

A Survey of Human-Centered Evaluations in Human-Centered Machine Learning: Supplementary Material

Fabian Sperrle¹, Mennatallah El-Assady¹, Grace Guo², Rita Borgo³, Duen Horng Chau², Alex Endert², and Daniel Keim¹

¹University of Konstanz ²Georgia Institute of Technology ³King's College London

1. Definitions of Coding Values





In the STAR, we provide definitions for all dimensions in the respective boxes. Here, we additionally provide definitions or clarifications for all coding values used to describe the aspects *evaluation setup* (Table 1), *models and explanations* (Table 2), *interaction* (Table 3) and *results* (Table 4).

2. Topic Model Backbone

The topic model backbone used to derive the IHTM of the results dimensions is shown in Table 5.

3. Coding Results

We provide our coding results in Table 7. For brevity, the annotation values have been shortened. For all relevant dimensions, the abbreviations are defined in Table 6. The free text results dimensions have been omitted due to the space requirements. For impact, age, and gender, ✓ encodes that the respective values were reported and have been coded.

Little squares indicate that a variable was *measured* , constituted a *condition* , or a combination thereof . We use  to encode that a property was motivated but not evaluated.

Dimension	Value(s)	Definition
Study Type	crowdsourcing	Feedback for the system (or more commonly, a smaller system component) is crowdsourced.
	pair analytics	Participants can be supported by a researcher (typically the study leader) answering questions, giving hints, helping with technical difficulties, and taking over system interactions [KF14]
	lab experiment	Systems are evaluated in a controlled lab environment, with defined research questions, hypotheses, conditions, and between- or within-subjects analysis.
	observation study	Systems are evaluated by observing how participants, usually experts, interact with them. Usually more open-ended and exploratory than lab experiments.
Result Processing	qualitative	Qualitative data (questionnaires, study protocols, etc.) was collected and analyzed.
	quantitative	Quantitative data (e.g., measurements) was collected and analyzed.
	both	Both qualitative and quantitative data were collected and analyzed.
Learning Phase	unguided exploration	Before starting the task in the study, participants were given time to freely explore the system
	training	Participants were trained to use the system, often by completing a systematic set of tasks
	walkthrough	Participants were familiarized with the system by watching a demonstration
	none	Participants were not given any opportunity to learn the system
Time Needed	number(s)	The time needed by participants, often specified as average study session length.
Domain/Dataset Expertise	low	Participants have rarely or never worked with the dataset or application domain.
	mid	Participants have some experience working with the dataset or application domain.
	high	Participants are experts of the dataset or application domain.
	study condition	Participants' expertise was varied as a study condition.
ML/AI Expertise	low	Participants have rarely or never worked with machine learning or AI.
	mid	Participants have some experience working with machine learning or AI.
	high	Participants are experts of machine learning or AI.
	study condition	Participants' expertise was varied as a study condition.
Background	free comment	Any information provided regarding participant background, like the domain of expertise, job titles, years of experience, etc.
Age	number(s)	The age range of all participants, or the average age if only one value is reported in the paper.
Gender	free comment	Any information reported about participants' gender, usually as counts of male, female, and other.
Analysis Task	diagnose	Diagnosing model issues and errors, typically during training.
	refine	Improving and refining model performance.
	understand	Understanding how the model works, or why certain decisions were made.
	explore	Exploring the model and results, can be related to understanding the model.
	hypothesize	Generating hypotheses about model behavior, often followed by testing the hypothesis.
	use	General usage of the system proposed, typically not with a training or refinement intent.
	compare	Comparing different model behaviours or model alternatives.
	justify	Justifying or explaining why the model behaves in a certain way.
Data Types	train	Tasks related to training the model
	text	Text data
	time series	Temporal data
	geo	Geographic data
	images	Image data
	video	Video data
	multivariate	Multivariate data

Table 1: Definitions of values used for evaluation setup dimensions during coding.

Dimension	Value(s)	Definition
Quality	measured	The quality dimension was a dependent variable.
	study condition	The quality dimension was a controlled variable.
	measured condition	The quality dimension was controlled variable that was also measured.
	motivated	The paper was motivated to make an impact on the quality dimension, but no evaluation was conducted.
	N/A	The paper did not discuss the quality dimension
Perceived Quality		
Transparency (model)		
Trustworthiness (model)	see Quality	Values and definitions for all model dimensions are identical to the quality dimension.
Interpretability		
Controllability		
Transparency (explanation)		
Trustworthiness (explanation)	see Quality	Values and definitions for all explanation dimensions are identical to the Quality dimension.
Effectiveness		
Fidelity		

Table 2: Definitions of values used for ML/AI models dimensions during coding.

