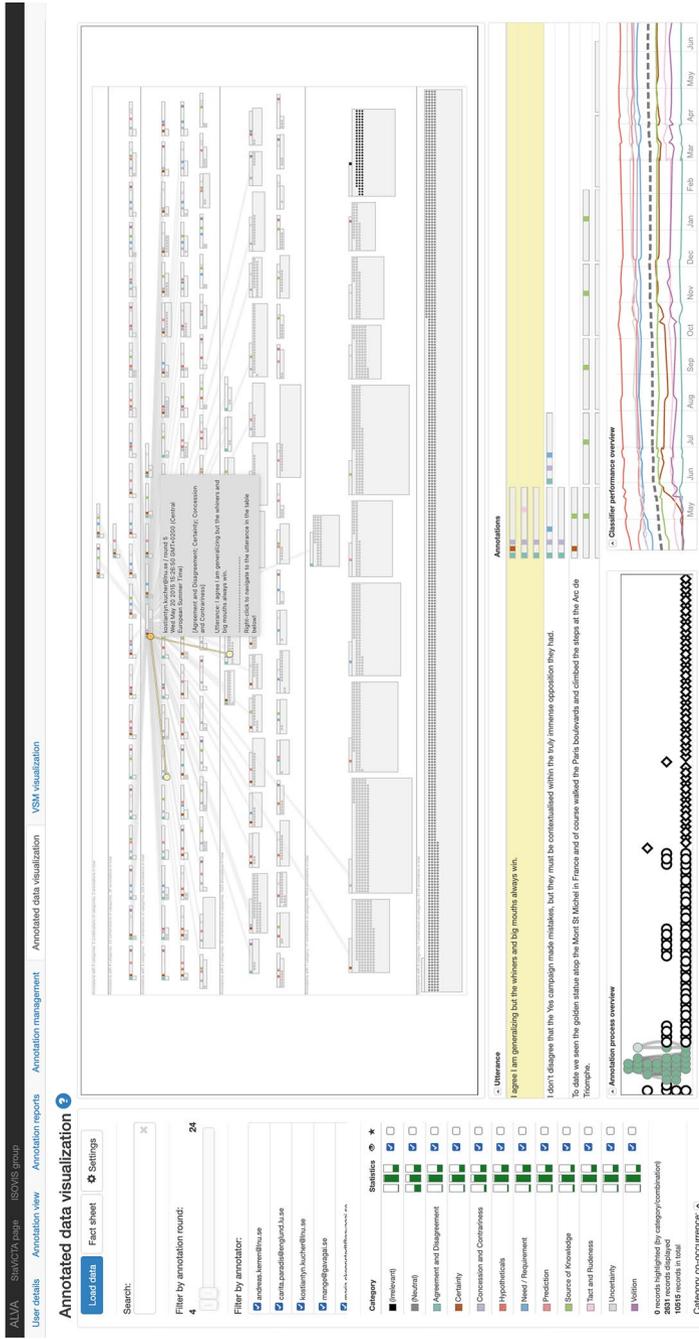


Fig. 3 The user interface of ALVA used for the visual analysis of annotated data, annotation process, and active learning classifier training for multi-label text classification

AI [65] has led to the initiatives on human-computer interaction research agenda for explainable, accountable, and intelligent systems [66], as well as the human-centered artificial intelligence agenda [67]. Amershi et al. proposed a set of 18 generally applicable guidelines for human-AI interaction applicable not only for research prototypes but also industrial products, divided into several main categories according to the training and use stages/situations: initial, ongoing, AI error/failure, and long-term use [68]. Liao et al. contributed an explainable AI question bank collected via discussions with industry design practitioners on the users' needs for explanations and explainability [69]. These guidelines and resources are relevant for ongoing and future work in ADR, as they elaborate on the constraints and preferences for AI models and tools that are used and/or overseen by humans.

The work within the InfoVis and especially visual analytic fields relevant to interpretable and explainable AI has resulted in a number of interactive visual approaches to support human-centered machine learning [70, 71], explainable deep learning [72], and trust in AI [73]. Several prominent examples to mention here include the *What-If Tool* for interactive model probing and counterfactual reasoning [74], *AttentionViz* for a global view on Transformer attention patterns [75], and the *explAiner* framework for generating visual explanations and suggestions for model steering [76].

Our own work in this area has resulted in design spaces and interactive survey browsers focusing on the topics of enhancing the trust in ML models with the use of visualization⁴ [77, 78] and the use of data embedding approaches in visual analytics⁵ [79]. These resources (see Fig. 4) position the existing work and allow researchers and practitioners alike to find the existing visual analytic techniques supporting tasks such as ML model construction, debugging, quality/bias control, interactive exploration and comparison of the embedding space, as well as representation and explanation of model decisions. Furthermore, we contributed visual analytic techniques and tools such as *StackGenVis* for training stacking ML classifiers [80] and *EVO* for text similarity analyses using ensembles of embeddings [81].

