Supplementary Material for Journal Article
"Explainable Software Systems: From Requirements Analysis to System Evaluation"

Larissa Chazette*, Wasja Brunotte*†, Timo Speith‡
*Leibniz University Hannover, Software Engineering Group, Hannover, Germany
†Leibniz University Hannover, Cluster of Excellence PhoenixD, Hannover, Germany
‡University of Bayreuth, Chair for Philosophy, Computer Science and Artificial Intelligence, Bayreuth, Germany
Email: {larissa.chazette, wasja.brunotte}@inf.uni-hannover.de, timo.speith@uni-bayreuth.de

CONTENTS

I. About

This is the supplementary material for the research paper *"Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue"* accepted at the IEEE 29th Requirements Engineering Conference [1]. Since the space allocated for research papers did not suffice to elaborate on all research results, this document serves to amend this lack. In particular, more details on the Systematic Literature Review (SLR) – e.g., inclusion/exclusion criteria, complete list of analyzed papers, codes, etc. – and our workshops can be found here. For a deeper understanding of the supplementary material we recommend to have a look at the corresponding paper [1].

The specific content of this work is structured as follows. In the next section (Sec. II), our SLR is explained in more detail. Specific details will be given about how we selected the venues for our manual search and, more generally, about our inclusion and exclusion criteria and the selection procedure. Furthermore, all papers found during the SLR will be mentioned.

Sec. III will link the final selection of our papers with how they answer our research questions (RQs). We provide two tables, one for RQ1 and the second for RQ2 and RQ3.

In Section IV, we will present the codes resulting from our coding process for the data related to RQ2 and RQ3. We deliberately do not present codes from RQ1, as we have coded more categories and data than presented here or in the corresponding research paper. We will address these categories codes in a future publication. For the purposes here, TABLE II should be sufficient to illustrate our codes.

Finally, we will elaborate the workshops in detail in Sec. V. This includes all activities that the participants had to do before and during the workshops.

II. Systematic Literature Review

The search strategy for our SLR consisted of a manual search followed by a snowballing process. Table I shows the sources that were part of our SLR. The *Total Retrieved* column contains the total number of papers that were analyzed for eligibility in the respective source. *Initial selection* indicates the number of papers that have been pre-selected for closer examination. The *Final Selection* column indicates the number of papers that were ultimately selected for our start set, as these papers met our inclusion criteria.

A. Found Papers

1) Manual Search: The following 104 papers are the result of our manual search and the application of our selection procedure: [2–105]

2) Snowballing: Since our review of the literature was partially based on the grounded theory approach to literature reviews proposed by [106], we stopped our snowballing process during the second snowballing iteration. Up to here, our data achieved a saturation and we have not gained any new insights nor have we discovered any new concepts or categories that arose from the data. The following 125 papers are the result of our forward & backward snowballing and the application of our selection procedure: [107–231]

TABLE I: Manual Search Sources and Paper Selection

Source	Total Retrieved	Initial Selection	Final Selection
International Requirements Engineering Conference (RE)	1312	15	4
Symposium on the Foundations of Software Engineering (FSE)	667	8	2
Information and Software Technology (IST)[1]	2668	23	0
Intelligent User Interfaces (IUI)	1158	52	18
Journal of Systems and Software (JSS)[2]	4121	8	0
Transaction on Software Engineering (TSE)	2910	23	0
Conference on Information and Knowledge Management (CIKM)	2789	8	2
International Working Conference on Requirement Engineering: Foundation for Software Quality (REFSQ)	328	4	0
Transactions on Software Engineering and Methodology (TOSEM)	615	4	1
Conference on Recommender Systems (RecSys)	521	15	6
Requirements Engineering Journal (REJ)	455	9	2
RE Workshops	21	4	1
REFSQ Workshops[3]	162	5	1
Minds and Machines	724	30	17
Big Data & Society	284	16	6
International Joint Conference on Artificial Intelligence - Workshop on eXplainable Artificial Intelligence[4]	63	41	34
Philosophy and Technology[5]	259	10	9
Ethics and Information Technology	538	1	1

B. Overview of the studies

Fig. 1(a) gives an overview of the number of primary and secondary studies in our SLR. Fig. 1(b) gives the number of papers in our SLR grouped by interdisciplinarity. We divided the field of computer science into human-computer interaction (HCI), RE, and computer science in general. We grouped studies from the fields of philosophy and psychology together, since these disciplines often publish in the same venues and are closely intertwined. The Interdisciplinary group combines studies from different fields such as Law, Business, and venues that are interdisciplinary in nature.

III. References for Answering the RQs

A. RQ1

We partially derived our presented definition related to RQ1 from the literature. As a first factor, the following sources highlight the importance of understanding: [4, 5, 11, 16, 20, 26, 29, 30, 32, 46, 51, 56, 69–72, 80, 87, 88, 91, 97, 98, 102, 104, 105, 107, 110, 112, 113, 115, 121, 122, 128, 131, 137, 138, 146–148, 151, 153, 156, 159, 163, 166, 167, 169–171, 174, 176, 179, 180, 186, 189–192, 196–198, 201, 202, 205, 207, 208, 211, 212, 220, 228, 231]. As a second factor, the following sources specified certain aspects of a system that should be explained:

[1]20 publications out of 2668 publications were not accessible.
[2]Seven publications out of 4121 publications were not accessible.
[3]Only the proceedings of the years 2000, 2001, 2006-2008, 2010, 2011 and 2015-2019 were accessible.
[4]Proceedings from 2017-2019 were considered
[5]Only the issues beginning from 2011 were accessible.

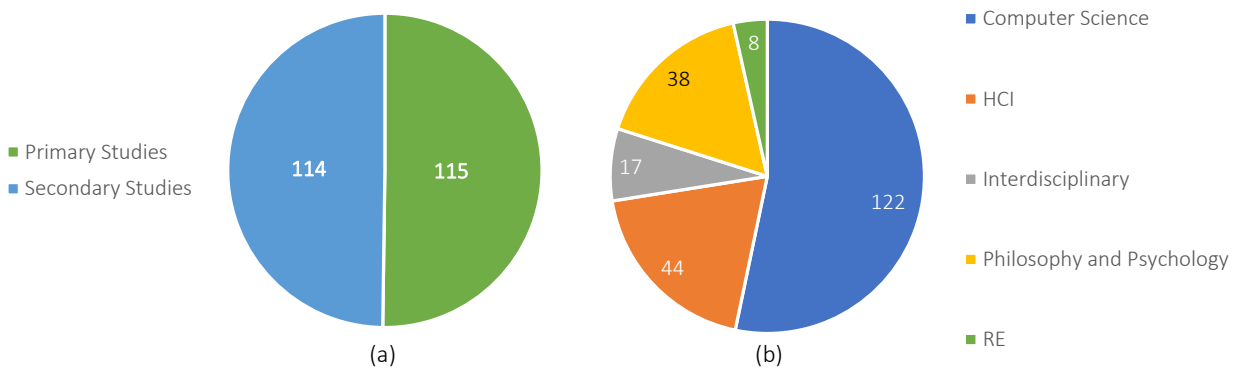

Fig. 1. **Overview of the studies** — (a) primary and secondary studies – (b) overview of interdisciplinarity

TABLE II: Aspects to Explain (RQ1)

Aspect to explain	**References**
System in General	[26, 88, 90, 105, 136, 144, 148, 151, 161, 171, 186, 189, 211, 212, 220]
System's Reasoning Processes	[42, 71, 101, 146, 153, 166, 170, 171, 178, 217, 220, 221]
System's Inner Logic	[26, 29, 39, 56, 64, 89, 97, 102, 137, 146, 170, 224, 230]
System's Model's Internals	[9, 26, 69, 84, 97, 98, 126, 158, 159, 170, 174]
System's Intention	[110, 180, 220]
System's Behavior	[88, 90, 208, 217, 225]
System's Decision	[26, 78, 80, 117, 157, 158, 168, 189, 192]
System's Performance	[101, 196, 220]
Knowledge about User or World	[80, 88, 90, 192, 221]

TABLE III: Types of Explanation (RQ1)

Type of explanation	**References**
Answer to "W"-Questions	[9, 69, 107, 116, 125, 146, 153, 154, 161, 170, 171, 186, 190, 204]
Scientific Explanation	[23, 28, 43, 46, 48, 69, 87, 88, 121, 126, 127, 139, 163, 173, 201, 204]
Causal Explanation	[9, 11, 28, 30, 33, 41, 44–46, 88, 107, 110, 121, 127, 135, 138, 142, 146, 153, 164, 165, 173, 199, 201, 206, 222]
Mechanistic Explanation	[2, 18, 37, 69, 72, 84, 104, 110, 162, 222]
Counterfactual Explanation	[22, 70, 104, 107, 131, 188]
Nomological Explanation	[23, 43, 46, 139, 173]
Functional Explanation	[44, 46, 104, 162, 222]
Reason Explanation	[9, 16, 21, 42, 45, 71, 98, 113, 115, 116, 121, 125, 161, 181, 195, 209, 228]
Contrastive Explanation	[20, 30, 131, 138, 153, 165, 188, 199, 225]
Arguments	[49, 63, 65, 120, 125, 168, 169, 201]

TABLE IV: Output Formats (RQ1)

Output Format	References
Visual	[9, 10, 15, 20, 22, 24–26, 34, 40, 49, 55, 61, 62, 65, 68, 70, 73, 74, 78–83, 94–96, 101, 103, 105, 114, 115, 117, 118, 120, 122, 125, 138, 141, 144, 147, 148, 150, 151, 154, 167, 171, 172, 174–176, 180–183, 185, 186, 189–194, 196, 204, 210, 216, 218, 220, 222, 224, 226, 231]
Textual	[4, 15, 16, 21, 26, 32, 33, 40, 49, 54, 59–63, 65, 67, 71, 74, 80–83, 85, 92–95, 97, 99, 100, 102, 107, 112, 115, 117, 118, 122, 133, 136, 138, 139, 141, 144, 151, 169–172, 175–178, 180, 182–184, 186, 187, 189, 190, 192, 193, 195–198, 204, 205, 207–209, 213–216, 221–224, 227–231]
Numerical	[70, 101, 172, 186, 190, 224, 228]
Audio	[107, 147, 171, 180, 188, 195, 226]
Mixed	[15, 26, 40, 61, 62, 80, 81, 118, 147, 175, 176, 225, 226]
Natural Language	[49, 55, 57, 115, 120, 121, 124, 139, 140, 148, 172, 174, 181, 188, 190, 217, 221, 224, 226, 227]
Rules	[71, 94, 115, 138, 190, 228, 229]
Model	[138, 151]
Arguments	[115, 120, 125, 169]

