KI - Künstliche Intelligenz (2022) 36:287–296
https://doi.org/10.1007/s13218-022-00780-8

PROJECT REPORTS

XAINES: Explaining AI with Narratives

Mareike Hartmann[1,2] · Han Du[3] · Nils Feldhus[4] · Ivana Kruijff-Korbayová[5] · Daniel Sonntag[1,6]

Received: 13 June 2022 / Accepted: 31 October 2022 / Published online: 21 November 2022

Abstract

Artificial Intelligence (AI) systems are increasingly pervasive: Internet of Things, in-car intelligent devices, robots, and virtual assistants, and their large-scale adoption makes it necessary to explain their behaviour, for example to their users who are impacted by their decisions, or to their developers who need to ensure their functionality. This requires, on the one hand, to obtain an accurate representation of the chain of events that caused the system to behave in a certain way (e.g., to make a specific decision). On the other hand, this causal chain needs to be communicated to the users depending on their needs and expectations. In this phase of explanation delivery, allowing interaction between user and model has the potential to improve both model quality and user experience. The XAINES project investigates the explanation of AI systems through narratives targeted to the needs of a specific audience, focusing on two important aspects that are crucial for enabling successful explanation: generating and selecting appropriate explanation content, i.e. the information to be contained in the explanation, and delivering this information to the user in an appropriate way. In this article, we present the project's roadmap towards enabling the explanation of AI with narratives.

Keywords Explainable AI · Interactive machine learning · Human–machine interaction · Conversational explanations

1 Introduction

AI systems have huge potential to improve our lives, especially when deployed in high stake scenarios such as healthcare applications or automated driving, where erroneous decisions can have severe consequences [65, 106]. Their impact on human lives comes hand in hand with our need to understand *why* a system behaves in a certain way, to verify that it works as intended, and to estimate the extent to which its decisions can be trusted. In order to enable the use of AI systems in real-world applications, we need to find appropriate ways for explaining their behaviour [29, 43, 97]. How to do that depends on the audience consuming the model explanations [5, 14, 24, 80]. For example, *Machine Learning (ML) developers* usually want to test and improve the system, and explanations provide a way of identifying model shortcomings to be fixed [48, 77]. For *domain experts*, such as medical staff or engineers using the system for domain-specific applications, explanations serve to improve the cooperation between the domain expert and the machine, e.g.,

✉ Mareike Hartmann
mareike.hartmann@dfki.de

Han Du
han.du@dfki.de

Nils Feldhus
nils.feldhus@dfki.de

Ivana Kruijff-Korbayová
Ivana.Kruijff-Korbayova@dfki.de

Daniel Sonntag
daniel.sonntag@dfki.de

[1] Department of Interactive Machine Learning, German Research Center for Artificial Intelligence, Saarbrücken, Germany

[2] Department of Language Science and Technology, Saarland University, Saarbrücken, Germany

[3] Department of Agents and Simulated Reality, German Research Center for Artificial Intelligence, Saarbrücken, Germany

[4] Department of Speech and Language Technology, German Research Center for Artificial Intelligence, Berlin, Germany

[5] Department of Multilinguality and Language Technology, German Research Center for Artificial Intelligence, Saarbrücken, Germany

[6] Applied Artificial Intelligence (AAI), Oldenburg University, Oldenburg, Germany

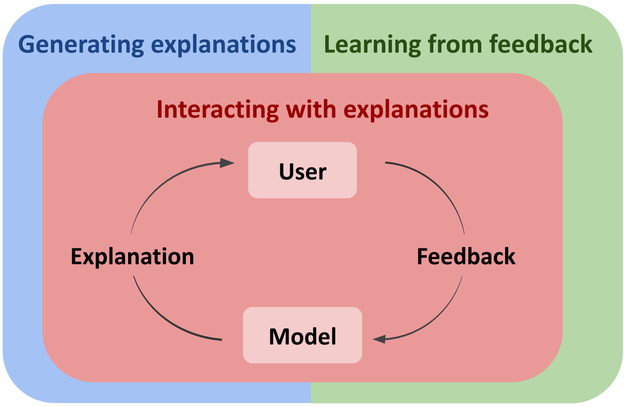

Fig. 1 Interaction with explanations (middle part) plays a central role for explaining AI systems, which requires the generation of model explanations (left part) and the integration of user feedback (right part)

by providing a way of evaluating the reliability of a model's decision, thereby increasing user trust in the system. In addition, domain experts might want to use explanations in order to learn from the AI by extracting knowledge that the AI acquired from large amounts of training data [80, 82].[1]

For explanation delivery, interaction between user and machine based on explanations is a central component (see Fig. 1 for an overview of the different tasks that need to be addressed in order to enable explanation in an interactive loop), where the model provides an explanation to the user, and the user provides feedback to the model based on the explanation [45, 93, 99]. For ML developers, providing feedback to the model allows to efficiently fix deficiencies that were identified based on model explanations [49]. For domain experts, the interaction with model explanations benefits the user and the way they use the system: the ability to provide feedback to the model increases user satisfaction [2, 86, 93] and their trust in the system [31]. Finally, the social sciences point out that explanations themselves should be embedded in interactive communication between the model as explainer and the user as explainee [61, 62].

The work presented here is part of the XAINES project[2], that aims at explaining AI systems through *narratives*. A narrative is a form of discourse conveying information about an event by giving an account of meaningfully connected events. In the context of explaining AI, explaining with narratives means to explain an event by recounting the events that caused it [66]. Communicating an explanation in the form of a narrative also addresses the fact that an event is usually affected by a set of causes that should be part of the explanation, rather than one factor in isolation [39].[3] As narratives are an elementary form of human expression [6], we hypothesize that they are an appropriate means to communicate explanations, in particular to users without ML background.

RQ1: **How to generate explanations?**
§ 2.1 … for image content via image captioning
§ 2.2 … for actions by aligning videos with descriptions

RQ2: **How to use explanations for improving models in an interactive machine learning (IML) setup?**
§ 3.1 Updating models with explanatory user feedback

RQ3: **How to model explanation delivery as interactive dialogue between user and machine?**
§ 3.2.1 Interactive exploration of NLP models

Fig. 2 Relation between the works presented in this article and the project's research questions. The grey squares indicate the respective section numbers

In the following, we present a summary of our accomplished, on-going and planned work on explanation generation (Sect. 2) and the interaction with explanations (Sect. 3), and outline how it contributes to approaching our ultimate goal of creating explainable AI. These works are separate contributions addressing different research questions which need to be answered in order to enable explainable AI. Figure 2 provides an overview over this article's structure and how the presented works relate to the project's research questions.