There is potential for further advances in the area by combining human-computer interaction, visual analytics, and ADR involving the challenges of interpretability, explainability, and trustworthiness of AI—not only within academia but also with further societal and industrial collaborations. This is highlighted by our recent conceptual framework on this topic [82] as well as a study on explainable AI interfaces for operators in process industries [83].

⁴<https://trustmlvis.lnu.se/>; last accessed on June 1, 2025.

⁵<https://va-embeddings-browser.ivis.itn.liu.se/>; last accessed on June 1, 2025.



Fig. 4 Further online survey browsers relevant to ADR topics. Left: TrustMLVis Browser, currently hosting a dataset with 542 categorized visualization techniques aimed at enhancing trust in ML models. Right: VA + Embeddings Browser, currently hosting a dataset with 123 categorized visualization techniques relevant to the use of data embedding methods at various stages of the visual analytic pipeline

3.2 *Joint Human-AI Control*

Some of the “ironies of automation” described above [13] remained open over the course of decades, resulting in the *automation conundrum* associated with the change of *level of automation* (LOA) from completely manual human control toward full automation [17]: “as more autonomy is added to a system, and its reliability and robustness increase, the lower the situation awareness of human operators and the less likely that they will be able to take over manual control when needed.” When considering AI instead of traditional automation techniques, new ironies emerge: according to Endsley, for instance, “The more intelligent and adaptive the AI, the less able people are to understand the system” and “The more capable the AI, the poorer people’s self-adaptive behaviours for compensating for shortcomings” [14]. Requirements for explainability, human interaction and oversight, training and skill retention, and joint testing of human-AI systems are mentioned among alleviations for these challenges [14].

Relying on the notions of abstraction hierarchy [10], situation awareness [18], and LOA [17], our own work proposed the *Joint Control Framework (JCF)* [84] that describes human-machine (including automation and AI) interaction scenarios with a focus on *cognitive joints*, moments in time and space where human and machine work together. JCF relies on three key components:

1. Process mapping that results in definition and categorization of relevant subjects, objects, and effects for the scenario at hand.
2. The framework of *Levels of Autonomy in Cognitive Control (LACC)* that defines six levels (1–2: *how?*, 3–4: *what?*, 5–6: *why?*) of autonomy and interaction in relation to situational awareness concepts such as frames and the event horizon.
3. *JCF Score*, a notation for temporal description resembling the musical notation used for sheet music. This score uses six horizontal lines to represent six LACC levels. Separate groups of lines can be used for separate subjects or objects, similarly to separate staves used for different instruments in sheet music.

While the framework and score can be applied on their own, the ongoing work on a prototype of *JCF Editor* depicted in Fig. 5 facilitates transcription and analyses of complex use case scenarios recorded with the existing human-interaction tools and users. In this example, future drone traffic is analyzed from the perspective of an operator who needs to inspect planned traffic before approving it [85]. Looking at Fig. 5, we immediately can see that the blue decision joints of the human subject are much higher up than their green and orange action and decision joints. This means that the subject must make inferences in abstraction, i.e., consequences of implemented plans on higher goals—and then back again from those higher-level goals to what adjustments on lower levels that must be made to fulfil the goals.

The relevance of this line of work to ADR lies in the ability to formalize and analyze complex scenarios including multiple subjects and objects over time, such as one or several operators and one or several automation/AI agents. The methodology can be applied for evaluation of existing techniques and planning of future