TABLE V: Implementation Strategy

Implementation	References
Meta-Explanation	[103, 120, 177]
Feature Relevance	[14, 20, 22, 24, 62, 68, 74, 78, 84, 95, 116, 128, 138, 146, 150, 186, 188, 189, 191, 195, 216–218, 220]
Examples	[17, 22, 59, 66, 83, 116, 186, 189, 222]
Structure Leveraging	[14, 16, 24, 31, 61, 71, 114, 115, 121, 133, 177, 189, 197]
Surrogates	[25, 51, 58, 71, 87, 89, 91, 114, 131, 138, 150, 176, 186, 189, 190, 222, 229]
Perturbations	[62, 128, 189, 231]
Explanation Module	[16, 22, 32, 51, 59, 60, 63, 70, 75, 80, 81, 92–97, 100, 107, 120, 124, 133, 139, 141, 147, 155, 183–185, 187, 188, 194, 197, 198, 209, 213, 217, 223, 224, 230]
Interactive	[54, 66, 155, 230]
Self-Explaining	[112, 114]
Post-Hoc	[9, 13, 15, 16, 22, 25, 32, 36, 42, 51, 51, 55, 58–63, 70, 71, 75, 80, 81, 87, 89, 91–93, 93–97, 99, 107, 114–116, 120, 124, 126, 128, 131, 133, 139, 141, 147, 150, 151, 155, 172, 175–177, 183–185, 187–190, 194, 195, 197, 198, 209, 213, 217, 218, 222, 227, 229–231]
Ante-Hoc	[13, 14, 24, 31, 62, 83, 89, 92, 112, 116, 125, 126, 150–152, 155, 188–190, 205, 221, 222]
Model-Agnostic	[62, 114, 126, 128, 172, 189, 231]
Model-Specific	[16, 22, 114–116, 121, 133, 189, 197, 227]

TABLE VI: Study Methods

Method	References
Interviews	[196, 207]

User-Study	[10, 22, 30, 33, 38–40, 51, 53, 55, 57, 59, 60, 65, 80–83, 85, 98–100, 116, 118, 130, 133, 139, 144–147, 152, 157, 160, 162, 168, 174, 177, 183, 184, 187, 188, 191, 193, 194, 197, 207, 209, 213–215, 218, 220, 221, 223, 230]
Case-Study	[34, 58, 102, 116, 177, 184, 210]
Empirical Study	[46, 54, 63, 67, 117, 124]
Questionnaire	[16, 35, 38, 51, 53, 57, 59, 74, 85, 100, 115, 122, 123, 133, 141, 155, 170, 174, 178, 192, 194, 207, 216, 221]
Baseline	[81, 115, 116, 144]
Alternative Explanations	[16, 24, 25, 49, 68, 80, 81, 116, 118, 122, 130, 139, 144, 160, 162, 213, 216, 224]
Mental Model Comparison	[10, 148]
AB-Tests	[50, 118, 141]
Task-Based	[99, 182, 209]

TABLE VII: Explanation Quality

Quality Measure	**References**
Clarity of Language	[113]
Soundness	[94, 114, 120, 145, 216]
Awareness	[16, 194]
Acceptance	[30, 132, 144, 146]
Actionability	[105, 132, 145]
Accessability	[132]
Accuracy	[16, 62, 74, 80, 94, 115, 118, 220]
Capability	[83]
Completeness	[94, 113, 114, 120, 132, 145, 157, 216, 221]
Compactness	[114]
Comprehensibility	[16, 30, 91, 114, 122, 129, 132, 194, 213, 227]
Complexity	[9, 91, 145]
Correctness	[24, 25, 221]
Coherence	[30, 145, 165]
Consistency	[46]
Comparison to Ground-Truth	[70, 185, 193]
Interactivity	[9]
Effort	[115]
Mental Model Accuracy	[10, 74, 180]
Plausibility	[16, 46]
Relatability	[16]
Persuasiveness	[130]
Usefulness	[27, 30, 94, 130, 138, 147, 224, 231]
Effectiveness	[22, 27, 31]
(Task) Efficiency	[74, 100, 223]
Realism	[22]
Timeliness	[132]

Believability	[132]
Relevancy	[30, 80, 132, 218]
Scope	[134]
Depth	[134]
Novelty	[145]
Naturalness	[147]
Simplicity	[30, 46, 165]
Task load	[146, 148]
Helpfulness	[152]
Faithfulness/Fidelity	[91, 172, 190, 216]
Generalizability	[190]
Diversity	[218]
Reliability	[217]

B. RQ2 and RQ3

TABLE VIII: Quality Aspects Impacted by Explainability (RQ2 & RQ3)

Quality Aspect	References
Accountability	Pos: [5, 15, 52, 55, 58, 78, 114, 117, 131, 138, 140, 145, 158, 159, 165, 167, 170, 184, 188, 193, 210, 217, 229]
Accuracy	Pos: [35, 66, 103, 121, 175, 182, 224] Neg: [6, 13, 29, 55, 62, 92, 126, 134, 140, 145, 146, 150, 153, 189, 221]
Adaptability	Workshop Exclusive
Auditability	Pos: [20, 21, 26, 34, 38, 48, 57, 78, 81, 83, 84, 117, 121, 122, 166, 180, 182, 186, 189, 203, 205, 209, 213, 229]
Complexity	Workshop Exclusive
Compliance	Pos: [15, 53, 57, 94, 143, 146, 159, 210]
Confidence in the System	Pos: [16, 41, 42, 96, 109, 115, 116, 121, 126, 129, 135, 144, 147, 151, 154, 156, 161, 166, 167, 170, 176, 178, 184, 191, 197, 198, 209, 228] Neg: [40, 184, 213]
Correctness	Pos: [15, 117, 192]
Customer Loyalty	Pos: [16, 27, 49, 75, 100, 115, 136, 142, 181, 197]
Debugging	Pos: [10, 15, 29, 48, 54, 115, 121, 122, 135, 138, 140, 146, 147, 164, 167, 172, 186, 188, 189, 191, 210, 216, 227, 229, 230]
Decision Justification	Pos: [15, 29, 80, 109, 184, 189, 214, 221]
Development Cost	Neg: [29, 98, 105, 121, 134, 180, 189]
Effectiveness	Pos: [53, 94, 138]
Efficiency	Neg: [134, 140, 189]
Ethics	Pos: [5, 126, 146, 159, 180, 190, 228]
Extensibility	Workshop Exclusive
Fairness	Pos: [11, 21, 38, 56, 64, 67, 84, 117, 126, 138, 143, 145, 146, 151, 157, 161, 170, 184, 186, 188–190, 196]
Guidance	Pos: [8, 26, 39, 53, 117, 224]
Human-Machine Cooperation	Pos: [38, 58, 107, 151, 156, 186, 219, 231]

Continued on next page ...

TABLE VIII – continued from previous page

Quality Aspect	References
Knowledge Discovery	Pos: [8, 146, 151, 180, 186, 189, 190, 196, 210]
Learnability	Pos: [15, 74, 121, 147, 154, 156, 158, 165, 170, 171, 203, 211, 212, 229]
Maintenability	Workshop Exclusive
Mental Model Accuracy	Pos: [40, 62, 74, 83, 102, 122, 147, 161, 165, 166, 170, 177, 181, 196, 208, 213]
Model Optimization	Pos: [61, 138, 150, 151, 157, 186, 205, 210, 214]
Perceived Usefulness	Pos: [81, 82, 85, 111, 115, 130, 136, 160, 170, 171, 224]
Perceived Value	Pos: [79, 83, 94, 100, 115, 121, 142, 161, 170, 171, 178, 183, 189, 193, 194, 207, 208, 215, 224]
Performance	Pos: [93, 170, 196, 207, 231] Neg: [9, 26, 29, 31, 133, 134, 151]
Persuasiveness	Pos: [39, 49, 62, 75, 80, 81, 81, 98, 115, 116, 121, 130, 136, 142, 144, 152, 160, 165, 170, 184, 190, 194, 207, 209, 221, 224, 224, 229]
Portability	Workshop Exclusive
Predictability	Pos: [41, 62, 118]
Privacy	Pos: [77, 126, 159, 174, 186] Neg: [57, 77, 145, 151, 167, 188, 190, 191, 200]
Privacy Awareness	Pos: [151, 186, 215]
Real-Time Capability	Workshop Exclusive
Reliability	Pos: [159, 186, 222]
Robustness	Pos: [95, 150, 151]
Safety	Pos: [11, 114, 118, 126, 146, 186]
Scrutability	Pos: [53, 99, 115, 136, 138, 183, 184, 211, 221]
Security	Pos: [42, 126, 150] Neg: [14, 29, 42, 55, 57, 84, 119, 145, 188]
Stakeholder Trust	Pos: [4, 8, 11, 12, 14, 16, 21, 22, 27, 29, 31–33, 36, 38, 41, 42, 49, 51, 57–59, 62, 70, 71, 75, 77, 79–82, 84, 87, 90, 91, 93–99, 101, 103, 107, 109, 113–116, 118, 119, 121, 122, 125, 126, 131, 135, 136, 138, 141, 145, 146, 148, 150, 152–154, 156, 158, 159, 161, 165, 167, 171, 172, 178–181, 184, 186, 187, 189–192, 194, 204, 208, 215, 219, 221–224, 226, 228, 229, 231] Neg: [42, 62, 82, 84, 96, 98, 138, 178, 184, 191, 213, 224]
Support Decision Making	Pos: [8, 26, 39, 49, 80, 105, 111, 115, 121, 122, 130, 138, 151, 152, 160, 161, 170, 182, 190, 196, 209, 211, 224]
System Acceptance	Pos: [6, 8, 39, 41, 70, 77, 81, 90, 100, 118, 121, 122, 129, 136, 154, 161, 170, 171, 176, 180, 192, 194, 196, 197, 210, 223, 224]
Testability	Pos: [121, 146, 186]
Trade Secrets	Neg: [26, 55, 57, 145, 151]
Transferability	Pos: [151, 180, 185, 190]
Transparency	Pos: [8, 26, 32, 39, 42, 81, 87, 94, 100, 102, 109, 115, 121, 136, 138, 141, 147, 148, 154, 160, 166, 178, 180, 183, 184, 207, 219, 221, 223, 224]
Trustworthiness	Pos: [5, 6, 29, 50, 114, 116, 121, 147, 151, 158, 228, 231]
Understandability	Pos: [6, 12, 15–17, 26, 29, 35, 38, 45, 50, 53, 60, 73, 75–78, 82, 91, 94, 96, 101, 103, 108, 111–113, 115, 116, 122–125, 130, 131, 134, 138, 148, 150, 155, 156, 161, 165, 167, 170, 174, 175, 178, 189, 191, 197, 198, 207, 210, 213, 214, 219, 222, 231] Neg: [26, 98, 132]

Continued on next page ...