2 Generating Explanations

Users request model explanations for different reasons and with different motivations in mind [28, 80, 82]. The XAINES project addresses these different user needs by distinguishing two types of explanations (see Fig. 3): *ML narratives* convey the causal chain leading to a model prediction,

[1] We mainly focus on explanations for ML developers and domain experts, rather than laypeople, as the project involves domain-specific application scenarios, such as medical decision support, where an AI would support a domain expert. Some parts of the project address explanations that might be targeted to laypeople, e.g. the interactive exploration of NLP models (Sect. 3.2.1).

[2] https://www.dfki.de/en/web/research/projects-and-publications/projects-overview/projekt/xaines/.

[3] Jacovi et al. [39] illustrate the set of causes affecting an event using an example of a self-driving car that crashed into a wall: The narrative explaining this event is, that the car was driving at 50 km/h instead of the allowed 20km/h, because it misidentified a speed sign due to debris covering its camera. A bump in the road caused the speeding car to go off the road and crash into the wall. Here, several causes (debris on camera, misidentification of speed sign, bump in the road) affected the event.

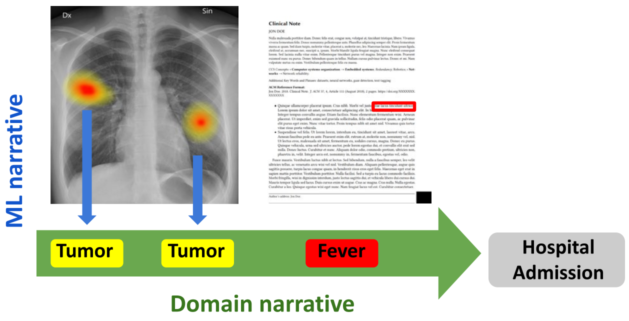

Fig. 3 Examples of ML and domain narratives for a medical decision support system

and can primarily be used to test and improve the model. For example, saliency maps as ML explanations [84] can reveal that a model picks up on irrelevant features to classify X-ray images [18]. *Domain narratives* describe sequences of domain-specific events that led to a specific outcome, and can for example be used by domain experts to assess if a model decision is justified.[4] This latter type of explanations should be model invariant and accessible to consumers without any knowledge about ML [8]. We explore the generation of both types of explanations in the context of processing visual content, focusing on two use-cases: (1) providing explanations for systems that process images, focusing on applications in the medical domain; (2) providing explanations for systems that process video content.

2.1 Linking Images with Language

AI systems developed for usage in the medical domain often involve image processing components, e.g. for speech-based image annotation [88] or medical decision support [73], where relevant information has to be extracted from domain-specific data in various forms, such as X-ray images and health records [87]. In order to describe relevant information in an image or sequences thereof, we focus on the tasks of image captioning [41, 59, 104] and visual story telling [38]. Image captions have previously been explored as a means to explain decisions of image classifiers [35, 51] and Visual Question Answering (VQA) models [49], whereas in our work we investigate their use as domain narratives. In particular, we focus on the challenges of selecting the most relevant information from the images, and of addressing the needs of the respective target audience by generating personalized image descriptions. In [10], we propose an image captioning model that conditions generation on selected visual information to model the fact that humans restrict their explanation of an event to a subset of selected causal connections [61]. In [9], we investigate the use of transfer learning and machine translation for generating image captions in German. Due to a lack of non-English image captioning resources, such cross-lingual transfer is necessary in order to make natural language domain narratives accessible to non-English speakers.

2.1.1 Image Captions as Explanations

Natural language is often pointed out as the most intuitive way of communicating an explanation, especially to non-ML experts [23, 46], hence image captions as natural language descriptions of image content appear to be an obvious choice for domain narratives. Moreover, recent progress in pre-training large multi-modal encoders on large multi-modal datasets [12, 40, 76] has pushed the state-of-the-art for image captioning [37, 64]. However, Rohrbach et al. [81] raised awareness for the phenomenon of *object hallucinations*, i.e. the description of objects that are not actually visible in an image. Such errors can potentially be very harmful when explaining image content in high-stake domains. One of the underlying research questions we aim to answer in the project is if image descriptions are suitable as domain narratives, and how their interplay with ML explanations impacts the explanation process. To give ML explanations for the generated descriptions, we will explore the use of saliency methods such as Grad-CAM [83], which we had previously used for explaining classifier decisions in the context of skin cancer recognition [64, 67], for highlighting image regions that affected the generation of specific words in the description. Whereas we so far focused on the technical challenges of generating the explanations, a next step will be to evaluate the quality and adequateness of image captions as domain explanations in a user study.

2.2 Linking Action with Language

Similar to how we use generated natural language sequences to describe image content, we can use natural language to describe actions performed by an embodied AI, for example a robot or an AI-driven digital human. AI-driven digital humans have widely been used in industry simulation, remote education, healthcare, and entertainment. For many applications, it is important to understand the intention of AI-driven characters [72, 78]. For example, in the digital simulation of autonomous driving, the autonomous car needs to understand the behaviour of simulated pedestrians, for example whether the pedestrian is going to cross the street or not. In addition, some sequences of actions require domain knowledge, for instance when skilled workers perform manual assembly tasks in workshops, and it is useful to include domain knowledge into motion generation models

[4] Biran and McKeown [8] refer to this type of information as *justification*.

[13, 57]. It requires experts' knowledge to explain why the actions should be executed in certain orders. We hypothesize that when providing domain narratives, users can better understand and interact with generated motion.