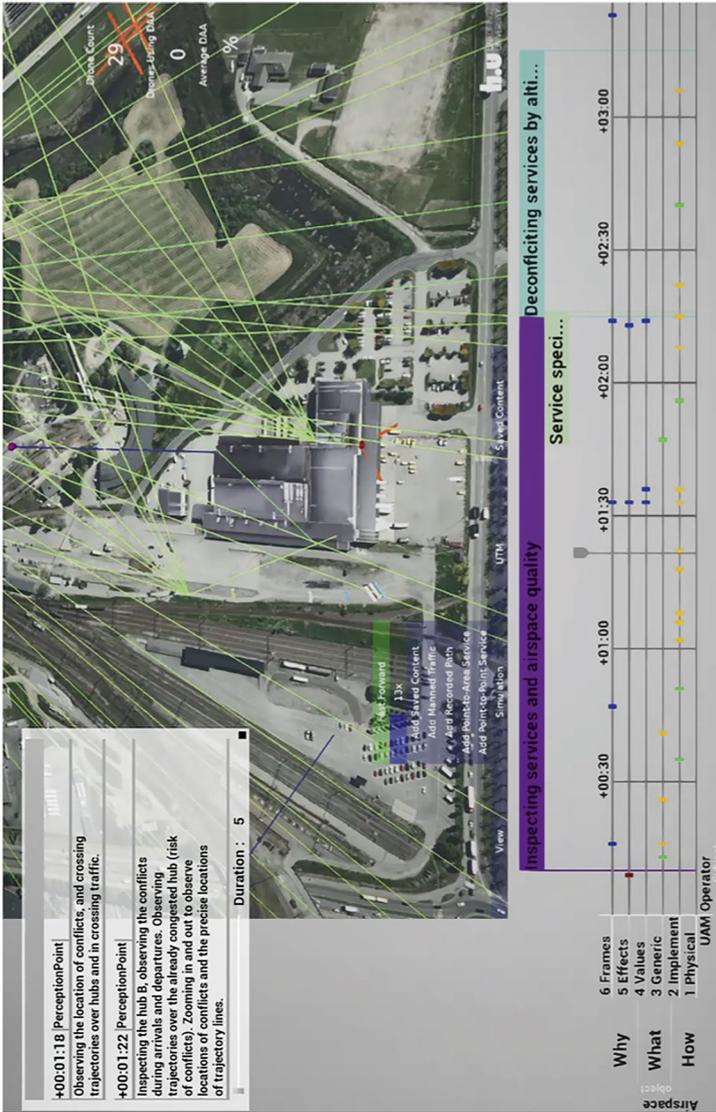


Fig. 5 The user interface of JCF Editor used for annotating and analyzing the data from a study on unmanned aerial vehicle traffic coordination [85]. Blue markers represent decision joints between human operators and the traffic process, green markers represent actions, and orange markers represent perception joints. The markers represent what the joint is about, e.g., is it about a plan (L3) or a goal (L5), and is this goal or plan explicitly represented so that it can be perceived and/or manipulated? Level differences between perception-decision-action mean that the subject does the interpretation themselves, e.g., cognitive or computational load

studies. A relevant example here involves the work on human-in-the-loop AI for future, advanced air traffic management systems combining manned and unmanned vehicles and services [86]. To accommodate the long-term planning and design for such eventual future systems (in the air traffic control domain and beyond), Nylin et al. described the concept of Reduced Autonomy Workspace that combines the notions of LOA and LACC to describe the extent of divided and collaborative work between subjects, including automation/AI agents [87]. The term “reduced autonomy” refers to the cases when the automation/AI agent needs to hand over some degree of control to the human operator (e.g., due to uncertainty or accountability requirements), and this handover should be adapted to accommodate the operator (e.g., checking whether the operator is highly occupied at the moment). Nylin et al. also introduced the design of glyphs for similar real-time scenarios in future systems that visually encodes time remaining to make the necessary decision [88]. The development of actual models, agents, and autonomous systems within the scope of ADR that would be involved in such traffic management or unmanned traffic scenarios provides clear opportunities for future contributions and collaborations across fields.

3.3 *Evaluation for Human-AI Decision Support Systems*

Our final topic is relevant to the context of the ongoing *AI4REALNET* project⁶ that aims to achieve human-in-the-loop decision-making for co-learning between AI and humans. The project focuses on application scenarios with critical network infrastructure systems (electricity, railway, and air traffic management) that are traditionally operated by humans. The project is tasked with not only purely technical (such as novel reinforcement learning methods for complex networks) but also sociotechnical goals to enhance transparency, safety, explainability, and acceptance of the developed AI solutions by human stakeholders (in particular, industrial operators). More specifically, the human-centered evaluations planned within the project are intended to assess the social-technical decision quality; AI acceptability, trust, and trustworthiness; human user experience; AI and human learning curves; task allocation balance; and long-term consequences of AI assistants.

As discussed above, the concerns related to human-centered AI and ML [67, 68, 71] are relevant for such scenarios in ADR. More specifically, Hoffman et al. describe the 7-item Explanation Goodness Checklist (to be used by other researchers/experts), the corresponding 7-item Explanation Satisfaction Scale with Likert items (to be used by study participants), and the 8-item XAI Trust Scale with Likert items (also to be used by study participants) [89]. Similar questionnaires are also used with respect to further (X)AI properties in the published work, and they are certainly useful; however, the work by Buçinca et al. indicated that proxy tasks and

⁶<https://ai4realnet.eu/>; last accessed on June 1, 2025.

subjective measurements may be misleading in evaluating explainable systems; thus, experimental design for such tasks can require careful planning [90]. The work by Nauta et al. defined 12 properties for explanations such as *correctness*, *completeness*, or *compactness* and provided first steps toward systematic evaluation of these properties by combining computational methods (e.g., randomly perturbing the internals of the model and checking whether the generated explanation would change) as well as user studies [91], which may also be relevant for this project.