TABLE VIII – continued from previous page

Quality Aspect	References
Usability	Pos: [4, 26, 27, 29, 35, 49, 59, 91, 100, 115, 121, 134, 136, 160, 170, 177, 186, 191, 196, 215, 222, 224] Neg: [26, 40, 148]
User Awareness	Pos: [170, 177, 178, 191]
User Control	Pos: [5, 115, 146, 154, 190, 215]
User Effectiveness	Pos: [45, 80, 115, 116, 121, 144, 219, 221, 224] Neg: [65, 115, 213]
User Efficiency	Pos: [62, 82, 115, 121, 136, 144, 147, 184, 221, 224] Neg: [31, 82, 155]
User Experience	Pos: [49, 50, 79, 99, 181, 186, 191, 207] Neg: [26, 83, 165]
User Performance	Pos: [45, 121, 142, 161, 168, 171, 178]
User Satisfaction	Pos: [27, 39–41, 49, 59, 80, 82, 83, 98, 115, 121, 122, 134, 136, 141, 144, 152, 168, 170, 171, 176, 181, 183, 184, 187, 194, 197, 224]
Validation	Pos: [87, 111, 115, 121]
Verifiability	Pos: [41, 84, 96, 121, 157, 161, 166, 170, 172, 186, 198, 214]

IV. Codes - RQ2 & RQ3

In the material below, "P" indicates that explainability has a positive impact on the corresponding quality aspect, and "N" stands for a negative impact. We defined a threshold for producing a code during the coding procedure. This means that only impacts were considered that were mentioned by at least *three* different sources. This approach was taken to ensure more reliable results. The material below is already cleaned up in this respect, such that codes that are not supported by at least three sources are not listed.

Accountability_P
- a critical piece in achieving accountability
- accountability
- allows to be held accountable
- assess accountability
- assignment of blame
- better manage liability
- enforce accountability under the law
- facilitate accountability
- fulfill the goal of accountability
- hold entities accountable
- important for responsibility of algorithms
- important for liability
- important to achieve responsible AI
- know where the blame lies when things go wrong
- legal accountability
- liability
- liability for machines
- means to promote accountability
- render decision-making more accountable
- render decisions more accountable
- shift responsibility from algorithm to participant
- used to determine legal culpability

Accuracy_N
- can negatively affect predictive accuracy
- cost of classification accuracy
- decrease in predictive performance
- inverse relationship between accuracy and explainability
- may conflict with precision
- requires a sacrifice for accuracy
- tension between accuracy and explainability
- the higher the accuracy, the lower the explainability
- the more explainable, the less accurate
- there is a well-known trade-off between accuracy and explainability
- trade off explainability against accuracy
- trade-off between accuracy and explanatory power
- yield lower accuracy

Accuracy_P
- improve classifiers
- improve prediction accuracy
- improve recommendation accuracy
- improved accuracy of decisions
- improving the accuracy
- improving the accuracy of VQA systems
- increase accuracy measures

Auditability_P
- allow the system to demonstrate partial capability
- allows for question-answering about the system's reasoning steps
- allows to look why the system has reached an anomalous result
- assess the reliability of systems
- auditability
- benefit the auditability
- describe the methods employed and the reasons for exploying them
- detect errors in input data that led to adverse decisions
- enables models to be audited
- estimate likelihood of errors in decision making
- examine the knowledge of the system
- help to identify errors and biases
- help users understand what contributed to the large error
- identify that the system has erred
- indicate prior reliability
- is an auditable and provable way to defend algorithmic decisions
- judge whether a prediction is correct
- satisfying the desire to know whethe an algorithm caused an error
- scrutinize individual cases
- support evaluation of system conclusions
- understand the scope of an error and solve discrepancies
- understand the underlying technicalities and models

Compliance_P
- ensure legal decisions are made
- implement a right to explanation
- make model performance and guideline compliance compatible
- necessary for compliance
- used to comply with regulatory and policy changes
- useful for increased compliance

Confidence in System_N
- decrease reliance if level of detail insufficient
- had a negative impact on user confidence
- question the system and lead to self-reliance

Confidence in System_P
- aim at acquiring confidence
- allow to develop appropriate reliance
- can lead to a higher level of confidence
- confidence
- confidence in decision making
- decides how much confidence to place in recommendation
- develop appropriate reliance
- higher confidence in system recommendations
- higher degrees of confidence
- improve confidence of the decision quality
- increase confidence in system's abilities
- increase reliance
- increase the end user’s confidence in the system’s conclusion
- increase user belief in system
- increase users' confidence in problem-solving competence
- increase user's confidence in the system
- increases their confidence
- instill confidence
- it can help build confidence
- might give us more confidence
- more user confidence
- significantly increase users’ confidence
- strenghten the confidence of users

Correctness_P
- help correct errors in input data
- necessary for correcting an error
- optimality and correctness

Customer Loyalty_P
- affect perceptions of the organization
- could help inspire loyalty
- decrease the likelihood of abandoning the system
- higher degrees of rapport
- higher intention to use the interface again
- increase the propability of using the system
- it can help build rapport
- make users more willing to continue the use of a system
- promote rapport
- rapport

Debugging_P
- better debug the system
- debug predictive models
- debug the procedure
- debugging
- debugging purposes
- determine and fix errors and faults
- diagnosis and refinement
- ease of debugging
- enable users to debug the intelligent agent
- enables models to be debugged
- end-user debugging
- for debugging
- for system debugging
- help in locating sources of error
- help to identify and correct errors
- likely to be used for debugging

- model debugging
- notifying users about the defects in the system
- play an important role in debugging systems
- potential for aiding program analysis
- support developers for system debugging
- support system debugging
- used in system diagnosis

Decision Justification_P
- justification (e.g., offering reasons for an action).
- justification (of individual recommendations)
- justification (why recommendation was made)
- justify a given decision
- justify actions and decisions
- justify the decision made
- justify the outcome of a particular classification
- provides the required information to justify results

Development Cost_N
- at the expense of development effort
- can be costly on multiple dimensions
- cost expensive
- could be costly
- development cost
- development time
- is expensive

Effectiveness_P
- increase overall effectiveness
- system effectiveness limited by inability to explain
- useful for increasing an intelligent system's overall effectiveness

Efficiency_N
- could result in less efficient systems
- trade off explainability against efficiency

Ethics_P
- achieve an evaluation of the ethical standards of a machine
- achieve ethical AI
- ensure ethical decisions are made
- ethical decision making
- ethics
- evaluate if a decision is not in conflict with ethical norms
- promote ethics
- support ethical reasons

Fairness_P
- a way to check fairness
- address concerns about lack of fairness
- assurance that fairness issues have been mitigated
- demonstrate a model's fairness
- detect discrimination
- develop solutions against algorithmic discrimination
- enable people to identify and address fairness
- enable the identification of discrimination
- enables to assess whether algorithm is just
- ensure algorithmic fairness
- ensure fair decisions are made
- ensure fairness
- ensure that algorithms do not encode discrimination
- fair decision making
- fairness
- fulfill the goal of fairness
- identify algorithmic biases
- important to ensure algorithmic fairness
- inspect fairness
- mitigate decision biases
- moderate assessments of fairness
- non-discrimination
- render decisions more fair
- support fairness in decision-making
- verify fairness
- way to defend algorithmic decisions as being fair

Guidance_P
- educate about product knowledge
- guide in solving problems
- guide users during the use of the system
- increase product knowledge
- provide guidance on how to reverse adverse decisions
- to better guide the user
- to educate users about product knowledge

Human-Machine Cooperation_P
- allow for a better human AI cooperation
- crucial for effective human machine interaction
- improve cooperation
- improve the human-machine interaction performance
- key to effective human-AI interaction
- necessary for human in the loop
- provide a more effective interface for the human in-the-loop
- the ability of a model to be interactive with the user

Knowledge Discovery_P
- causality
- gain further insights or evidence
- helpful for knowledge discovery
- helpful tool to gain knowledge
- knowledge generation
- knowledge/scientific discovery

Learnability_P
- aid learning
- allow developers to not make the same mistake again
- allow users to learn something from the system
- allows the humans to learn
- can facilitate learning
- educate the user about the process of generating a recommendation
- facilitate learning
- help developers and humans working with a system learn from it
- help users to learn about how the system works
- improve user learning
- learn how the system operates
- learning how the system works
- teach something and learn
- teach users how to achieve a task

Mental Model Accuracy_P
- a tool to build and refine inner knowledge models
- align and adapt the mental models of the participating parties
- assist participants in the conception of an accurate mental model
- be aware of the system's limitations
- can help users to develop better mental models of systems
- create a shared understanding of decisions
- discover when a system is pushed over the bound of its expertise
- form accurate picture about the system's reliability
- gradually improve the mental model
- help build useful mental models
- help the user to build a theory of mind
- help users to understand when they can rely on the system
- helps users determine whether their needs are known by the system
- helps users to revise their theory of mind
- keep agents on the same page with respect to their beliefs
- know what the limitations of the system are
- learn the methods and knowledge used in the problem solving process
- users can judge whether their goal match those of the system

Model Optimization_P
- detect bias in training data
- detecting model bias
- determine what should be learned by a model
- identify bias in training data
- identify problems in training data
- improve and develop ML models

- lead to the design of better models

Perceived Usefulness_P
- assess whether a recommendations is useful
- crucial aspects for system utility
- effect on perceived usefulness of recommendations
- enhance perceived usefulness of advice
- increase perceived information usefulness
- increase the perceived usefulness of a recommender system
- perceived as a useful system
- perceived system usefulness
- usefulness

Perceived Value_P
- a sense of control for the user
- adds to business value
- affect perceptions of fairness
- be perceived as easy to use
- being perceived as a trustworthy system
- benefit perceived reliability
- better user perception
- convince users of of a system's competence
- help improve user perceptions during system errors
- help increase perceived benevolence
- increase perceived control
- increase perceived efficiency
- increase perceived fairness
- increase perceived recommendation quality
- increase perceived safety
- increase users' perception of system competence
- increases users' perception of the interface's competence
- influence the extent to which the decision is viewed as fair
- more positive perceptions of a system
- perceive as more helpful
- perceive the interface as more competent
- perceived to be fair
- perception of system ease-of-use
- positively affect perceptions of system effectiveness
- promote more positive user perception of the system
- raise perception of integrity
- result in more positive user perceptions
- system perceived as more competent
- useful for increasing perceived value

Performance_N
- drawbacks in terms of performance
- loading memory issue
- loading time issue
- may conflict with performance
- require high computational costs
- take time to compute
- trade-off at the expense of performance
- trade-off with the performance of a model
- use of CPU, memory, storage, and battery

Performance_P
- help to decide between models
- improve AI performance
- improve overall system performance
- increase system performance
- learning time can be reduced