2.2.1 Alleviating the Data Bottleneck

For activity recognition, existing methods [7, 26, 108] usually require labeled 3D motion data as ground truth for model training. However, annotating 3D motion capture data with narrative explanations is cost-intensive and time-consuming, even more so for domain-specific activities such as martial arts or dancing, where experts' knowledge is required. One promising way to tackle this challenge is to use existing collections of video data [53, 68, 79]. There are huge amounts of videos available online that contain well-explained activities as subtitles in the timeline. For example, on the video sharing platform *Youtube*, people can learn physical skills with instructional visual movements and narrative textual explanation. In the XAINES project, our goal is to alleviate the issue of limited availability of labeled 3D data by leveraging existing video data with narrative explanation. This way, domain-specific knowledge can be integrated into the motion generation model, which can synthesize target motion with narrative explanations.

In order to model 3D movement with textual explanation, we first apply state-of-the-art 3D motion estimation approaches [30, 95] to reconstruct 3D movements from the 2D videos. The textual annotation is then automatically aligned with the estimated 3D motion based on video time stamps. To include the rich variations of natural human movement, we apply deep generative model Variational Autoencoder (VAE) [44] to capture the statistical property of human movement [22]. In our work, the 3D motion data and textual annotation are jointly modeled together. Given high-level targets, our motion synthesis framework can create the required motion from the textual explanation. For motion recognition, the synthetic motion generated from our model can serve as ground truth to improve model training.

2.2.2 Multi-modal Embeddings for Motion and Text

A common approach to model inputs from both modalities is learning joint embeddings for the multi-modal data. In [27], we propose a joint embedding model to learn the mapping between 3D motion and narrative description. Two autoencoders are deployed to learn the representation of 3D motion and natural language separately. For motion data, we use a hierarchical pose model to address the kinematic structure of the human model. For textual input, we apply the BERT model [19] which is pre-trained on large text corpora to create contextualized embeddings. Both inputs are then combined in a joint embedding space for pose and language. Given a textual description, our model can produce the corresponding motion using the hierarchical pose decoder. Theoretically, our model can also be used for generating a narrative explanation given the 3D motion.

Our model in [27] is trained on the KIT Motion-Language Dataset [71], which contains 3D pose data with human-annotated sentences. However, the type of actions in this dataset is limited and the language annotation is quite simple. In XAINES, we plan to test our approach for complex martial arts actions such as Tai Chi or Capoeira with more detailed textual descriptions. Our model will automatically generate descriptive explanations to describe the motions in multiplayer games. The goal of each player can be derived from descriptive explanations. We also plan to investigate the performance of our approach on video data compared to 3D motion capture data. Our ultimate goal is to animate semantic-aware, high-fidelity AI-driven characters that can interact with users, while being explainable via textual descriptions.

3 Interacting with Explanations

Our work presented above focuses on the generation of explanations for different AI-related components. Once the explanations are generated, the selected information needs to be communicated to the user. In this step of explanation delivery, we focus on making use of interaction between user and machine: First, we investigate how explanations can be delivered in an explanation-feedback loop, that aims at improving the model based on human feedback, and allows personalization of explanations. Second, we explore how to move beyond a one-way broadcast of explanation content by modelling explanation as a conversational interaction between user and machine.

3.1 Explanation-Based Feedback Loop

In this part of the project, we explore the interaction with explanations of classifier decisions in the Interactive Machine Learning (IML) framework, which serves to improve ML models based on feedback gained from interaction with users. On the one hand, Explainable AI (XAI) is often considered a prerequisite for enabling meaningful interaction between user and machine, allowing the user to provide useful feedback based on which the model can be improved [31, 94, 99]. On the other hand, IML might be a necessary component of optimal XAI systems, as users provided with model explanations desire to provide feedback in order to adjust the model [86]. Hence, we hypothesize that investigating the application of IML approaches in an XAI context and vice versa can serve the goals of both paradigms. Building on related work exploring the explanation-feedback loop [45, 89, 98], we will address the open questions of the best mechanism for integrating feedback into the model [1], the type of feedback that is most helpful for model improvement, and how to best evaluate the framework, either in terms of model accuracy, or in terms of user-centric metrics. In [33], we provide a survey on improving Natural Language Processing (NLP) models with different types of human explanations. We consider human explanations as a promising type of human feedback, as models can be trained more efficiently with human explanations compared to label-level feedback. The two most prominent types of human explanations used to improve NLP models are *highlight* explanations, i.e. subsets of input elements that are deemed relevant for a prediction, and *free-text* explanations [103], i.e. natural language statements answering the question why an instance was assigned a specific label. We plan to focus our future efforts on learning from feedback in the form of natural language explanations, as users generally perceive natural language as preferred way of interacting with models, and natural language explanations are less constrained and can consequently have a higher information content than highlight explanations. In addition to enabling IML through XAI, we ask how IML methods can be used for best rendering domain narratives. Along with providing a means for general model improvement, the interaction between user and model can be exploited to adapt explanations, e.g. as personalized image descriptions that take into account the user's active vocabulary [15] or other features such as their preferred sentence length or level of detail. Our experiments in [10] show promising initial results for caption personalization using interactive re-ranking of decoder output, which we plan to explore further in the future. In [32], we outline an approach for using text- and image-based data augmentation to efficiently adapt image captioning models to new data based on user feedback. We plan to gain first insights on the effectiveness of these approaches based on simulated feedback, and to then consolidate findings in an interactive user study.

3.2 Conversational Interaction as Narrative Explanation of AI

Human explanations are interactive and incremental, allowing participants to challenge, query, negotiate, discuss and clarify the explanation content, ideally until mutual understanding and agreement is achieved [56]. In this part of the project, we aim at modelling this important aspect of explanation as a goal-oriented dialog between the user and the machine, where the goal is to achieve mutual understanding with respect to the explanation [46, 61, 80]. We envision the dialog system to be adaptive with respect to the user, as the amount of detail of the explanatory dialogue should be conditioned on their abilities and expectations [61]. Oversimplified explanations that lead to unjustified trust must be avoided [28], therefore one challenge is to find a trade-off between persuasive and descriptive explanation strategies [35]. Other challenges include how to best present the narrative, e.g. by splitting it into multiple installments [16], and how to adapt user representations over time. We hypothesize that such questions can best be answered observing human conversational interaction, ideally in explanatory dialogue. To this end, we are currently in the progress of collecting resources that contain such interactive explanations between humans. So far, we identified three data types that we expect to contain such explanatory dialogue: datasets for information-seeking dialogue [54, 69, 74], datasets for teacher-student interactions [20, 90], and video tutorials [60]. Our planned next steps are to analyse to which extent explanatory dialogue is present in these datasets and if they constitute a suitable resource for our purpose.