It is also interesting to note within the context of ADR, especially robotics, that some of the prior work focusing on human-robot interaction discusses measurements and evaluation methods applicable for human-AI interaction involving software AI agents. Nikolaidis et al. evaluate human-robot *mutual adaptation*, including direct measurements of human participants' self-confidence in their ability to complete the task as well as their assessment of trustworthiness of the robot (and several further measurements in separate experiments) [92]. Hoffman discusses both subjective (based on self-reported questionnaire results) and objective metrics of *fluency* in human-robot interaction; the latter include the percentage of concurrent activity, the human's idle time, the robot's functional delay, and the robot's idle time [93]. These measures are interesting to consider for other human-AI interaction scenarios; however, for various tasks involving monitoring of ongoing activities, measures such as idle time may not be relevant. Perception of robots and other AI agents by humans is also the subject of the Artificial-Social-Agent Questionnaire [94]. The resulting short version of the questionnaire comprises 24 Likert-scale items that range from the appearance of the agent to its usability, social presence, and perceived user-agent alliance. Finally, Hauptman et al. discuss an evaluation of perceptions of *adaptive* autonomous agents for human-AI teaming [95]. In their study, AI teammates could be either considered manually set to a particular LOA or considered able to adapt their LOA dynamically. The results of the factorial survey indicate that predictability and degree of team experience can enable higher levels of autonomy in such scenarios. These studies as well as further work related to sociotechnical evaluation of human-AI systems will be relevant to the evaluations within the scope of the *AI4REALNET* project, and we hope that the outcomes of the project (including experimental protocols) will nurture future work in ADR as well as interdisciplinary collaborations.

4 Conclusions and Opportunities

In this chapter, we have outlined the basic concepts as well as several scenarios and prior results in human-AI interaction and visualization that are relevant to ADR. We have described methods and techniques that originate from these fields and would be directly relevant and helpful for ADR researchers and practitioners. *We conclude that combining these would have broad impact into several areas of concern such as developing user interfaces and visual analytic solutions to facilitate ADR development, debugging, and monitoring and analyzing the existing and potential*

scenarios of multi-agent interactions. We especially see promise in addressing the involvement of human stakeholders by approaching the evaluation of ADR solutions as sociotechnical problems, with human-centered concerns being as important as model performance.

Besides these direct applications of human-centered methods, we would also like to mention the following challenges and directions where collaborations between researchers and practitioners in ADR and human-centered fields have the potential to make impact:

4.1 Reaching Out to Audiences in Critical Application Fields

As discussed in this chapter, operators in highly specialized fields such as air traffic control or process industries are trained to follow protocols that emphasize safety and predictability. Introducing black-box ADR methods directly at the level of full automation for such users and fields may cause short-term and especially long-term issues including loss of situational awareness, under-/over-trust, and deskilling. The study with decision-makers from the industry by Bedué and Fritzsche shows that AI adoption intention among such respondents is affected by several dimensions of trust (including *access to knowledge*, *transparency*, *explainability*, and *reliability*) as well as perceived benefits and risks [96]. Peres et al. highlight the requirements for keeping human in the loop using human-machine technology, achieving data availability and quality, and ensuring interpretability and trust for AI as major challenges for industrial AI research and adoption, especially considering higher LOA [97]. Making use of human-centered methods to introduce ADR techniques and models may thus be the best way to find common ground with such stakeholders and users.

4.2 Reaching Out to Wider Audiences

Interestingly, not only audiences in highly specialized domains but also the general public can benefit from collaborations across human-centered methods and ADR, considering the ubiquity and impact of ADR technologies on the society and everyday life. Sinderman et al. describe the Attitude Towards Artificial Intelligence Scale based on five Likert-scale items with questions such as “I fear AI” and “AI will benefit humankind” [98]. Human-centered methods can thus be helpful for introducing and educating the general public on various aspects of ADR, which would allow them to form informed opinions.