Persuasiveness_P
- affect task-taking motivation
- convince the user that the agent is correct
- convince the user to accept the result
- convince users that the suggested alternative is appropriate
- convince users to adopt recommentation
- convince users to try or buy
- critical to acceptance of systems' decisions
- greater adherence to system recommendation
- help users to understand and accept a recommendation
- improve persuasiveness
- improve user acceptance of recommendations
- improve users' acceptance of recommendations
- increase acceptance of system conclusions
- increase persuasiveness of the recommended messages
- increase probability of accepting a suggestion
- increase the acceptance of recommendations
- increase the system's persuasiveness
- increased the acceptance of the recommendations
- increasing system persuasivness
- influence applicant reactions
- invoke user's interest in recommendation
- make recommendations more persuasive
- may accept the system's choices
- more compliance with a request
- motivate users to consume items
- nudge the user in a certain direction
- persuade users to make a certain choice
- persuade users to try or purchase a recommended item
- persuasion
- persuasiveness
- positively affect users' purchasing beghavior
- suggestions being accepted by users

Predictability_P
- ability to predict the system's performance correctly
- determine when the system will make a mistake
- increase the user's ability to predict the classifier decision
- makes users more likely to correctly predict the model's success

Privacy Awareness_P
- lower privacy concerns
- privacy awareness - ability to assess privacy
- provides a way to check privacy

Privacy_N
- can demolish stakeholder's anonymity
- can hurt privacy
- can threaten privacy
- could jeopardize privacy
- data privacy
- negative consequences regarding privacy
- pose a privacy risk
- privacy implications
- privacy issues regarding training data
- privacy might be violated

Privacy_P
- can help privacy
- ensure privacy is uphold
- ensure that sensitive information is protected
- privacy
- promote privacy

Reliability_P
- improve reliability
- reliability

Robustness_P
- facilitate robustness
- improve model robustness
- instrumental in developing more robust systems

Safety_P
- guarantee safety concerns
- guarantee safety concerns are met
- help create safer products
- help increase safety
- improve safety
- increase safety
- safety

Scrutability_P
- alter which features of an item are deemed most important
- customize recommendations to fit users' preferences
- explain corrections back to the system
- increase scrutability
- provide scrutability
- scrutability

Security_N
- allow for gaming
- can threaten security
- enable individuals to game the system
- enhance capabilities of attackers
- facilitate manipuations of the system
- introduce security vulnerabilities
- may conflict with security
- pose a security risk
- security implications
- trade off with security

Security_P
- bridge the gap between 'actual security' and 'perceived security'
- identify illegitimate intrusion, malware, and other attacks
- protection mechanism can be improved by public scrutiny
- security

Stakeholder Trust_N
- bad explanation-for-trust may fail to create trust
- bad explanations may lead to distrust
- can undermine trust
- decrease trust
- effects on trust
- erode trust
- may lead to mistrust
- may lose trust
- minimum explanations can potentially harm user trust
- too much explanatio eroded user trust
- too much explanation can degrade trust

Stakeholder Trust_P
- achieve trust
- achieve user trust
- affect user trust
- agents tend to trust
- appropriately trust ML-based solutions
- assessing trust
- boost user trust
- breed trust
- build more trust in the interface
- build trust in the agent's choices
- build trust towards a machine's performance
- can address trust concerns
- can aim at acquiring trust
- can engender trust
- can promote trust in the system
- correctly calibrate trust in systems
- could help inspire user trust
- critical tool to build public trust in AI
- easier to trust
- enable domain experts and users to trust the model
- enable them to trust results
- enable to have appropriate level of trust
- engender trust in AI
- enhance trust in the classifier
- essential to assure trust
- foster trust
- foster user trust
- fundamental mechanism to increase user trust
- gain user trust
- generating trust in the model
- have a positive effect on user trust
- have the potential to build a trust relationship
- help gain user's trust
- help inspire trust in a recommender
- help promote trust
- help to establish a more appropriate level of trust
- impact perceptions of trust in the company
- important in assessing trust
- important in order to build human trust in systems
- improve user trust
- improving trust in decision by AI system
- improving trust in the advice
- increase trust
- increase trust in the recommender system
- increase trust in the system
- increases user trust
- increasing user trust
- instill trust
- means to establishing trust
- moderate trust to an appropriate level
- more likely to trust the underlying ML systems
- positive effects on organizational trust
- provide a way to check trust
- significant improvement in user trust
- significantly higher trust
- significantly increase users' trust
- students trust the system more
- support trust
- trust
- will enhance trust at the user side
- will trust the agent in the future

Support Decision Making_P
- aid the explainee in performing a task
- allow to make more accurate decsiosn
- allow to make more informed and accurate decisions
- allow users to make more informed decisions
- better decision whether to choose the recommended item
- employ for decision-making
- empower users to make informed choices
- enable users to make more accurate judgment on a decision
- enable users to make more informed and confident decisions
- enable users to make more informed decisions
- evaluate possible alternatives
- give enough information to decide
- help decide whether to engage with item further
- help facilitate improved user decisions
- help users make accurate decisions
- help users make to better decisions
- help users to evaluate items correctly
- help users to make more accurate decisions
- help users to resolve preference conflict
- helping users to make better decisions
- helping users to make good decisions
- improve the quality of user decisions
- improve users' decision making
- increased problem solving competence
- influence decision making
- make informed decisions
- might be critical to aid a person in making final decisions
- needed to make rational choices
- provide users with knowledge to make decisions
- support to make more accurate decisions
- support user decision-making

System Acceptance_P
- affect acceptance
- beneficial to the system's acceptance
- bring closer to acceptance by end users
- can enhance acceptability of systems
- can improve acceptance of recommender systems
- convince to use system
- determine the ovewrall adoption
- essential for gaining acceptance

- gain user acceptance
- greater user acceptance
- higher degree of willingness to use systems
- higher level of willingness to use autonomous systems
- important for acceptance of recommender systems
- important for user acceptance
- increase acceptance
- increase user acceptance
- increasing user acceptance
- influential for acceptance of intelligent systems
- key towards acceptance of AI systems
- may increase system acceptance
- means of influencing user acceptance
- open the chance of adoption
- pivotal to acceptance by society
- potential to increase acceptance
- promote acceptance of the system
- requirement for model acceptance
- user acceptance

Testability_P
- enables models to be tested
- help system designers to evaluate or test a system
- helpful for testing

Trade Secrets_N
- cannot protect proprietary information
- disclose proprietary knowledge
- model confidentiality
- trade secret

Transferability_P
- improve transfer learning
- support transferability
- transferability

Transparency_P
- be perceived as transparent
- can provide transparency
- contribute to transparency
- could provide transparency
- foster transparency
- fulfill the goal of transparency
- improve transparency
- improve transparency of the system
- increase the system's transparency
- increase transparency
- increased transparency
- increasing perceived transparency
- increasing system transparency
- increasing transparency of system reasoning
- make more transparent
- offer transparency
- perceived to be transparent
- provide transparency
- provide transparency of how system works
- providing system transparency
- render decisions more transparent
- support transparency
- system transparency
- transparency

Trustworthiness_P
- achieve complete trustworthiness
- create more trustable products
- essential for becoming trustworthy
- foster trustworthiness
- important to achieve trustworthy AI
- improve untrustworthy models
- increase trustworthiness
- make it more trustworthy
- trustworthiness
- trustworthy AI

Understandability_N
- create false binaries
- hinder understanding if not appropriate
- may cause information overload

Understandability_P
- ad in understanding network mistakes
- aid in understanding an automated vehicle's function
- aid in understanding the decision's consequences
- allow its users to better understand its outputs
- allow to understand the behavior of a DNN
- allow users to understand why prediction is made
- assist understanding how the system processes data
- better understand what a model has learned
- can be easily comprehended
- can increase our understanding of the models
- can strongly infuence users' understanding of complex data
- concerned with enabling human understanding
- effects on user understanding
- enable domain experts and users to understand the model
- enable human users to understand
- enable human users to understand AI systems
- enable understanding
- enhance understanding of training situations
- enrich people's understanding
- equip users to understand the system
- facilitate the understanding of the system
- get insights into predictions
- help a user to understand why an item is recommended
- help the human to understand the behavior
- help the human to understand why an agent failed
- help the user better to understand the rationale
- help user to understand agent behavior
- help user to understand why item is recommended
- help users to understand system's output
- help users understand the suggestions made
- help users understand the system's operation domain
- helpful for understanding intelligent system's reasoning process
- hint to better understanding of the job recommendations
- important for program comprehension
- important to understand how the system operates
- improve understandability of ML
- improve user understanding of the system
- improve users' understanding
- improving user understanding
- interpretability
- it can help build understanding
- key to understanding models better
- know the system's reasoning
- make decisions understandable
- make opaque algorithms more understandable
- positive effect on understandability
- provide insights into the model
- show a significant improvement in user understanding
- support intelligibility
- tell us something about the decison-making process
- tend to cause understanding
- tool to build understanding of the technology
- understand and contextualize the system strategy
- understand and evaluate the model
- understand system behavior
- understand system's reasoning
- understand the relevance feedback algorithm
- understanding
- understanding a system
- understanding how the application works
- understanding of why the outcome arose
- understanding system behavior
- way to better understand decisions
- way to interpret the decisions

Usability_N
- not every software should provide explanations
- reducing the sense of ease of use
- the focus is not on usability

Usability_P
- better utilize the AI
- can help reduce complex cognitive efforts
- crucial aspect for usability
- ease of use of the recommender system
- ensure effective interactions with ML systems
- help to improve the usability of the system
- improve usability
- increase ease of use
- leads to more effective use
- leads to more efficient use
- make better use of the system
- make effective use of system
- make it quicker and easier for users to find what they want
- reduce cognitive burden
- save user's cognitive effort
- save user's interaction efforts
- substantially affect usability
- to make a better use of a system's outputs
- usability

User Awareness_P
- can promote awareness
- increase situational awareness
- serve awareness

User Control_P
- control
- enable users to control their interaction more actively
- give greater control to the user
- human control over the decision
- meaningful human control over algorithms
- provide control
- starting point for better user control

User Effectiveness_N
- can be conflicting with effectiveness
- lead to agreement with incorrect system suggestions
- lead to automation bias when familiar with problem

User Effectiveness_P
- allow to assess whether the alternative is appropriate
- can lead to greater accuracy in decision making
- effectiveness
- improve users' decision effectiveness
- necessary for the user to carry out his or her task
- user effectiveness

User Efficiency_N
- require significant human effort
- taking more time
- users need to take time to analyze explanations

User Efficiency_P
- decreasing the time needed to reach a judgement
- help users make decisions faster
- help users make faster decisions
- reduce time for decision making
- user efficiency
- user's decision efficiency
- users need less time understanding the interface

User Experience_N
- become distracting over time
- can pollute user interface
- may have contributed to confusion or surprise
- too much explanation can create confusion

User Experience_P
- change user attitude towards the system
- improve users' experience
- influences actions, emotions and beliefs of the recipient
- lead to more positive user attitudes
- make the system more engaging
- play an important role in improving user experience
- promote feelings of familiarity
- user experience

User Performance_P
- affect test performance
- better user performance
- can increase performance on information retrieval tasks
- improve task performance
- improve user performance
- improved decision quality
- improved problem solving performance
- user performance