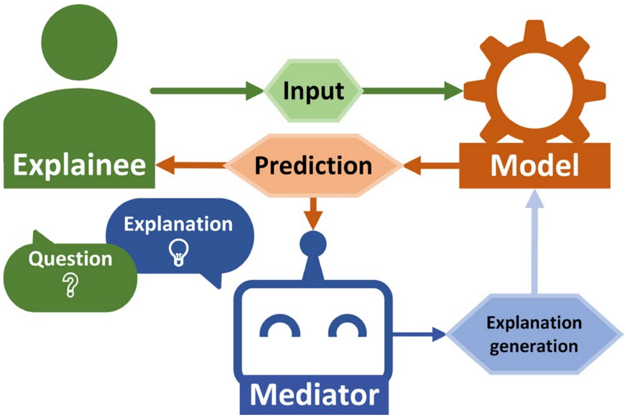

Fig. 4 Simplified concept of a Mediator [25] explaining the predictions of a Model to the human Explainee. Step 1: The Explainee provides input to the Model. Step 2: The Model outputs a prediction based on the input. Step 3: The Mediator generates candidate explanations based on the prediction and grey-box access to the Model. Step 4: The Mediator starts off the explanation dialogue with the Explainee. Step 5: The Explainee acts upon the explanation and asks follow-up questions until satisfied. Meanwhile, the Mediator keeps track of the dialogue state and the user's mental model

The proposed dialog system should also be able to recognize user intent, by matching a user query with an appropriate explanation method [50, 100, 102]. A query like *Which parts of the input contributed most to model output?* matches with an explanation method highlighting the salient parts of the input, e.g. based on input gradients [96]. In contrast, a query like *What (general) patterns in the (training) data are responsible for an output?* matches with an explanation resulting from a probing task [17]. For matching intent to explanation, we plan to explore standard intent classification [52, 107] and textual similarity models [34, 105]. We established the aforementioned desiderata for text-based conversational agents explaining the behavior of NLP models as *Mediators* in [25] depicted in Fig. 4.

We are planning to investigate the above mentioned research questions associated with the implementation of an AI system explainable via conversational explanations within the use case of an interactive NLP model explorer, with a proof of concept in text classification and language modeling tasks, which we will describe in the following.[5]

3.2.1 Interactive NLP Model Exploration

Many types of explanation-generating methods can be employed to diagnose the behaviour of NLP models [39, 55]. Our work builds on top of applications allowing to explore language models interactively [11, 47, 70, 91, 92, 98]. Our goal is to provide users with easy access to a better understanding of NLP model behaviour via conversational agents that can draw from a pool of explanations in a task- and model-agnostic manner. This means such an agent is trained to handle NLP models of different sizes and training objectives. Although the task and model chosen by the user might be transparent, the pitfall that the agent has to circumvent is the one of generalization: For example, in feature attribution for sentiment analysis, words that are salient for one task and model might not be in a different context. The agent has to be able to abstract away these biases. At the same time, the agent as well as the underlying NLP model receive rich feedback from the dialog history [94] that can be utilized for improvement and better alignment with the user. The modality of natural language lends itself to very comprehensive explanations involving counterfactuals and insights about training data and dynamics that are not easily understood by people outside the NLP domain. Contrary to the previously described works in XAINES, our conversational interactions present the narrative bit by bit, i.e. with each turn of the dialog, and with the simplest parts first, so users are not overwhelmed and are animated to ask follow-up questions. This enables human studies with laypeople. The two most pressing issues we identified are the lack of explanation dialog datasets [4, 101, 103] and evaluation standards [3, 58, 103]. Both of them require a solid foundation through human evaluation: When constructing datasets, human annotators should be tasked to judge generated explanations according to their preferences and additionally edit them to make them more natural and aligned with human expectations [103]. For evaluation, participants in user studies should be capable of simulating the underlying model [21] after the narrative has been presented and a mutual understanding has been reached. We also hope to close the gap of applying explainability methods to NLP problems beyond text classification, such as summarization, machine translation and open-domain question answering. Our proposed framework will require us to come up with solutions.

4 Outlook

In this project description, we presented several parts of the on-going XAINES project and how they connect in order to explain AI with narratives. The project's runtime is scheduled until August 2024, and we want to conclude our contribution with a brief summary and an outlook on planned future work. For explanation generation, we focus on visual content: images and predictions of image classifiers, and (synthesized) motion in video data. So far, we completed work on image caption generation and synthesizing motion from textual descriptions, which can serve as integral components for implementing explainable AI for concrete use-cases, which we see in the medical domain and the development of automated driving. We are currently in the process of creating a resource of dancing and martial arts videos annotated with textual descriptions, which can be used for training both text-to-motion and motion-to-text models. For communicating explanations, we focus on interaction between user and machine: First, we exploit interaction in the IML framework, where we aim to improve explainable models based on user feedback. This feedback can take many forms, and currently we focus on learning from feedback in the form of an explanation from user to machine, which has the potential to improve both the model and the model's explanations. Planned next steps are to develop methods for learning classifiers from natural language explanations for tabular and multi-modal data as supported by the CLUES [75] and e-ViL [42] datasets. Second, we want to model the process of explaining as a conversational interaction between human and machine. Feldhus et al. [25] introduce a blueprint of such a system, and a next step towards enabling conversational explanations will be to implement the system conceptualized there. While there exists an open-source implementation for dialogue-based explanations based on tabular classification datasets by Slack et al. [85], the transfer to more challenging applications requires the collection of task-specific datasets, which is another planned step in our future research outline.

Acknowledgements The XAINES project is funded by the German Federal Ministry of Education and Research [grant number 01IW20005].

Funding Open Access funding enabled and organized by Projekt DEAL.

[5] Note that this framework does not represent a demonstrator for all of XAINES, but is rather conceived for its own NLP use cases such as sentiment analysis. However, we are also exploring synergies with other parts of the project.