Dimension	Value(s)	Definition
Direct/Indirect	direct	The system supports interaction in the form of direct manipulation.
	indirect	The system supports interaction in the form of indirect manipulation or semantic interaction.
	both	The system supports both direct and indirect interaction.
Interaction Type	free comment	Any information reported about how participants interact with the underlying machine learning model.
Impact	free comment	Any information reported about how interactions affect the model and other results (e.g., perceived model or explanation quality dimensions) from the study.
Time/Phase	data selection	Users interact with the system during the data selection phase before model training starts.
	data preprocessing	Users interact with the system during data preprocessing before model training starts.
	training	Users interact with the system during model training.
	post training	Users interact with the system during the post training or model refinement phase.
Frequency	throughout	User input drives the work of the system, and user interaction is required throughout.
	on-demand	User interaction is not required to drive the analysis process or model adaptation. Users work with the system without frequently interacting directly with the underlying machine learning model.
Degree (guidance)	orienting	
	directing	All guidance degrees are as defined in [CGM*17]
	prescribing	
Knowledge Gap (guidance)	data	
	task	
	VA method	All guidance knowledge gap dimensions are as defined in [CGM*17]
	user	
Adaptation (guidance)	infrastructure	
	content	The system adapts the provided guidance in terms of content (e.g., based on semantic similarity after relevance feedback).
	context	The system adapts in which analysis contexts or situations it provides guidance.
	timing	The system adapts the timing of the provided guidance (e.g., updating cooldown periods or inactivity timers).

Table 3: Definitions of values used for interaction dimensions during coding.

Dimension	Value(s)	Definition
Main HCML Finding	free comment	Any information provided regarding the main finding concerning HCML
User Interface Feedback	free comment	Any information provided regarding user feedback for the user interface
Interaction Feedback	free comment	Any information provided regarding user feedback for the interaction design

Table 4: Definitions of values used for results dimensions during coding.

Backbone	Keywords
#1	quality, observed quality, accuracy
#2	transparency, transparent
#3	trustworthiness, trust, trusted
#4	interpretability, interpretation, interpret
#5	controllability, control, steer
#6	explain, explanation, explainable
#7	filtering, selection, filter, select
#8	debug, debugging,
#9	interpret, interpretation
#10	explore, exploration, investigation, investigate
#11	understand, understanding, comprehend, comprehension
#12	refine, refinement, optimization, optimize, tune, tuning, fine-tune, fine-tuning

Table 5: The backbone topics used to prime the ITHM. Each topic consists only of a few keywords. Each backbone topic is only a suggestion to the algorithm allowing the encoding of domain knowledge and does not force the inclusion of the topic in the final result.

Dimension	Shortcuts
Domain/Dataset Expertise	1 = low, 2 = mid, 3 = high
ML/AI Expertise	1 = low, 2 = mid, 3 = high
Analysis Task	D = Diagnose, R = Refine, U = Understand, E = Explore, H = Hypothesize, Use = Use, C = Compare, J = Justify, T = Train
Data Types	Text = Text Data, Geo = Geo, Img = Images, Vid = Video, MVD = Multivariate Data
Study Type	P = Pair Analytics, LTO = long term observation study, O = observational study, L = lab experiment, F = field study, C = crowdsourcing
Result Processing	QL = qualitative, QN = quantitative, B = both
Learning Phase	U = unguided exploration, T = training, W = walkthrough, N = none
Direct / Indirect	D = direct, I = indirect, N/A
Time/Phase	DS = data selection, DP = data preprocessing, T = training, PT = post training/refinement
Frequency	T = throughout, OD = on demand, N/A
Degree	O = orienting, D = directing, P = prescribing, N/A
Knowledge Gap	D = Data, T = Task, VA = VA methods, U = user, I = infrastructure
Adaptation	C = suggestion content, CX = context, T = timing

Table 6: The abbreviations used in the coding results.

Table 7: Overview of the coding results.