User Satisfaction_P
- be more comfortable
- brings enjoyment to users
- can increase user satisfaction
- can lead to higher level of satisfaction
- contributes to user satisfaction
- decrease frustration
- feel more comfortable
- has a positive effect on satisfaction
- higher degrees of satisfaction
- important for user satisfaction
- improved user satisfaction
- improving user satisfaction
- increase enjoyment
- increase satisfaction
- increase user satisfaction
- increase users' satisfaction
- increased participants' satisfaction with the recommendation
- lead to better user satisfaction
- lead to greater satisfaction
- lead to greater user satisfaction
- positive effect on satisfaction
- positive relationship with user satisfaction
- promote feelings of comfort
- satisfaction
- significantly impact satisfaction
- user satisfaction

Validation_P
- help users assess if the recommended alternative is truly adequate for them
- validate system knowledge
- way of validating the decision process of an AI
- way to verify the validity of a decision

Verifiability_P
- assess whether a prediction is accurate
- check the system by examining the way it reasons
- enable users to verify that an algorithm is accurate
- ensure that algorithms are performing as expected
- ensure that algorithms perform as expected
- evaluate correctness of system ouputs
- evaluate the goodness of a match
- evaluate the proposed solution
- help users evaluate the accuracy of the system's prediction
- increase the ability to assess whether a prediction is accurate
- verification
- verify accuracy
- verify correct functioning
- verify that the system works as it should
- verify the correctness of the knowledge base or structure

V. Workshops

A. Homework

For the workshop with requirements engineers, we designed the following hypothetical scenarios. Note that the workshops were originally in German, so the following scenario descriptions are translations.

a) Loan Issuing Software: Bank Tresor Holdings Ltd. wants to change its loan issuing software. In the past, the bank has often come under suspicion of including gender or ethnicity when determining creditworthiness. For this reason, the new system is to be equipped with a feature that will allow the employee to find out why a particular person gets the calculated credit score. The system should be able to explain to the employee what the most important factor was that contributed to that exact classification, and what the customer would have to change to get a different (better) credit score. Tresor Holdings employees are experts in their field and can work through even complicated issues in a way that customers can understand. The customers themselves do not get to see anything from the system.

b) Medical Diagnosis System: Saint Health hospital not only wants to help doctors make better judgments, but also to upgrade the doctor-patient interaction. To this end, it plans to introduce a new interactive disease prognosis system that not only reliably explains to doctors how a particular prognosis came about, but can also provide this information tailored to patients. With the help of the system, doctors will be able to better educate patients about their diagnoses so they can accept or reject a particular treatment in an informed and justified manner. Beyond being understandable to patients, of course, the system should also be able to provide doctors with additional specialist information that patients might not be able to relate to.

c) Applicant Selection Software: The company GoodJobs wants to automate most of its hiring process. Before a select few applicants get a real interview, an initial screening is to take place completely online. The final result should not only be presented to the applicants, but also made comprehensible for them. Especially the rejected applicants should be able to understand what led to their rejection and how they can improve for the next time. The recipients of the explanations should not only be the applicants, but also the legal department, which must determine whether certain rejections were justified.

d) Autonomous Driving: The car manufacturer AutoNomous GmbH is the first car manufacturer in the world to offer fully autonomous vehicles. As with normal vehicles, however, accidents do occur from time to time with autonomous vehicles, but much less frequently than with conventional automobiles driven by humans. In order to be able to resolve these rare accidents without the help of experts, an explanatory component is to be added to the software. The company can no longer afford to provide an expert to analyze and assess each accident. The software extension should make it easy for accident investigators to find out for themselves what led to the accident and who is responsible.

Acknowledgments

This work was supported by the research initiative Mobilise between the Technical University of Braunschweig and Leibniz University Hannover, funded by the Ministry for Science and Culture of Lower Saxony and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Cluster of Excellence PhoenixD (EXC 2122, Project ID 390833453). Work on this paper was also funded by the Volkswagen Foundation grant AZ 98514 "Explainable Intelligent Systems" (EIS) and by the DFG grant 389792660 as part of TRR 248.

References

[1] L. Chazette, W. Brunotte, and T. Speith, "Exploring explainability: A definition, a model, and a knowledge catalogue," in *IEEE 29th International Requirements Engineering Conference (RE)*. IEEE, 2021.

[2] A. N. Venturelli, "A cautionary contribution to the philosophy of explanation in the cognitive neurosciences," *Minds and Machines*, vol. 26, no. 3, pp. 259–285, 2016.

[3] J. Sliwinski, M. Strobel, and Y. Zick, "A characterization of monotone influence measures for data classification," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 48–52.

[4] C. Brinton, "A framework for explanation of machine learning decisions," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 14–18.

[5] S. Robbins, "A misdirected principle with a catch: Explicability for ai," *Minds and Machines*, vol. 29, no. 4, pp. 495–514, 2019.

[6] R. Pierrard, J.-P. Poli, and C. Hudelot, "A new approach for explainable multiple organ annotation with few data," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 101–107.

[7] A. I. Antón and J. B. Earp, "A requirements taxonomy for reducing web site privacy vulnerabilities," *Requirements engineering*, vol. 9, no. 3, pp. 169–185, 2004.

[8] A. Richardson and A. Rosenfeld, "A survey of interpretability and explainability in human-agent systems," in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence (XAI 2018)*, 2018, pp. 137–143.

[9] M. Hall, D. Harborne, R. Tomsett, V. Galetic, S. Quintana-Amate, A. Nottle, and A. Preece, "A systematic method to understand requirements for explainable AI (XAI) systems," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 21–27.

[10] T. Eiter, Z. G. Saribatur, and P. Schüller, "Abstraction for zooming-in to unsolvability reasons of grid-cell problems," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 7–13.

[11] M. Krishnan, "Against interpretability: A critical examination of the interpretability problem in machine learning," *Philosophy & Technology*, pp. 1–16, 2019.

[12] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, "AI4people—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.

[13] P. B. De Laat, "Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability?" *Philosophy & Technology*, vol. 31, no. 4, pp. 525–541, 2018.

[14] E. S. Dahl, "Appraising black-boxed technology: the positive prospects," *Philosophy & Technology*, vol. 31, no. 4, pp. 571–591, 2018.

[15] I. Monteath and R. Sheh, "Assisted and incremental medical diagnosis using explainable artificial intelligence," in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence (XAI 2018)*, 2018, pp. 104–108.

[16] U. Ehsan, P. Tambwekar, L. Chan, B. Harrison, and M. O. Riedl, "Automated rationale generation: a technique for explainable ai and its effects on human perceptions," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 263–274.

[17] A. Aggarwal, P. Lohia, S. Nagar, K. Dey, and D. Saha, "Black box fairness testing of machine learning models," in *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. New York, NY, USA: Association for Computing Machinery, 2019, p. 625–635. [Online]. Available: https://doi.org/10.1145/3338906.3338937

[18] K. L. Theurer, "Compositional explanatory relations and mechanistic reduction," *Minds and Machines*, vol. 23, no. 3, pp. 287–307, 2013.

[19] D. J. Buller, "Confirmation and the computational paradigm (or: Why do you think they call it artificial intelligence?)," *Minds and Machines*, vol. 3, no. 2, pp. 155–181, 1993.

[20] A. Lucic, H. Haned, and d. R. Maarten, "Contrastive explanations for large errors in retail forecasting predictions through monte carlo simulations," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 66–72.

[21] K. Sokol and P. A. Flach, "Conversational explanations of machine learning predictions through class-contrastive counterfactual statements." in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence (XAI 2018)*, 2018, pp. 5785–5786.

[22] M. L. Olson, L. Neal, F. Li, and W.-K. Wong, "Counterfactual states for atari agents via generative deep learning," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 87–93.

[23] H. A. Simon, "Discovering explanations," *Minds and Machines*, vol. 8, no. 1, pp. 7–37, Feb 1998.

[24] L. Hiley, A. Preece, Y. Hicks, D. Marshall, and H. Taylor, "Discriminating spatial and temporal relevance in deep taylor decompositions for explainable activity recognition," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 35–39.

[25] Y. Coppens, K. Efthymiadis, T. Lenaerts, A. Nowé, T. Miller, R. Weber, and D. Magazzeni, "Distilling deep reinforcement learning policies in soft decision trees," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 1–6.

[26] L. Chazette, O. Karras, and K. Schneider, "Do end-users want explanations? analyzing the role of explainability as an emerging aspect of non-functional requirements," in *IEEE 27th International Requirements Engineering Conference (RE)*, 2019, pp. 223–233.

[27] N. Tintarev and J. Masthoff, "Effective explanations of recommendations: user-centered design," in *Proceedings of the 2007 ACM conference on Recommender systems*, 2007, pp. 153–156.

[28] S. T. Mckinlay, "Evidence, explanation and predictive data modelling," *Philosophy & Technology*, vol. 30, no. 4, pp. 461–473, 2017.

[29] M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender, "Explainability as a non-functional requirement," in *IEEE 27th International Requirements Engineering Conference (RE)*, 2019, pp. 363–368.

[30] T. Miller, P. Howe, and L. Sonenberg, "Explainable AI: Beware of inmates running the asylum. or: How I learnt to stop worrying and love the social and behavioural sciences," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 36–42.

[31] P. S. Kumar, M. Saravanan, and S. Suresh, "Explainable classification using clustering in deep learning models," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 115–121.

[32] M. Fox, D. Long, and D. Magazzeni, "Explainable planning," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 24–30.

[33] P. Madumal, T. Miller, L. Sonenberg, and F. Vetere, "Explainable reinforcement learning through a causal lens," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 73–79.

[34] Z. Juozapaitis, A. Koul, A. Fern, M. Erwig, and F. Doshi-Velez, "Explainable reinforcement learning via reward decomposition," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 47–53.

[35] F. Gutiérrez, S. Charleer, R. De Croon, N. N. Htun, G. Goetschalckx, and K. Verbert, "Explaining and exploring job recommendations: a user-driven approach for interacting with knowledge-based job recommender systems," in *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 60–68.

[36] R. O. Weber, H. Hong, and P. Goel, "Explaining citation recommendations: Abstracts or full texts?" in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 136–142.

[37] N. Angius and G. Tamburrini, "Explaining engineered computing systems' behaviour: the role of abstraction and idealization," *Philosophy & Technology*, vol. 30, no. 2, pp. 239–258, 2017.

[38] J. Dodge, Q. V. Liao, Y. Zhang, R. K. Bellamy, and C. Dugan, "Explaining models: an empirical study of how explanations impact fairness judgment," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 275–285.

[39] L. Chen and F. Wang, "Explaining recommendations based on feature sentiments in product reviews," in *Proceedings of the 22nd international conference on intelligent user interfaces*, 2017, pp. 17–28.

[40] C.-H. Tsai and P. Brusilovsky, "Explaining recommendations in an interactive hybrid social recommender," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 391–396.