References

1. Active learning in image captioning. https://iml.dfki.de/active-learning-in-image-captioning/
2. Amershi S, Cakmak M, Knox WB, Kulesza T (2014) Power to the people: the role of humans in interactive machine learning. AI Mag 35(4):105–120. https://doi.org/10.1609/aimag.v35i4.2513
3. Arora S, Pruthi D, Sadeh N, Cohen WW, Lipton ZC, Neubig G (2021) Explain, edit, and understand: Rethinking user study design for evaluating model explanations. https://doi.org/10.48550/ARXIV.2112.09669.
4. Attari N, Heckmann M, Schlangen D (2019) From explainability to explanation: using a dialogue setting to elicit annotations with justifications. In: Proceedings of the 20th annual SIGdial meeting on discourse and dialogue. Association for Computational Linguistics, Stockholm, Sweden, p 331–335. https://doi.org/10.18653/v1/W19-5938.
5. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012
6. Barthes R, Duisit L (1975) An introduction to the structural analysis of narrative. New Literary Hist 6(2):237–272. https://doi.org/10.2307/468419
7. Berretti S, Daoudi M, Turaga P, Basu A (2018) Representation, analysis, and recognition of 3D humans: a survey. ACM Trans Multimedia Comput Commun Appl. https://doi.org/10.1145/3182179
8. Biran O, McKeown K (2014) Justification narratives for individual classifications. In: Proceedings of the AutoML workshop at ICML, vol 2014, p 1–7
9. Biswas R, Barz M, Hartmann M, Sonntag D (2021) Improving German image captions using machine translation and transfer learning. In: International conference on statistical language and speech processing. Springer, p 3–14. https://doi.org/10.1007/978-3-030-89579-2_1
10. Biswas R, Barz M, Sonntag D (2020) Towards explanatory interactive image captioning using top-down and bottom-up features, beam search and re-ranking. KI-Künstliche Intelligenz 34(4):571–584. https://doi.org/10.1007/s13218-020-00679-2
11. Bove C, Aigrain J, Lesot MJ, Tijus C, Detyniecki M (2022) Contextualization and exploration of local feature importance explanations to improve understanding and satisfaction of non-expert users. In: 27th International conference on intelligent user interfaces, IUI '22. Association for Computing Machinery, New York, NY, USA, p 807–819. https://doi.org/10.1145/3490099.3511139
12. Bugliarello E, Cotterell R, Okazaki N, Elliott D (2021) Multimodal pretraining unmasked: a meta-analysis and a unified framework of vision-and-language BERTs. Trans Assoc Comput Linguist 9:978–994. https://doi.org/10.1162/tacl_a_00408
13. Busemann S, Steffen J, Herrmann E (2016) Interactive planning of manual assembly operations: from language to motion. Procedia CIRP 41:224–229. https://doi.org/10.1016/j.procir.2015.12.106. Research and innovation in manufacturing: key enabling technologies for the factories of the future—proceedings of the 48th CIRP conference on manufacturing systems
14. Calegari R, Ciatto G, Dellaluce J, Omicini A (2019) Interpretable narrative explanation for ML predictors with LP: a case study for XAI. In: WOA, p 105–112
15. Chunseong Park C, Kim B, Kim G (2017) Attend to you: personalized image captioning with context sequence memory networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 895–903
16. Clark HH (1996) Using language. Cambridge University Press, Cambridge
17. Conneau A, Kruszewski G, Lample G, Barrault L, Baroni M (2018) What you can cram into a single $&!#* vector: probing sentence embeddings for linguistic properties. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers). Association for Computational Linguistics, Melbourne, Australia, p 2126–2136. https://doi.org/10.18653/v1/P18-1198
18. DeGrave AJ, Janizek JD, Lee SI (2021) Ai for radiographic covid-19 detection selects shortcuts over signal. Nat Mach Intell. https://doi.org/10.1038/s42256-021-00338-7
19. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, p 4171–4186. https://doi.org/10.18653/v1/N19-1423.
20. Di Eugenio B, Fossati D, Ohlsson S, Cosejo D (2009) Towards explaining effective tutorial dialogues. Cognitive Science - COGSCI
21. Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning . https://doi.org/10.48550/ARXIV.1702.08608.
22. Du H, Herrmann E, Sprenger J, Fischer K, Slusallek P (2019) Stylistic locomotion modeling and synthesis using variational generative models. In: Motion, interaction and games, MIG '19. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3359566.3360083
23. Ehsan U, Harrison B, Chan L, Riedl MO (2018) Rationalization: a neural machine translation approach to generating natural language explanations. In: Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society, AIES '18. Association for Computing Machinery, New York, NY, USA, p 81–87. https://doi.org/10.1145/3278721.3278736
24. Ehsan U, Passi S, Liao QV, Chan L, Lee IH, Muller M, Riedl MO (2021) The who in explainable AI: how AI background shapes perceptions of AI explanations. https://doi.org/10.48550/ARXIV.2107.13509
25. Feldhus N, Ravichandran AM, Möller S (2022) Mediators: conversational agents explaining NLP model behavior. In: IJCAI 2022 workshop on explainable artificial intelligence (XAI)
26. Fortes Rey V, Garewal KK, Lukowicz P (2021) Translating videos into synthetic training data for wearable sensor-based activity recognition systems using residual deep convolutional networks. Appl Sci 11(7):3094. https://doi.org/10.3390/app11073094