4.3 Supporting Data-Centric and Human-Centric AI Paradigms

One of the recent developments directly relevant to ADR is the emergence of the *data-centric* AI paradigm that shifts the focus from algorithm/model performance toward the role of data in the AI life cycle; as Gröger put it, “There is no AI without data” [99]. Sambasivan et al. discuss the risks rising from cascades of negative, downstream effects from data issues and highlight the role of data as a first-class citizen in AI [100]. Jarrahi et al. describe the principles for data-centric AI, which explicitly mention “Human-Centeredness of ‘Data Work,’” “AI as a Sociotechnical System,” and “Continuous and Substantive Interactions Between AI and Domain Experts,” which align with our recommendations for deeper collaboration between researchers in human-centered fields and ADR [101].

4.4 Preparing Appropriate Responses to ADR Cybersecurity and Privacy Challenges

Peres et al. highlight the role of cybersecurity and privacy for industrial AI [97]; however, we would argue that ADR applications in further critical application fields must make use of all feasible strategies to ensure the highest degree of cybersecurity. West and Aydin discuss how the process of aligning AI models with user preferences and values can be manipulated by malicious actors [102]. Machine learning models used for dimensionality reduction to represent data can also be attacked and manipulated in some scenarios, as discussed in our recent work [103]. There is thus room for extensive research and innovation on the intersection of ADR, human-centered methods, and cybersecurity as well as privacy.

4.5 Contributing Novel ADR-Driven Interfaces and Interactions

Finally, besides applying human-centered methods for ADR, ADR methods could also be highly beneficial for human-computer interaction and visualization research and applications. The existing work on automation in visualization [104], guidance [105], adaptive visualization [106], and further opportunities of integrating ML and AI in visualization [107–109] provides us with clear evidence for the interest existing for such collaborations among our fields.

Acknowledgments The work on this chapter was carried out within the scope of the AI4REALNET project. AI4REALNET has received funding from European Union’s Horizon Europe Research and Innovation Programme under Grant Agreement No. 101119527 and from the

Swiss State Secretariat for Education, Research and Innovation (SERI). Views and opinions expressed are however those of the authors 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.

References

1. Enqvist, L. (2023). ‘Human oversight’ in the EU artificial intelligence act: What, when and by whom? *Law, Innovation and Technology*, 15(2), 508–535.
2. Curry, E., Heintz, F., Irgens, M., Smeulders, A. W., & Stramigioli, S. (2022). Partnership on AI, data, and robotics. *Communications of the ACM*, 65(4), 54–55.
3. Karwowski, W. (2005). Ergonomics and human factors: The paradigms for science, engineering, design, technology and management of human-compatible systems. *Ergonomics*, 48(5), 436–463.
4. Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., Wilson, J. R., & Van der Doelen, B. (2012). A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*, 55(4), 377–395.
5. Militello, L. G., Dominguez, C. O., Lintern, G., & Klein, G. (2010). The role of cognitive systems engineering in the systems engineering design process. *Systems Engineering*, 13(3), 261–273.
6. Hoffman, R. R., Feltovich, P. J., Ford, K. M., Woods, D. D., Klein, G., & Feltovich, A. (2002). A rose by any other name... would probably be given an acronym. *IEEE Intelligent Systems*, 17(4), 72–80.
7. Ebert, A., Gershon, N. D., & van der Veer, G. C. (2012). Human-computer interaction: Introduction and overview. *KI-Künstliche Intelligenz*, 26, 121–126.
8. Dix, A. (2017). Human-computer interaction, foundations and new paradigms. *Journal of Visual Languages and Computing*, 42, 122–134.
9. Bennett, K. B., & Flach, J. (2019). Ecological interface design: Thirty-plus years of refinement, progress, and potential. *Human Factors*, 61(4), 513–525.
10. Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors*, 44(1), 62–78.
11. Zohrevandi, E., Westin, C. A., Lundberg, J., & Ynnerman, A. (2022). Design and evaluation study of visual analytics decision support tools in air traffic control. *Computer Graphics Forum*, 41(1), 230–242.
12. Zohrevandi, E., Westin, C. A., Vrotsou, K., & Lundberg, J. (2022). Exploring effects of ecological visual analytics interfaces on experts’ and novices’ decision-making processes: A case study in air traffic control. *Computer Graphics Forum*, 41(3), 453–464.
13. Bainbridge, L. (1982). Ironies of automation. *IFAC Proceedings Volumes*, 15(6), 129–135.
14. Endsley, M. R. (2023). Ironies of artificial intelligence. *Ergonomics*, 66(11), 1656–1668.
15. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
16. Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, 55(13s), 283.
17. Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors*, 59(1), 5–27.
18. Lundberg, J. (2015). Situation awareness systems, states and processes: A holistic framework. *Theoretical Issues in Ergonomics Science*, 16(5), 447–473.
19. McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In *Readings in human-computer interaction* (pp. 152–169). Morgan Kaufmann.