[41] O. Biran and C. Cotton, "Explanation and justification in machine learning: A survey," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 8–13.

[42] W. Pieters, "Explanation and trust: what to tell the user in security and ai?" *Ethics and Information Technology*, vol. 13, no. 1, pp. 53–64, 2011.

[43] A. Gopnik, "Explanation as orgasm," *Minds and machines*, vol. 8, no. 1, pp. 101–118, 1998.

[44] J. Fernández, "Explanation by computer simulation in cognitive science," *Minds and Machines*, vol. 13, no. 2, pp. 269–284, 2003.

[45] H. Johnson and P. Johnson, "Explanation facilities and interactive systems," in *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI)*. New York, NY, USA: Association for Computing Machinery, 1993, pp. 159–166.

[46] W. F. Brewer, C. A. Chinn, and A. Samarapungavan, "Explanation in scientists and children," *Minds and Machines*, vol. 8, no. 1, pp. 119–136, 1998.

[47] A. Rueger, "Explanations at multiple levels," *Minds and Machines*, vol. 11, no. 4, pp. 503–520, 2001.

[48] J. De Winter, "Explanations in software engineering: The pragmatic point of view," *Minds and Machines*, vol. 20, no. 2, pp. 277–289, 2010.

[49] N. Tintarev, "Explanations of recommendations," in *Proceedings of the 2007 ACM Conference on Recommender Systems*. New York, NY, USA: Association for Computing Machinery, 2007, pp. 203–206.

[50] J. McInerney, B. Lacker, S. Hansen, K. Higley, H. Bouchard, A. Gruson, and R. Mehrotra, "Explore, exploit, and explain: personalizing explainable recommendations with bandits," in *Proceedings of the 12th ACM conference on recommender systems*, 2018, pp. 31–39.

[51] I. Lage, D. Lifschitz, F. Doshi-Velez, and O. Amir, "Exploring computational user models for agent policy summarization," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 59–65.

[52] C. Lucero, B. Coronado, O. Hui, and D. S. Lange, "Exploring explainable artificial intelligence and autonomy through provenance," in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence (XAI 2018)*, 2018, p. 85.

[53] V. Putnam and C. Conati, "Exploring the need for explainable artificial intelligence (xai) in intelligent tutoring systems (its)." in *IUI Workshops*, vol. 19, 2019.

[54] A. J. Ko and B. A. Myers, "Extracting and answering why and why not questions about java program output," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 20, no. 2, pp. 1–36, 2010.

[55] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck, "Fair, transparent, and accountable algorithmic decision-making processes," *Philosophy & Technology*, vol. 31, no. 4, pp. 611–627, 2018.

[56] M. Veale and R. Binns, "Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data," *Big Data & Society*, vol. 4, no. 2, pp. 1–17, 2017.

[57] M. Hosseini, A. Shahri, K. Phalp, and R. Ali, "Four reference models for transparency requirements in information systems," *Requirements Engineering*, vol. 23, no. 2, pp. 251–275, 2018.

[58] G. J. Nalepa, M. van Otterlo, S. Bobek, and M. Atzmueller, "From context mediation to declarative values and explainability," in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence*

(XAI 2018), 2018.

[59] S. J. Green, P. Lamere, J. Alexander, F. Maillet, S. Kirk, J. Holt, J. Bourque, and X.-W. Mak, "Generating transparent, steerable recommendations from textual descriptions of items," in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 281–284.

[60] J. Schaffer, P. Giridhar, D. Jones, T. Höllerer, T. Abdelzaher, and J. O'donovan, "Getting the message? a study of explanation interfaces for microblog data analysis," in *Proceedings of the 20th international conference on intelligent user interfaces*, 2015, pp. 345–356.

[61] J. C.-T. Ho, "How biased is the sample? reverse engineering the ranking algorithm of facebook's graph application programming interface," *Big Data & Society*, vol. 7, no. 1, pp. 1–15, 2020.

[62] A. Papenmeier, G. Englebienne, and C. Seifert, "How model accuracy and explanation fidelity influence user trust in ai," in *Proceedings of the IJCAI 2019 Workshop on Explainable Artificial Intelligence (XAI)*, 2019, pp. 94–100.

[63] T. Barik, D. Ford, E. Murphy-Hill, and C. Parnin, "How should compilers explain problems to developers?" in *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2018, pp. 633–643.

[64] J. Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data & Society*, vol. 3, no. 1, pp. 1–12, 2016.

[65] J. Schaffer, J. O'Donovan, J. Michaelis, A. Raglin, and T. Höllerer, "I can do better than your ai: expertise and explanations," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 240–251.

[66] H. Liu, J. Wen, L. Jing, J. Yu, X. Zhang, and M. Zhang, "In2rec: Influence-based interpretable recommendation," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1803–1812.

[67] B. Ghosh, D. Malioutov, and K. S. Meel, "Interpretable classification rules in relaxed logical form," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 14–20.

[68] T. Huber, D. Schiller, and E. André, "Introducing selective layer wise relevance propagation to dueling deep q learning," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019))*, 2019.

[69] W. Bechtel, "Levels of description and explanation in cognitive science," *Minds and Machines*, vol. 4, no. 1, pp. 1–25, 1994.

[70] X. Watts and F. Lécué, "Local score dependent model explanation for time dependent covariates," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 129–135.

[71] L. Michael, "Machine coaching," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 80–86.

[72] R. McClamrock, "Marr's three levels: A re-evaluation," *Minds and Machines*, vol. 1, no. 2, pp. 185–196, 1991.

[73] M. T. Stuart and N. J. Nersessian, "Peeking inside the black box: a new kind of scientific visualization," *Minds and Machines*, vol. 29, no. 1, pp. 87–107, 2019.

[74] T. Kulesza, M. Burnett, W.-K. Wong, and S. Stumpf, "Principles of explanatory debugging to personalize interactive machine learning," in *Proceedings of the 20th international conference on intelligent user interfaces*, 2015, pp. 126–137.

[75] P. Sawyer, N. Bencomo, J. Whittle, E. Letier, and A. Finkelstein, "Requirements-aware systems: A research agenda for re for self-adaptive systems," in *2010 18th IEEE International Requirements Engineering Conference*. IEEE, 2010, pp. 95–103.

[76] A. Vogelsang and M. Borg, "Requirements engineering for machine learning: Perspectives from data scientists," in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2019, pp. 245–251.

[77] L. M. Cysneiros, M. Raffi, and J. C. S. do Prado Leite, "Software transparency as a key requirement for self-driving cars," in *2018 IEEE 26th international requirements engineering conference (RE)*. IEEE, 2018, pp. 382–387.

[78] C. Zednik, "Solving the black box problem: A normative framework for explainable artificial intelligence," *Philosophy & Technology*, pp. 1–24, 2019.

[79] R. Häuslschmid, M. von Buelow, B. Pfleging, and A. Butz, "Supporting trust in autonomous driving," in *Proceedings of the 22nd international conference on intelligent user interfaces*, 2017, pp. 319–329.

[80] J. Vig, S. Sen, and J. Riedl, "Tagsplanations: Explaining recommendations using tags," in *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI*. New York, NY, USA: Association for Computing Machinery, 2009, pp. 47–56.

[81] K. Ehrlich, S. E. Kirk, J. Patterson, J. C. Rasmussen, S. I. Ross, and D. M. Gruen, "Taking advice from intelligent systems: The double-edged sword of explanations," in *Proceedings of the 16th International Conference on Intelligent User Interfaces (IUI)*. New York, NY, USA: Association for Computing Machinery, 2011, pp. 125–134.

[82] V. Dominguez, P. Messina, I. Donoso-Guzmán, and D. Parra, "The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 408–416.

[83] C. J. Cai, J. Jongejan, and J. Holbrook, "The effects of example-based explanations in a machine learning interface," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 258–262.

[84] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, pp. 1–21, 2016.

[85] M. Zanker, "The influence of knowledgeable explanations on users' perception of a recommender system," in *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 269–272.

[86] L. Floridi, "The method of levels of abstraction," *Minds and machines*, vol. 18, no. 3, pp. 303–329, 2008.

[87] A. Páez, "The pragmatic turn in explainable artificial intelligence (XAI)," *Minds & Machines*, vol. 29, pp. 441–459, 2019.

[88] R. A. Wilson and F. Keil, "The shadows and shallows of explanation," *Minds and machines*, vol. 8, no. 1, pp. 137–159, 1998.

[89] M. T. Keane and E. M. Kenny, "The twin-system approach as one generic solution for xai: An overview of ann-cbr twins for explaining deep learning," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 54–58.

[90] A. Glass, D. L. McGuinness, and M. Wolverton, "Toward establishing trust in adaptive agents," in *Proceedings of the 13th international conference on Intelligent user interfaces*, 2008, pp. 227–236.

[91] C. Henin and L. M. Daniel, "Towards a generic framework for black-box explanation methods," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 28–34.

[92] J. Clos, N. Wiratunga, and S. Massie, "Towards explainable text classification by jointly learning lexicon and modifier terms," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 19–23.

[93] H. Wicaksono, C. Sammut, and R. Sheh, "Towards explainable tool creation by a robot," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2017)*, 2017, pp. 63–67.

[94] V. Putnam, L. Riegel, and C. Conati, "Towards personalized xai: A case study in intelligent tutoring systems," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 108–114.

[95] R. Borgo, M. Cashmore, and D. Magazzeni, "Towards providing explanations for ai planner decisions," in *Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence (XAI 2018)*, 2018.

[96] J. Zhou and F. Chen, "Towards trustworthy human-ai teaming under uncertainty," in *Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence (XAI 2019)*, 2019, pp. 143–147.

[97] J. Zerilli, A. Knott, J. Maclaurin, and C. Gavaghan, "Transparency in algorithmic and human decision-making: is there a double standard?" *Philosophy & Technology*, vol. 32, no. 4, pp. 661–683, 2019.

[98] H. Felzmann, E. F. Villaronga, C. Lutz, and A. Tamò-Larrieux, "Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns," *Big Data & Society*, vol. 6, no. 1, pp. 1–14, 2019.

[99] X. He, T. Chen, M.-Y. Kan, and X. Chen, "Trirank: Review-aware explainable recommendation by modeling aspects," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 1661–1670.

[100] P. Pu and L. Chen, "Trust building with explanation interfaces," in *Proceedings of the 11th international conference on Intelligent user interfaces*, 2006, pp. 93–100.

[101] J. Drozdal, J. Weisz, D. Wang, G. Dass, B. Yao, C. Zhao, M. Muller, L. Ju, and H. Su, "Trust in automl: Exploring information needs for establishing trust in automated machine learning systems," in *Proceedings of the 25th International Conference on Intelligent User*

Interfaces, 2020, pp. 297–307.

[102] D. Holliday, S. Wilson, and S. Stumpf, "User trust in intelligent systems: A journey over time," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 2016, pp. 164–168.