27. Ghosh A, Cheema N, Oguz C, Theobalt C, Slusallek P (2021) Synthesis of compositional animations from textual descriptions. In: Proceedings of the IEEE/CVF international conference on computer vision, p 1396–1406
28. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L (2018) Explaining explanations: an overview of interpretability of machine learning. In: 2018 IEEE 5th international conference on data science and advanced analytics (DSAA), p 80–89. IEEE . https://doi.org/10.1109/DSAA.2018.00018
29. Goebel R, Chander A, Holzinger K, Lecue F, Akata Z, Stumpf S, Kieseberg P. Holzinger A (2018) Explainable AI: the new 42? In: International cross-domain conference for machine learning and knowledge extraction. Springer, p 295–303. https://doi.org/10.1007/978-3-319-99740-7_21
30. Gong K, Li B, Zhang J, Wang T, Huang J, Mi MB, Feng J, Wang X (2022) Posetriplet: co-evolving 3D human pose estimation, imitation, and hallucination under self-supervision. https://doi.org/10.48550/ARXIV.2203.15625.
31. de Graaf MMA, Malle BF (2017) How people explain action (and autonomous intelligent systems should too). In: 2017 AAAI fall symposia, Arlington, Virginia, USA, November 9-11, 2017. AAAI Press, p 19–26. https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16009
32. Hartmann M, Anagnostopoulou A, Sonntag D (2022) Interactive machine learning for image captioning. https://doi.org/10.48550/ARXIV.2202.13623
33. Hartmann M, Sonntag D (2022) A survey on improving NLP models with human explanations. In: Proceedings of the first workshop on learning with natural language supervision. Association for Computational Linguistics, Dublin, Ireland, p 40–47. https://doi.org/10.18653/v1/2022.lnls-1.5
34. He P, Liu X, Chen W, Gao J (2019) A hybrid neural network model for commonsense reasoning. In: Proceedings of the first workshop on commonsense inference in natural language processing. Association for Computational Linguistics, Hong Kong, China, p 13–21. https://doi.org/10.18653/v1/D19-6002
35. Hendricks LA, Akata Z, Rohrbach M, Donahue J, Schiele B, Darrell T (2016) Generating visual explanations. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016. Springer International Publishing, Cham, p 3–19. https://doi.org/10.1007/978-3-319-46493-0_1
36. Herman B (2017) The promise and peril of human evaluation for model interpretability. https://doi.org/10.48550/ARXIV.1711.07414
37. Hu X, Gan Z, Wang J, Yang Z, Liu Z, Lu Y, Wang L (2022) Scaling up vision-language pre-training for image captioning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), p 17980–17989
38. Huang THK, Ferraro F, Mostafazadeh N, Misra I, Agrawal A, Devlin J, Girshick R, He X, Kohli P, Batra D, Zitnick CL, Parikh D, Vanderwende L, Galley M, Mitchell M (2016) Visual storytelling. In: Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, San Diego, California, p 1233–1239. https://doi.org/10.18653/v1/N16-1147
39. Jacovi A, Bastings J, Gehrmann S, Goldberg Y, Filippova K (2022) Diagnosing AI explanation methods with folk concepts of behavior. https://doi.org/10.48550/ARXIV.2201.11239
40. Jia C, Yang Y, Xia Y, Chen YT, Parekh Z, Pham H, Le Q, Sung YH, Li Z, Duerig T (2021) Scaling up visual and vision-language representation learning with noisy text supervision. In: International conference on machine learning, p 4904–4916. PMLR
41. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 3128–3137 . https://doi.org/10.1109/CVPR.2015.7298932
42. Kayser M, Camburu OM, Salewski L, Emde C, Do V, Akata Z, Lukasiewicz T (2021) e-ViL: a dataset and benchmark for natural language explanations in vision-language tasks. In: Proceedings of the IEEE/CVF international conference on computer vision, p 1244–1254
43. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D (2019) Key challenges for delivering clinical impact with artificial intelligence. BMC Med 17(1):1–9. https://doi.org/10.1186/s12916-019-1426-2
44. Kingma DP, Welling M (2014) Auto-encoding variational bayes. In: 2nd International conference on learning representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, conference track proceedings
45. Kulesza T, Burnett M, Wong WK, Stumpf S (2015) Principles of explanatory debugging to personalize interactive machine learning. In: Proceedings of the 20th International Conference on Intelligent User Interfaces, IUI ’15. Association for Computing Machinery, New York, NY, USA, p 126–137. https://doi.org/10.1145/2678025.2701399
46. Lakkaraju H, Slack D, Chen Y, Tan C, Singh S (2022) Rethinking explainability as a dialogue: a practitioner’s perspective . https://doi.org/10.48550/ARXIV.2202.01875
47. Lee M, Liang P, Yang Q (2022) Coauthor: designing a human–AI collaborative writing dataset for exploring language model capabilities. In: CHI conference on human factors in computing systems, CHI ’22. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3491102.3502030
48. Lertvittayakumjorn P, Toni F (2021) Explanation-based human debugging of NLP models: a survey. Trans Assoc Comput Linguist 9:1508–1528. https://doi.org/10.1162/tacl_a_00440
49. Li Q, Fu J, Yu D, Mei T, Luo J (2018) Tell-and-answer: Towards explainable visual question answering using attributes and captions. In: Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, p 1338–1346. https://doi.org/10.18653/v1/D18-1164
50. Liao QV, Gruen D, Miller S (2020) Questioning the AI: informing design practices for explainable AI user experiences. Association for Computing Machinery, New York, NY, USA, p 1–15. https://doi.org/10.1145/3313831.3376590
51. Lipton ZC (2018) The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 16(3):31–57. https://doi.org/10.1145/3236386.3241340
52. Liu C, Xu P, Sarikaya R (2015) Deep contextual language understanding in spoken dialogue systems. In: Sixteenth annual conference of the international speech communication association
53. Liu W, Mei T (2022) Recent advances of monocular 2D and 3D human pose estimation: a deep learning perspective. ACM Comput Surv. https://doi.org/10.1145/3524497 (**accepted**)
54. Lowe R, Pow N, Serban I, Pineau J (2015) The Ubuntu dialogue corpus: a large dataset for research in unstructured multi-turn dialogue systems. In: Proceedings of the 16th annual meeting of the special interest group on discourse and dialogue. Association for Computational Linguistics, Prague, Czech Republic, p 285–294. https://doi.org/10.18653/v1/W15-4640
55. Madsen A, Reddy S, Chandar APS (2021) Post-hoc interpretability for neural NLP: a survey. ArXiv **abs/2108.04840** . https://doi.org/10.48550/arXiv.2108.04840
56. Madumal P, Miller T, Sonenberg L, Vetere F (2019) A grounded interaction protocol for explainable artificial intelligence. In: Proceedings of the 18th international conference on autonomous agents and multiagent systems, p 1033–1041