20. Purchase, H. C. (2012). *Experimental human-computer interaction: A practical guide with visual examples*. Cambridge University Press.
21. Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 345–352). ACM.
22. Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29.
23. Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: When there's no time for the SUS. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 2099–2102). ACM.
24. Card, S. K., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.
25. Card, S. (2007). Information visualization. In *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (2nd ed., pp. 509–543). CRC Press.
26. Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 921–928.
27. Miksch, S., & Aigner, W. (2014). A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38, 286–290.
28. Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2431–2440.
29. Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages* (pp. 336–343). IEEE.
30. Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *Proceedings of the IEEE Symposium on Information Visualization* (pp. 111–117). IEEE.
31. Görg, C., Pohl, M., Qeli, E., & Xu, K. (2007). Visual representations. In *Human-Centered Visualization Environments: GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5–8, 2006, Revised Lectures* (pp. 163–230). Springer.
32. Munzner, T. (2014). *Visualization analysis and design*. CRC Press.
33. Fikkert, W., D'Ambros, M., Bierz, T., & Jankun-Kelly, T. J. (2007). Interacting with visualizations. In *Human-Centered Visualization Environments: GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5–8, 2006, Revised Lectures* (pp. 77–162). Springer.
34. Hoque, E., & Islam, M. S. (2025). Natural language generation for visualizations: State of the art, challenges and future directions. *Computer Graphics Forum*, 44(1), e15266.
35. Voigt, H., Alaçam, Ö., Meuschke, M., Lawonn, K., & Zarrieß, S. (2022). The why and the how: A survey on natural language interaction in visualization. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 348–374). ACL.
36. Yi, J. S., ah Kang, Y., Stasko, J., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1224–1231.
37. Kosara, R., Hauser, H., & Gresh, D. L. (2003). An interaction view on information visualization. In *EuroGraphics'03—State of the art reports*. Eurographics.
38. Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59–67.
39. Kucher, K., & Kerren, A. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the IEEE Pacific Visualization Symposium* (pp. 117–121). IEEE.
40. Liu, X., Alharbi, M. S., Chen, J., Diehl, A., Rees, D., Firat, E. E., Wang, Q., & Laramée, R. S. (2023). Visualization resources: A survey. *Information Visualization*, 22(1), 3–30.

41. Liu, S., Maljovec, D., Wang, B., Bremer, P. T., & Pascucci, V. (2016). Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3), 1249–1268.
42. Espadoto, M., Martins, R. M., Kerren, A., Hirata, N. S., & Telea, A. C. (2019). Toward a quantitative survey of dimension reduction techniques. *IEEE Transactions on Visualization and Computer Graphics*, 27(3), 2153–2173.
43. Roberts, J. C. (2007). State of the art: Coordinated & multiple views in exploratory visualization. In *Proceedings of the International Conference on Coordinated and Multiple Views in Exploratory Visualization* (pp. 61–71). IEEE.
44. Borgo, R., Kehrer, J., Chung, D. H., Maguire, E., Laramée, R. S., Hauser, H., Ward, M., & Chen, M. (2013). Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *EuroGraphics'13—State of the art reports* (pp. 39–63). Eurographics.
45. Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer*, 1, 69–91.
46. Carpendale, S. (2008). Evaluating information visualizations. In *Information visualization: Human-centered issues and perspectives* (pp. 19–45). Springer.
47. Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., & Möller, T. (2013). A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2818–2827.
48. Elmqvist, N., & Yi, J. S. (2015). Patterns for visualization evaluation. *Information Visualization*, 14(3), 250–269.
49. Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9), 1520–1536.
50. Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information visualization: Human-centered issues and perspectives* (pp. 154–175). Springer.
51. Amar, R. A., & Stasko, J. T. (2005). Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4), 432–442.
52. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., & Keim, D. A. (2014). Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1604–1613.
53. Wu, A., Deng, D., Chen, M., Liu, S., Keim, D., Maciejewski, R., Miksch, S., Strobelt, H., Viégas, F., & Wattenberg, M. (2023). Grand challenges in visual analytics applications. *IEEE Computer Graphics and Applications*, 43(5), 83–90.
54. Satyanarayan, A., Lee, B., Ren, D., Heer, J., Stasko, J., Thompson, J., Brehmer, M., & Liu, Z. (2019). Critical reflections on visualization authoring systems. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 461–471.
55. Behrisch, M., Streeb, D., Stoffel, F., Seebacher, D., Matejek, B., Weber, S. H., Mittelstädt, S., Pfister, H., & Keim, D. (2018). Commercial visual analytics systems—Advances in the big data analytics field. *IEEE Transactions on Visualization and Computer Graphics*, 25(10), 3011–3031.
56. Froese, M. E., & Tory, M. (2016). Lessons learned from designing visualization dashboards. *IEEE Computer Graphics and Applications*, 36(2), 83–89.
57. Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2018). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 682–692.
58. Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2022). Dashboard design patterns. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 342–352.
59. van den Elzen, S., & van Wijk, J. J. (2011). BaobabView: Interactive construction and analysis of decision trees. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology* (pp. 151–160). IEEE.