[103] N. F. Rajani and R. J. Mooney, "Using explanations to improve ensembling of visual question answering systems," in *Proceedings of the IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI)*, 2017, pp. 43–47.

[104] D. van Eck, "Validating function-based design methods: An explanationist perspective," *Philosophy & Technology*, vol. 28, no. 4, pp. 511–531, 2015.

[105] S. Jasanoff, "Virtual, visible, and actionable: Data assemblages and the sightlines of justice," *Big Data & Society*, vol. 4, no. 2, pp. 1–15, 2017.

[106] J. F. Wolfswinkel, E. Furtmueller, and C. P. M. Wilderom, "Using grounded theory as a method for rigorously reviewing literature," *European journal of Information Systems*, vol. 22, no. 1, pp. 45–55, 2013.

[107] P. Madumal, T. Miller, L. Sonenberg, and F. Vetere, "A grounded interaction protocol for explainable artificial intelligence," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. Richland, SC, USA: International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 1033–1041.

[108] D. A. Wilkenfeld, "Functional explaining: A new approach to the philosophy of explanation," *Synthese*, vol. 191, pp. 3367–3391, 2014.

[109] L. Viganò and D. Magazzeni, "Explainable security," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, 2020, pp. 293–300.

[110] M. Milkowski, "A mechanistic account of computational explanation in cognitive science," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, Eds. Cognitive Science Society, 2013, pp. 3050–3055. [Online]. Available: https://mindmodeling.org/cogsci2013/papers/0545/index.html

[111] D. Billsus and M. J. Pazzani, "A personal news agent that talks, learns and explains," in *Proceedings of the third annual conference on Autonomous Agents*, 1999, pp. 268–275.

[112] M. Harbers, K. van den Bosch, and J.-J. C. Meyer, "A study into preferred explanations of virtual agent behavior," in *International Workshop on Intelligent Virtual Agents*. Springer, 2009, pp. 132–145.

[113] M.-A. Clinciu and H. Hastie, "A survey of explainable ai terminology," in *Proceedings of the 1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI 2019)*. Association for Computational Linguistics, 2019, pp. 8–13.

[114] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2019.

[115] I. Nunes and D. Jannach, "A systematic review and taxonomy of explanations in decision support and recommender systems," *User Modeling and User-Adapted Interaction*, vol. 27, no. 3-5, pp. 393–444, 2017.

[116] G. Friedrich and M. Zanker, "A taxonomy for generating explanations in recommender systems," *AI Magazine*, vol. 32, no. 3, pp. 90–98, 2011.

[117] A. Datta, S. Sen, and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *2016 IEEE symposium on security and privacy (SP)*. IEEE, 2016, pp. 598–617.

[118] *An Evaluation of the Human-Interpretability of Explanation*, 2018.

[119] L. M. Cysneiros and V. M. B. Werneck, "An initial analysis on how software transparency and trust influence each other." in *WER*. Citeseer, 2009.

[120] K. Čyras, D. Letsios, R. Misener, and F. Toni, "Argumentation for explainable scheduling," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 2752–2759, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/4126

[121] K. Darlington, "Aspects of intelligent systems explanation," *Universal Journal of Control and Automation*, vol. 1, no. 2, pp. 40–51, 2013.

[122] B. Y. Lim and A. K. Dey, "Assessing demand for intelligibility in context-aware applications," in *Proceedings of the 11th international conference on Ubiquitous computing*, 2009, pp. 195–204.

[123] E. S. Vorm, "Assessing demand for transparency in intelligent systems using machine learning," in *2018 Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, 2018, pp. 1–7.

[124] Y. Fukuchi, M. Osawa, H. Yamakawa, and M. Imai, "Autonomous self-explanation of behavior for interactive reinforcement learning agents," in *Proceedings of the 5th International Conference on Human Agent Interaction*, 2017, pp. 97–101.

[125] Z. Zeng, C. Miao, C. Leung, and J. J. Chin, "Building more explainable artificial intelligence with argumentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[126] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, pp. 1–13, 2019.

[127] J. Y. Halpern and J. Pearl, "Causes and explanations: A structural-model approach. part ii: Explanations," *The British Journal for the Philosophy of Science*, vol. 56, no. 4, pp. 889–911, 2005. [Online]. Available: https://doi.org/10.1093/bjps/axi148

[128] T. Laugel, M.-J. Lesot, C. Marsala, X. Renard, and M. Detyniecki, "Comparison-based inverse classification for interpretability in machine learning," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2018, pp. 100–111.

[129] A. A. Freitas, "Comprehensible classification models: A position paper," *SIGKDD Explorations Newsletter*, vol. 15, no. 1, pp. 1–10, 2014.

[130] M. Sato, K. Nagatani, T. Sonoda, Q. Zhang, and T. Ohkuma, "Context style explanation for recommender systems," *Journal of Information Processing*, vol. 27, pp. 720–729, 2019.

[131] R. M. Byrne, "Counterfactuals in explainable artificial intelligence (xai): Evidence from human reasoning," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, 2019, pp. 6276–6282.

[132] M. Hosseini, A. Shahri, K. Phalp, and R. Ali, "Crowdsourcing transparency requirements through structured feedback and social adaptation," in *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2016, pp. 1–6.

[133] A. J. Ko and B. A. Myers, "Debugging reinvented: Asking and answering why and why not questions about program behavior," in *Proceedings of the 30th International Conference on Software Engineering*, ser. ICSE '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 301–310. [Online]. Available: https://doi.org/10.1145/1368088.1368130

[134] R. Sheh and I. Monteath, "Defining explainable ai for requirements analysis," *KI-Künstliche Intelligenz*, vol. 32, no. 4, pp. 261–266, 2018.

[135] U. Chajewska and J. Y. Halpern, "Defining explanation in probabilistic systems," in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997, p. 62–71.

[136] N. Tintarev and J. Masthoff, "Designing and evaluating explanations for recommender systems," in *Recommender systems handbook*. Springer, 2011, pp. 479–510.

[137] T. Hirsch, K. Merced, S. Narayanan, Z. E. Imel, and D. C. Atkins, "Designing contestability: Interaction design, machine learning, and mental health," in *Proceedings of the 2017 Conference on Designing Interactive Systems*, ser. DIS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 95–99.

[138] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable ai," in *Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–15.

[139] D. Walton, "Dialogical models of explanation." *ExaCt*, vol. 2007, pp. 1–9, 2007.

[140] R. Sheh, "Different xai for different hri," in *AAAI Fall Symposium*. AAAI Press, 2017, pp. 114–117.

[141] M. Ter Hoeve, M. Heruer, D. Odijk, A. Schuth, and M. de Rijke, "Do news consumers want explanations for personalized news rankings," in *FATREC Workshop on Responsible Recommendation Proceedings*, 2017.

[142] D. M. Truxillo, T. E. Bodner, M. Bertolino, T. N. Bauer, and C. A. Yonce, "Effects of explanations on applicant reactions: A meta-analytic review," *International Journal of Selection and Assessment*, vol. 17, no. 4, pp. 346–361, 2009.

[143] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a "right to explanation"," *AI magazine*, vol. 38, no. 3, pp. 50–57, 2017.

[144] N. Tintarev and J. Masthoff, "Evaluating the effectiveness of explanations for recommender systems," *User Modeling and User-Adapted Interaction*, vol. 22, no. 4-5, pp. 399–439, 2012.

[145] K. Sokol and P. A. Flach, "Explainability fact sheets: A framework for systematic assessment of explainable approaches," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 56–67.

[146] A. Rosenfeld and A. Richardson, "Explainability in human–agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 33, no. 6, pp. 673–705, 2019.

[147] S. Anjomshoae, A. Najjar, D. Calvaresi, and K. Främling, "Explainable agents and robots: Results from a systematic literature review," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. Richland, SC, USA: International Foundation for Autonomous Agents and Multiagent Systems, 2019, p. 1078–1088.

[148] J. Zhu, A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood, "Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation," in *IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2018, pp. 1–8.

[149] F. Kamiran and I. Žliobaitė, "Explainable and non-explainable discrimination in classification," in *Discrimination and Privacy in the Information Society*. Springer, 2013, pp. 155–170.

[150] S. M. Mathews, "Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review," in *Intelligent Computing – Proceedings of the Computing Conference*. Springer, 2019, pp. 1269–1292.

[151] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[152] N. Wang, H. Wang, Y. Jia, and Y. Yin, "Explainable recommendation via multi-task learning in opinionated text data," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 165–174.

[153] H. K. Dam, T. Tran, and A. Ghose, "Explainable software analytics," in *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 53–56.

[154] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW)*. New York, NY, USA: Association for Computing Machinery, 2000, pp. 241–250.

[155] H.-F. Cheng, R. Wang, Z. Zhang, F. O'Connell, T. Gray, F. M. Harper, and H. Zhu, "Explaining decision-making algorithms through ui: Strategies to help non-expert stakeholders," in *Proceedings of the 2019 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12.

[156] R. R. Hoffman, G. Klein, and S. T. Mueller, "Explaining explanation for "explainable ai"," vol. 62, no. 1, 2018, pp. 197–201.

[157] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *IEEE 5th International Conference on Data Science and Advanced Analytics DSAA*, 2018, pp. 80–89.

[158] B. D. Mittelstadt, C. Russell, and S. Wachter, "Explaining explanations in AI," in *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 279–288.

[159] L. H. Gilpin, C. Testart, N. Fruchter, and J. Adebayo, "Explaining explanations to society," in *NIPS Workshop on Ethical, Social and Governance Issues in AI*, 2018, pp. 1–6.

[160] M. Bilgic and R. J. Mooney, "Explaining recommendations: Satisfaction vs. promotion," in *Beyond Personalization Workshop, IUI*, vol. 5, 2005, p. 153.

[161] K. Cotter, J. Cho, and E. Rader, "Explaining the news feed algorithm: An analysis of the "news feed fyi" blog," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA)*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 1553–1560.

[162] T. Lombrozo and N. Z. Gwynne, "Explanation and inference: Mechanistic and functional explanations guide property generalization," *Frontiers in Human Neuroscience*, vol. 8, p. 700, 2014.

[163] K. McCain, "Explanation and the nature of scientific knowledge," *Science & Education*, vol. 24, no. 7, pp. 827–854, 2015.

[164] F. C. Keil, "Explanation and understanding," *Annual Review of Psychology*, vol. 57, no. 1, pp. 227–254, 2006.

[165] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[166] F. Sørmo, J. Cassens, and A. Aamodt, "Explanation in case-based reasoning–perspectives and goals," *Artificial Intelligence Review*, vol. 24, no. 2, pp. 109–143, 2005.

[167] G. Ras, v. G. Marcel, and P. Haselager, "Explanation methods in deep learning: Users, values, concerns and challenges," in *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Springer, 2018, pp. 19–36.

[168] E. I. Sklar and M. Q. Azhar, "Explanation through argumentation," in *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 277–285.