57. Manns M, Fischer K, Du H, Slusallek P, Alexopoulos K (2018) A new approach to plan manual assembly. Int J Comput Integr Manuf 31(9):907–920. https://doi.org/10.1080/0951192X.2018.1466396
58. Mehri S, Choi J, D'Haro LF, Deriu J, Eskénazi M, Gasic M, Georgila K, Hakkani-Tür DZ, Li Z, Rieser V, Shaikh S, Traum DR, Yeh YT, Yu Z, Zhang Y, Zhang C (2022) Report from the NSF future directions workshop on automatic evaluation of dialog: research directions and challenges. ArXiv **abs/2203.10012**. https://doi.org/10.48550/arXiv.2203.10012
59. Melas-Kyriazi L, Rush A, Han G (2018) Training for diversity in image paragraph captioning. In: Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, p 757–761. https://doi.org/10.18653/v1/D18-1084
60. Miech A, Zhukov D, Alayrac JB, Tapaswi M, Laptev I, Sivic J (2019) Howto100m: Learning a text-video embedding by watching hundred million narrated video clips. In: Proceedings of the IEEE/CVF International conference on computer vision (ICCV)
61. Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007
62. Mittelstadt B, Russell C, Wachter S (2019) Explaining explanations in AI. In: Proceedings of the conference on fairness, accountability, and transparency, FAT* '19. Association for Computing Machinery, New York, NY, USA, p 279–288. https://doi.org/10.1145/3287560.3287574
63. Mokady R, Hertz A, Bermano AH (2021) Clipcap: clip prefix for image captioning. https://doi.org/10.48550/ARXIV.2111.09734
64. Nguyen DM, Nguyen TT, Vu H, Pham Q, Nguyen MD, Nguyen BT, Sonntag D (2022) TATL: task agnostic transfer learning for skin attributes detection. Med Image Anal 78:102359. https://doi.org/10.1016/j.media.2022.102359
65. Nielsen TAS, Haustein S (2018) On sceptics and enthusiasts: What are the expectations towards self-driving cars? Transport Policy 66:49–55. https://doi.org/10.1016/j.tranpol.2018.03.004
66. Norris S, Guilbert S, Smith M, Hakimelahi S, Phillips L (2005) A theoretical framework for narrative explanation in science. Science Education 89:535–563. https://doi.org/10.1002/sce.20063
67. Nunnari F, Kadir MA, Sonntag D (2021) On the overlap between grad-cam saliency maps and explainable visual features in skin cancer images. In: International cross-domain conference for machine learning and knowledge extraction. Springer, p 241–253. https://doi.org/10.1007/978-3-030-84060-0_16
68. Pavllo D, Feichtenhofer C, Grangier D, Auli M (2019) 3D human pose estimation in video with temporal convolutions and semi-supervised training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 7753–7762
69. Penha G, Balan A, Hauff C (2019) Introducing mantis: a novel multi-domain information seeking dialogues dataset . https://doi.org/10.48550/ARXIV.1912.04639
70. Perez E, Huang S, Song F, Cai T, Ring R, Aslanides J, Glaese A, McAleese N, Irving G (2022) Red teaming language models with language models . https://doi.org/10.48550/ARXIV.2202.03286
71. Plappert M, Mandery C, Asfour T (2016) The KIT motion-language dataset. Big Data 4(4):236–252. https://doi.org/10.1089/big.2016.0028
72. Poibrenski A, Sprenger J, Müller C (2018) Towards a methodology for training with synthetic data on the example of pedestrian detection in a frame-by-frame semantic segmentation task. In: Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems, SEFAIS '18. Association for Computing Machinery, New York, NY, USA, p 31-34. https://doi.org/10.1145/3194085.3194093
73. Prange A, Barz M, Sonntag D (2017) Speech-based medical decision support in vr using a deep neural network (demonstration). In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17, p 5241–5242. https://doi.org/10.24963/ijcai.2017/777
74. Qu C, Yang L, Croft WB, Trippas JR, Zhang Y, Qiu M (2018) Analyzing and characterizing user intent in information-seeking conversations. In: The 41st International ACM SIGIR conference on research and development in information retrieval, SIGIR '18. Association for Computing Machinery, New York, NY, USA, p 989–992. https://doi.org/10.1145/3209978.3210124
75. R. Menon R, Ghosh S, Srivastava S (2022) CLUES: a benchmark for learning classifiers using natural language explanations. In: Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Association for Computational Linguistics, Dublin, Ireland, p 6523–6546. https://doi.org/10.18653/v1/2022.acl-long.451
76. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. In: Meila M, Zhang T (eds) Proceedings of the 38th International conference on machine learning, Proceedings of machine learning research, vol 139. PMLR, p 8748–8763. https://proceedings.mlr.press/v139/radford21a.html
77. Ramos G, Meek C, Simard P, Suh J, Ghorashi S (2020) Interactive machine teaching: a human-centered approach to building machine-learned models. Human-Computer Interaction 35(5–6):413–451. https://doi.org/10.1080/07370024.2020.1734931
78. Rebolledo-Mendez G, de Freitas S, Gaona ARG (2009) A model of motivation based on empathy for ai-driven avatars in virtual worlds. In: 2009 Conference in Games and Virtual Worlds for Serious Applications, pp. 5–11. IEEE . https://doi.org/10.1109/VS-GAMES.2009.33
79. Rempe D, Birdal T, Hertzmann A, Yang J, Sridhar S, Guibas LJ (2021) Humor: 3d human motion model for robust pose estimation. In: Proceedings of the IEEE/CVF international conference on computer vision, p 11488–11499. https://doi.org/10.1109/ICCV48922.2021.01129
80. Ribera M, Lapedriza À (2019) Can we do better explanations? a proposal of user-centered explainable ai. In: IUI Workshops
81. Rohrbach A, Hendricks LA, Burns K, Darrell T, Saenko K (2018) Object hallucination in image captioning. In: Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, p 4035–4045. https://doi.org/10.18653/v1/D18-1437
82. Samek W, Wiegand T, Müller KR (2017) Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. ITU J 1:1–10. ICT Discoveries - Special Issue 1 - The Impact of Artificial Intelligence (AI) on Communication Networks and Services
83. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, p 618–626. https://doi.org/10.1109/ICCV.2017.74
84. Simonyan K, Vedaldi A, Zisserman A (2013) Deep inside convolutional networks: visualising image classification models and saliency maps . https://doi.org/10.48550/ARXIV.1312.6034
85. Slack D, Krishna S, Lakkaraju H, Singh S (2022) Talktomodel: Understanding machine learning models with open ended dialogues. CoRR **abs/2207.04154**
86. Smith-Renner A, Fan R, Birchfield M, Wu T, Boyd-Graber J, Weld DS, Findlater L (2020) No explainability without accountability: an empirical study of explanations and feedback in interactive ML. Association for Computing Machinery, New York, NY, USA, p 1–13. https://doi.org/10.1145/3313831.3376624