60. Heimerl, F., Koch, S., Bosch, H., & Ertl, T. (2012). Visual classifier training for text document retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2839–2848.
61. Höferlin, B., Netzel, R., Höferlin, M., Weiskopf, D., & Heidemann, G. (2012). Interactive learning of ad-hoc classifiers for video visual analytics. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology* (pp. 23–32). IEEE.
62. Bernard, J., Zeppelzauer, M., Sedlmair, M., & Aigner, W. (2018). VIAL: A unified process for visual interactive labeling. *The Visual Computer*, 34, 1189–1207.
63. Kucher, K., Paradis, C., Sahlgren, M., & Kerren, A. (2017). Active learning and visual analytics for stance classification with ALVA. *ACM Transactions on Interactive Intelligent Systems*, 7(3), 14.
64. Kucher, K., Sultanum, N., Daza, A., Simaki, V., Skeppstedt, M., Plank, B., Fekete, J. D., & Mahyar, N. (2022). An interdisciplinary perspective on evaluation and experimental design for visual text analytics: Position paper. In *Proceedings of the IEEE Workshop on Evaluation and Beyond—Methodological Approaches for Visualization* (pp. 28–37). IEEE.
65. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics* (pp. 80–89). IEEE.
66. Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM.
67. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504.
68. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., & Teevan, J. (2019). Guidelines for human-AI interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM.
69. Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM.
70. Andrienko, N., Andrienko, G., Adilova, L., & Wrobel, S. (2022). Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications*, 42(1), 123–133.
71. Sperrle, F., El-Assady, M., Guo, G., Borgo, R., Chau, D. H., Endert, A., & Keim, D. (2021). A survey of human-centered evaluations in human-centered machine learning. *Computer Graphics Forum*, 40(3), 543–568.
72. La Rosa, B., Blasilli, G., Bourqui, R., Auber, D., Santucci, G., Capobianco, R., Bertini, E., Giot, R., & Angelini, M. (2023). State of the art of visual analytics for explainable deep learning. *Computer Graphics Forum*, 42(1), 319–355.
73. Beauxis-Aussalet, E., Behrisch, M., Borgo, R., Chau, D. H., Collins, C., Ebert, D., El-Assady, M., Endert, A., Keim, D. A., Kohlhammer, J., & Oelke, D. (2021). The role of interactive visualization in fostering trust in AI. *IEEE Computer Graphics and Applications*, 41(6), 7–12.
74. Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., & Wilson, J. (2020). The What-If Tool: Interactive probing of machine learning models. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 56–65.
75. Yeh, C., Chen, Y., Wu, A., Chen, C., Viégas, F., & Wattenberg, M. (2024). AttentionViz: A global view of transformer attention. *IEEE Transactions on Visualization and Computer Graphics*, 30(1), 262–272.
76. Spinner, T., Schlegel, U., Schäfer, H., & El-Assady, M. (2020). explAiner: A visual analytics framework for interactive and explainable machine learning. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 1064–1074.
77. Chatzimparmpas, A., Kucher, K., & Kerren, A. (2024). Visualization for trust in machine learning revisited: The state of the field in 2023. *IEEE Computer Graphics and Applications*, 44(3), 99–113.
78. Chatzimparmpas, A., Martins, R. M., Jusufi, I., Kucher, K., Rossi, F., & Kerren, A. (2020). The state of the art in enhancing trust in machine learning models with the use of visualizations. *Computer Graphics Forum*, 39(3), 713–756.