[169] D. Walton, "Explanations and arguments based on practical reasoning." in *ExaCt*, 2009, pp. 72–83.

[170] E. Rader, K. Cotter, and J. Cho, "Explanations as mechanisms for supporting algorithmic transparency," in *Proceedings of the 2018 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–13.

[171] S. Gregor and I. Benbasat, "Explanations from intelligent systems: Theoretical foundations and implications for practice," *MIS Quarterly*, vol. 23, no. 4, pp. 497–530, 1999.

[172] S. Anjomshoae, K. Främling, and A. Najjar, "Explanations of black-box model predictions by contextual importance and utility," in *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*. Springer, 2019, pp. 95–109.

[173] E. Weber, J. Van Bouwel, and R. Vanderbeeken, "Forms of causal explanation," *Foundations of Science*, vol. 10, no. 4, pp. 437–454, 2005.

[174] F. Hohman, A. Head, R. Caruana, R. DeLine, and S. M. Drucker, "Gamut: A design probe to understand how data scientists understand machine learning models," in *Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–13.

[175] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell, "Generating visual explanations," in *European Conference on Computer Vision*. Springer, 2016, pp. 3–19.

[176] R. Sevastjanova, F. Beck, B. Ell, C. Turkay, R. Henkin, M. Butt, D. A. Keim, and M. El-Assady, "Going beyond visualization: Verbalization as complementary medium to explain machine learning models," in *VIS Workshop on Visualization for AI Explainability (VISxAI)*, 2018, pp. 1–6.

[177] S. Sreedharan, T. Chakraborti, and S. Kambhampati, "Handling model uncertainty and multiplicity in explanations via model reconciliation," *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 28, no. 1, 2018.

[178] R. F. Kizilcec, "How much information? effects of transparency on trust in an algorithmic interface," in *Proceedings of the 2016 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 2390–2395.

[179] M. M. De Graaf and B. F. Malle, "How people explain action (and autonomous intelligent systems should too)," in *2017 AAAI Fall Symposium Series*, 2017.

[180] J. Hois, D. Theofanou-Fuelbier, and A. J. Junk, "How to achieve explainability and transparency in human ai interaction," in *International Conference on Human-Computer Interaction (HCI)*. Springer, 2019, pp. 177–183.

[181] M. O. Riedl, "Human-centered artificial intelligence and machine learning," *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 33–36, 2019.

[182] O. Biran and K. McKeown, "Human-centric justification of machine learning predictions," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, p. 1461–1467.

[183] B. Lamche, U. Adıgüzel, and W. Wörndl, "Interactive explanations in mobile shopping recommender systems," in *Joint Workshop on Interfaces and Human Decision Making in Recommender Systems*,

vol. 14, 2014.
[184] R. Binns, V. K. Max, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'it's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions," in *Proceedings of the 2018 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–14.
[185] J. Chen, F. Lecue, J. Z. Pan, I. Horrocks, and H. Chen, "'knowledge-based transfer learning explanation," in *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*, 2018, pp. 349–358.
[186] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, 2019.
[187] B. P. Knijnenburg and A. Kobsa, "Making decisions about privacy: information disclosure in context-aware recommender systems," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 3, no. 3, pp. 1–23, 2013.
[188] K. Sokol and P. Flach, "One explanation does not fit all," *KI-Künstliche Intelligenz*, vol. 34, no. 2, pp. 235–250, 2020.
[189] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018.
[190] J. Schneider and J. P. Handali, "Personalized explanation for machine learning: A conceptualization," in *27th European Conference on Information Systems (ECIS)*, 2019.
[191] J. Zhou, H. Hu, Z. Li, K. Yu, and F. Chen, "Physiological indicators for user trust in machine learning with influence enhanced fact-checking," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2019, pp. 94–113.
[192] T. Chakraborti, S. Sreedharan, S. Grover, and S. Kambhampati, "Plan explanations as model reconciliation," in *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 258–266.
[193] M. K. Lee, A. Jain, H. J. Cha, S. Ojha, and D. Kusbit, "Procedural justice in algorithmic fairness," *Proceedings of the 2019 ACM on Human-Computer Interaction*, vol. 3, pp. 1–26, 2019.
[194] S. Nagulendra and J. Vassileva, "Providing awareness, explanation and control of personalized filtering in a social networking site," *Information Systems Frontiers*, vol. 18, no. 1, pp. 145–158, 2016.
[195] M. J. Druzdzel, "Qualitative verbal explanations in bayesian belief networks," *AISB QUARTERLY*, pp. 43–54, 1996.
[196] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the ai: informing design practices for explainable ai user experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15.
[197] U. Ehsan, B. Harrison, L. Chan, and M. O. Riedl, "Rationalization: A neural machine translation approach to generating natural language explanations," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 81–87.
[198] M. R. Wick and W. B. Thompson, "Reconstructive expert system explanation," *Artificial Intelligence*, vol. 54, no. 1-2, pp. 33–70, 1992.
[199] J. Van Bouwel and E. Weber, "Remote causes, bad explanations?" *Journal for the Theory of Social Behaviour*, vol. 32, no. 4, pp. 437–449, 2002.
[200] O. Zinovatna and L. M. Cysneiros, "Reusing knowledge on delivering privacy and transparency together," in *2015 IEEE fifth international workshop on requirements patterns (RePa)*. IEEE, 2015, pp. 17–24.
[201] J. F. Osborne and A. Patterson, "Scientific argument and explanation: A necessary distinction?" *Science Education*, vol. 95, no. 4, pp. 627–638, 2011.
[202] J. D. Trout, "Scientific explanation and the sense of understanding," *Philosophy of Science*, vol. 69, no. 2, pp. 212–233, 2002.
[203] L. Edwards and M. Veale, "Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for," *Duke L. & Tech. Rev.*, vol. 16, p. 18, 2017.
[204] I. Baaj, J.-P. Poli, and W. Ouerdane, "Some insights towards a unified semantic representation of explanation for explainable artificial intelligence," in *Proceedings of the 2019 Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI)*. Association for Computational Linguistics, 2019, pp. 14–19.
[205] M. Hind, D. Wei, M. Campbell, N. C. F. Codella, A. Dhurandhar, A. Mojsilović, K. N. Ramamurthy, and K. R. Varshney, "Ted: Teaching ai to explain its decisions," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 123–129.
[206] M. Strevens, "The causal and unification approaches to explanation unified—causally," *Noûs*, vol. 38, no. 1, pp. 154–176, 2004.
[207] H. Cramer, V. Evers, S. Ramlal, V. S. Maarten, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga, "The effects of transparency on trust in and acceptance of a content-based art recommender," *User Modeling and User-adapted interaction*, vol. 18, no. 5, p. 455, 2008.
[208] F. Nothdurft, T. Heinroth, and W. Minker, "The impact of explanation dialogues on human-computer trust," in *International Conference on Human-Computer Interaction (HCI)*. Springer, 2013, pp. 59–67.
[209] L. R. Ye and P. E. Johnson, "The impact of explanation facilities on user acceptance of expert systems advice," *Mis Quarterly*, pp. 157–172, 1995.
[210] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural computing and applications*, pp. 1–15, 2019.
[211] T. Lombrozo, "The instrumental value of explanations," *Philosophy Compass*, vol. 6, no. 8, pp. 539–551, 2011.
[212] J. Trout, "The psychology of scientific explanation," *Philosophy Compass*, vol. 2, no. 3, pp. 564–591, 2007.
[213] A. Bussone, S. Stumpf, and D. O'Sullivan, "The role of explanations on trust and reliance in clinical decision support systems," in *2015 international conference on healthcare informatics*. IEEE, 2015, pp. 160–169.
[214] K. McCarthy, J. Reilly, L. McGinty, and B. Smyth, "Thinking positively-explanatory feedback for conversational recommender systems," in *Proceedings of the European Conference on Case-Based Reasoning (ECCBR-04) Explanation Workshop*, 2004, pp. 115–124.
[215] T.-W. Chen and S. S. Sundar, "This app would like to use your current location to better serve you: Importance of user assent and system transparency in personalized mobile services," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13.
[216] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W.-K. Wong, "Too much, too little, or just right? Ways explanations impact end users' mental models," in *IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 2013, pp. 3–10.
[217] M. Sridharan and B. Meadows, "Towards a theory of explanations for human–robot collaboration," *KI-Künstliche Intelligenz*, vol. 33, no. 4, pp. 331–342, 2019.
[218] V. Dominguez, P. Messina, C. Trattner, and D. Parra, "Towards explanations for visual recommender systems of artistic images," in *Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*, 2018.
[219] M. Atzmueller, "Towards socio-technical design of explicative systems: Transparent, interpretable and explainable analytics and its perspectives in social interaction contexts information," in *Proceedings of the 2019 Workshop on Affective Computing and Context Awareness in Ambient Intelligence (AfCAI)*, 2019, pp. 1–8.
[220] R. Iyer, Y. Li, H. Li, M. Lewis, R. Sundar, and K. Sycara, "Transparency and explanation in deep reinforcement learning neural networks," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 144–150.
[221] K. Balog, F. Radlinski, and S. Arakelyan, "Transparent, scrutable and explainable user models for personalized recommendation," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 265–274.
[222] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, "Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda," in *Proceedings of the 2018 Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–18.
[223] P. Pu and L. Chen, "Trust-inspiring explanation interfaces for recommender systems," *Knowledge-Based Systems*, vol. 20, no. 6, pp. 542–556, 2007.
[224] L. Chen, D. Yan, and F. Wang, "User evaluations on sentiment-based recommendation explanations," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 9, no. 4, pp. 1–38, 2019.
[225] M. A. Neerincx, J. van der Waa, F. Kaptein, and J. van Diggelen, "Using perceptual and cognitive explanations for enhanced human-agent team performance," in *International Conference on Engineering*

Psychology and Cognitive Ergonomics. Springer, 2018, pp. 204–214.

[226] J. Bidot, S. Biundo-Stephan, T. Heinroth, W. Minker, F. Nothdurft, and B. Schattenberg, "Verbal plan explanations for hybrid planning," in *MKWI*, 2010.

[227] S. Rosenthal, S. P. Selvaraj, and M. Veloso, "Verbalization: Narration of autonomous robot experience," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, ser. IJCAI'16. AAAI Press, 2016, p. 862–868.

[228] D. Doran, S. Schulz, and T. R. Besold, "What does explainable ai really mean? a new conceptualization of perspectives," in *Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML (CEX 2017)*, 2017.

[229] R. Sheh, "" why did you do that?" explainable intelligent robots," in *AAAI Workshop-Technical Report*, 2017, pp. 628–634.

[230] T. Kulesza, S. Stumpf, W.-K. Wong, M. M. Burnett, S. Perona, A. Ko, and I. Oberst, "Why-oriented end-user debugging of naive bayes text classification," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 1, no. 1, pp. 1–31, 2011.

[231] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 1135–1144.