87. Sonntag D, Profitlich HJ (2018) An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation. Artif Intell Med 93:13–28. https://doi.org/10.1016/j.artmed.2018.08.003
88. Sonntag D, Schulz C, Reuschling C, Galarraga L (2012) Radspeech's mobile dialogue system for radiologists. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI '12. Association for Computing Machinery, New York, NY, USA, p 317–318. https://doi.org/10.1145/2166966.2167031
89. Spinner T, Schlegel U, Schäfer H, El-Assady M (2019) explainer: A visual analytics framework for interactive and explainable machine learning. IEEE transactions on visualization and computer graphics 26(1):1064–1074. https://doi.org/10.1109/TVCG.2019.2934629
90. Stasaski K, Kao K, Hearst MA (2020) CIMA: A large open access dialogue dataset for tutoring. In: Proceedings of the difteenth workshop on innovative use of NLP for building educational applications. Association for Computational Linguistics, Seattle, WA, USA (Online), p 52–64. https://doi.org/10.18653/v1/2020.bea-1.5.
91. Strobelt H, Hoover B, Satyanaryan A, Gehrmann S (2021) LMdiff: A visual diff tool to compare language models. In: Proceedings of the 2021 conference on empirical methods in natural language processing: system demonstrations. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, p 96–105. https://doi.org/10.18653/v1/2021.emnlp-demo.12
92. Strobelt H, Kinley J, Krueger R, Beyer J, Pfister H, Rush AM (2022) Genni: Human-aI collaboration for data-backed text generation. IEEE Trans Visual Comput Graph 28(1):1106–1116. https://doi.org/10.1109/TVCG.2021.3114845
93. Stumpf S, Rajaram V, Li L, Burnett M, Dietterich T, Sullivan E, Drummond R, Herlocker J (2007) Toward harnessing user feedback for machine learning. In: Proceedings of the 12th international conference on intelligent user interfaces, IUI '07. Association for Computing Machinery, New York, NY, USA, p 82–91. https://doi.org/10.1145/1216295.1216316
94. Stumpf S, Rajaram V, Li L, Wong WK, Burnett M, Dietterich T, Sullivan E, Herlocker J (2009) Interacting meaningfully with machine learning systems: three experiments. Int J Hum Comput Stud 67(8):639–662. https://doi.org/10.1016/j.ijhcs.2009.03.004
95. Sun Y, Bao Q, Liu W, Fu Y, Black MJ, Mei T (2021) Monocular, one-stage, regression of multiple 3D people. In: Proceedings of the IEEE/CVF international conference on computer vision, p. 11179–11188. https://doi.org/10.1109/ICCV48922.2021.01099
96. Sundararajan M, Taly A, Yan Q (2017) Axiomatic attribution for deep networks. In: Proceedings of the 34th international conference on machine learning—vol 70, ICML'17, p 3319–3328. JMLR.org
97. Teach RL, Shortliffe EH (1981) An analysis of physician attitudes regarding computer-based clinical consultation systems. Compute Biomed Res 14(6):542–558. https://doi.org/10.1016/0010-4809(81)90012-4
98. Tenney I, Wexler J, Bastings J, Bolukbasi T, Coenen A, Gehrmann S, Jiang E, Pushkarna M, Radebaugh C, Reif E, Yuan A (2020) The language interpretability tool: extensible, interactive visualizations and analysis for NLP models. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations. Association for Computational Linguistics, p 107–118. Online
99. Teso S, Kersting K (2019) Explanatory interactive machine learning. In: Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society, AIES '19. Association for Computing Machinery, New York, NY, USA, p 239–245. https://doi.org/10.1145/3306618.3314293
100. Wang D, Yang Q, Abdul A, Lim BY (2019) Designing theory-driven user-centric explainable AI. In: Proceedings of the 2019 CHI conference on human factors in computing systems, CHI '19. Association for Computing Machinery, New York, NY, USA, p 1–15. https://doi.org/10.1145/3290605.3300831
101. Weitz K, Vanderlyn L, Vu NT, André E (2021) "it's our fault!": insights into users' understanding and interaction with an explanatory collaborative dialog system. In: Proceedings of the 25th conference on computational natural language learning. Association for Computational Linguistics, p 1–16. Online. https://doi.org/10.18653/v1/2021.conll-1.1
102. Weld DS, Bansal G (2019) The challenge of crafting intelligible intelligence. Commun ACM 62(6):70–79. https://doi.org/10.1145/3282486
103. Wiegreffe S, Marasović A (2021) Teach me to explain: a review of datasets for explainable natural language processing. https://doi.org/10.48550/ARXIV.2102.12060
104. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, p 2048–2057. PMLR
105. Yang Y, Yuan S, Cer D, Kong Sy, Constant N, Pilar P, Ge H, Sung YH, Strope B, Kurzweil R (2018) Learning semantic textual similarity from conversations. In: Proceedings of the third workshop on representation learning for NLP. Association for Computational Linguistics, Melbourne, Australia, p 164–174. https://doi.org/10.18653/v1/W18-3022.
106. Yu KH, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2(10):719–731. https://doi.org/10.1038/s41551-018-0305-z
107. Zhang C, Li Y, Du N, Fan W, Yu P (2019) Joint slot filling and intent detection via capsule neural networks. In: Proceedings of the 57th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Florence, Italy, p 5259–5267. https://doi.org/10.18653/v1/P19-1519
108. Zhou B, Wang P, Wan J, Liang Y, Wang F, Zhang D, Lei Z, Li H, Jin R (2021) Decoupling and recoupling spatiotemporal representation for RGB-D-based motion recognition. arXiv preprint arXiv:2112.09129