79. Huang, Z., Witschard, D., Kucher, K., & Kerren, A. (2023). VA + Embeddings STAR: A state-of-the-art report on the use of embeddings in visual analytics. *Computer Graphics Forum*, 42(3), 539–571.
80. Chatzimparmpas, A., Martins, R. M., Kucher, K., & Kerren, A. (2020). StackGenVis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1547–1557.
81. Witschard, D., Jusufi, I., Martins, R. M., Kucher, K., & Kerren, A. (2022). Interactive optimization of embedding-based text similarity calculations. *Information Visualization*, 21(4), 335–353.
82. Kucher, K., Zohrevandi, E., & Westin, C. A. (2025). Towards visual analytics for explainable AI in industrial applications. *Analytix*, 4(1), 7.
83. Zhang, Y., Methnani, L., Brorsson, E., Zohrevandi, E., Darnell, A., & Kucher, K. (2025). Designing explainable and counterfactual-based AI interfaces for operators in process industries. In *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (GRAPP, HUCAPP and IVAPP)* (Vol. 1, pp. 831–841). SciTePress.
84. Lundberg, J., & Johansson, B. J. (2021). A framework for describing interaction between human operators and autonomous, automated, and manual control systems. *Cognition, Technology & Work*, 23(3), 381–401.
85. Lundberg, J., Arvola, M., & Palmerius, K. L. (2021). Human autonomy in future drone traffic: Joint human–AI control in temporal cognitive work. *Frontiers in Artificial Intelligence*, 4, article 704082.
86. Lundberg, J., Bång, M., Johansson, J., Cheaitou, A., Josefsson, B., & Tahboub, Z. (2019). Human-in-the-loop AI: Requirements on future (unified) air traffic management systems. In *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference*. IEEE.
87. Nylin, M., Johansson Westberg, J., & Lundberg, J. (2022). Reduced autonomy workspace (RAW)—An interaction design approach for human-automation cooperation. *Cognition, Technology & Work*, 24(2), 261–273.
88. Nylin, M., Lundberg, J., Bång, M., & Kucher, K. (2024). Glyph design for communication initiation in real-time human-automation collaboration. *Visual Informatics*, 8(4), 23–35.
89. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, 5, 1096257.
90. Buçinca, Z., Lin, P., Gajos, K. Z., & Glassman, E. L. (2020). Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proceedings of the ACM International Conference on Intelligent User Interfaces* (pp. 454–464). ACM.
91. Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., Van Keulen, M., & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. *ACM Computing Surveys*, 55(13s), 295.
92. Nikolaidis, S., Hsu, D., & Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *International Journal of Robotics Research*, 36(5–7), 618–634.
93. Hoffman, G. (2019). Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3), 209–218.
94. Fitrianie, S., Bruijnes, M., Li, F., Abdulrahman, A., & Brinkman, W. P. (2022). The artificial-social-agent questionnaire: Establishing the long and short questionnaire versions. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents*. ACM.
95. Hauptman, A. I., Schelble, B. G., McNeese, N. J., & Madathil, K. C. (2023). Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. *Computers in Human Behavior*, 138, 107451.
96. Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, 35(2), 530–549.

97. Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial artificial intelligence in Industry 4.0—Systematic review, challenges and outlook. *IEEE Access*, 8, 220121–220139.
98. Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI-Künstliche Intelligenz*, 35(1), 109–118.
99. Gröger, C. (2021). There is no AI without data. *Communications of the ACM*, 64(11), 98–108.
100. Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L.M. (2021). “Everyone wants to do the model work, not the data work”: Data cascades in high-stakes AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM.
101. Jarrahi, M. H., Memariani, A., & Guha, S. (2023). The principles of data-centric AI. *Communications of the ACM*, 66(8), 84–92.
102. West, R., & Aydin, R. (2025). The AI alignment paradox. *Communications of the ACM*, 68(3), 24–26.
103. Fujiwara, T., Kucher, K., Wang, J., Martins, R. M., Kerren, A., & Ynnerman, A. (2025). Adversarial attacks on machine learning-aided visualizations. *Journal of Visualization*, 28(1), 133–151.
104. Domova, V., & Vrotsou, K. (2023). A model for types and levels of automation in visual analytics: A survey, a taxonomy, and examples. *IEEE Transactions on Visualization and Computer Graphics*, 29(8), 3550–3568.
105. Collins, C., Andrienko, N., Schreck, T., Yang, J., Choo, J., Engelke, U., Jena, A., & Dwyer, T. (2018). Guidance in the human–machine analytics process. *Visual Informatics*, 2(3), 166–180.
106. Yanez, F., Conati, C., Ottley, A., & Nobre, C. (2025). The state of the art in user-adaptive visualizations. *Computer Graphics Forum*, 44(1), e15271.
107. Endert, A., Ribarsky, W., Turkay, C., Wong, B. W., Nabney, I., Blanco, I. D., & Rossi, F. (2017). The state of the art in integrating machine learning into visual analytics. *Computer Graphics Forum*, 36(8), 458–486.
108. Wang, Q., Chen, Z., Wang, Y., & Qu, H. (2022). A survey on ML4VIS: Applying machine learning advances to data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 5134–5153.
109. Wu, A., Wang, Y., Shu, X., Moritz, D., Cui, W., Zhang, H., Zhang, D., & Qu, H. (2022). AI4VIS: Survey on artificial intelligence approaches for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 5049–5070.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