Publication	Evaluation Setup				ML/AI Models							Interaction																
	Domain Expertise	ML/AI Expertise	Age	Gender	#Participants	Analysis Task	Data Types	Study Type	Result Processing	Learning Phase	Time Needed	Quality	Perceived Quality	Transparency	Trustworthiness	Interpretability	Controllability	Transparency	Trustworthiness	Effectiveness	Fidelity	Direct / Indirect	Impact	Time/Phase	Frequency	Degree	Knowledge Gap	Adaptation
[HKBE12]	2				12	Use,T	Text	O B N				■	■		■		□					D		PT	T			
[dSBD*12]					1	D	MVD	L B N				■	■									D	✓	PT	T			
[MP13]	3				11	H	MVD	O QL			2M											D		T	T			
[BAL*15]		2		✓	20	D,R,C,U	Text	L B U,W				■						■				D,I	✓	T,PT	T			
[KPN16]	3	3			5	U, E, D	MVD	O QL N			4M											D		PT	OD			
[SMD*16]	■		✓			Use	Vid	C QL T				■										D		DP	T	D	D	C
[SDMT16]		2	✓	✓	6	E,C, Use	MVD	L QL U			45											D		PT	T			
[SDMT16]		1	✓	✓	4	E, C, Use	MVD	L QL U			45											D		PT	T			
[ARO*17]	3					Use	Text	O QL N														I	✓	T	T			
[DLW*17]	1	1			34	C, Use, R	MVD	L QN				■	■		■			■	■			I	✓	PT	T			
[LSL*17]	3	3			2	U,R	Img	O QL N														D		PT				
[MCZ*17]		3			4	U,E,D	Text	O QL				□			□	□						D		PT	OD			
[RAL*17]		2			24	U	MVD	L B T			21.41	■	■											PT	T			
[RAL*17]		2			24	U	MVD	L B T			.17	■	■											PT	T			
[BHZ*18]	1	2	✓	✓	16	Use	MVD	L QN W				■										D,I	✓	DP	T	O	D	
[CVL*18]	3				26	Use, R	Text	C B T				■	■	■	■	■		■	■	■		D		PT	T	O,D	D	C
[ESKC18]	■	■			5	E,T,C,U,Use	Text	P QL W			120	■	■	□	■	■		□	■			D		T	T		VA	
[ESS*18]	3	■			6	R	Text	P B W			30	■	□	□	□	□						I		T	T			
[EKSK18]			✓	✓	32	U	MVD	O QL T							□	■		■	■	□		I		PT	OD			
[KAC18]	3	3		✓	3	U,E,D	MVD	O QL W			60	□			□	□						D		PT	OD	O	D	
[KEV*18]	3	3			7	Use,C,J	MVD	O QL N				□			□	□						I	✓	T	T	D	D	
[LLS*18]		3			1	D	Img	O QL N																PT	OD			
[LSC*18]	3	3			2	U,D,R	Img	O QL				■										I		PT	OD			
[LXL*18]		3			9	E,R	MVD	O B				■										I	✓	PT	OD			
[MLMP18]	2	3	✓	✓	4	R, C	MVD	O QL				■										I	✓	PT	OD	D	T	
[SKB*18]	3				1	J	Text	P QL N				□							■			D		T	T	O	D, T, VA	CX
[WGYS18]		3			3	E,R	Img	O QL W				□			□							D,I		PT	OD			
[WSW*18]		3			8	E,U	MVD	C QL N							□	□		■				I		PT	OD			
[CRH*19]	3				12	R,H,D	Img	L B W			150			□	■	□						D,I		T	T			
[CHH*19]	3				9	E,C,T		L QL W			1440											I	✓	T,PT	T			
[CD19]	2		✓	✓	12	E	MVD	L B T			25																	
[CWZ*19]		■			199	U	MVD	C QN N			20				□	■		■				I	✓	PT	OD			
[DVH*19]	3					E, R, U, Use	MVD	O QL T			60											D			VA	O	VA	
[ESD*19]	3					R	Text	P QL N				■	□	□	□							D		T		O	T	
[GLC*19]	2		✓	✓	84	Use	Vid	L QN W			30	■	■	■								D,I		T	T	P	T	C
[GLC*19]	2		✓	✓	14	Use	Vid	P B W				■	■	■				■				D,I		T	T	P	T	C
[HOW*19]			✓	✓	30	U, R	MVD	L QL W														D	✓	T	T			
[HHC*19]		2		✓	12	H,U,J,R	MVD	O QL W			90			■	■			■				D,I	✓	PT	OD			
[JSR*19]	3				2	U,R,E	Text	O QL N														I		T,PT	OD	O	T	CX
[KAY*19]		1	✓	✓	40	H,J	Vid	L B W														D,I		PT	T			
[LLL*19]	3	■			5	E	Text	L QL W,U														D	✓	PT	OD			
[MQB19]	1	2	✓		9	U,H	MVD	L B U,W							□	□						I		PT	OD			
[PZDD19]			✓		16	E	Geo	L B W			∞	■	■									D		PT	OD			
[SSKE19]	3				5	Use	Text	O QL W			90	□										D,I	✓	T	T	O	T	C
[SSBC19]	3				6	E,R,Use	Text	L QL W			40			■								D		PT	OD			
[WGSY19]		3			3	U,C,H,E	Img	O B				□			□	■						D,I		T,PT	OD			

Continued on next page

Table 7 – continued from previous page

Publication	Evaluation Setup				ML/AI Models										Interaction													
	Domain Expertise	ML/AI Expertise	Age	Gender	#Participants	Analysis Task	Data Types	Study Type	Result Processing	Learning Phase	Time Needed	Quality	Perceived Quality	Transparency	Trustworthiness	Interpretability	Controllability	Transparency	Trustworthiness	Effectiveness	Fidelity	Direct / Indirect	Impact	Time/Phase	Frequency	Degree	Knowledge Gap	Adaptation
[WGZ*19]		3			4	J,U	Img	O	QL													D		PT	OD			
[WMJ*19]		2	✓	✓	13	U,R,C	MVD	L	QN	W	40			□	□		□					I		T	T	O		
[XXM*19]	3	3			2	E	MVD	L	QL	T	90											I	✓	PT	T			
[ZWLC19]		2	✓	✓	10	U,D,J	MVD	L	B	W			□	□		□						D		PT	OD			
[LGM*20]	3				6	Use,D	MVD	O	QL	N	52											I		PT	OD	D		
[BSP20]	1	1	✓	✓	12	E, U,C, Use	MVD	O	QL	T	30						□			□		D				O	D,VA	
[BSP20]	3	3	✓		1	E, U, C, D, Use	MVD	O	QL	W							□	□	□	□	□	D				O	D,VA	
[CYL*20]	3	3			4	U	Img	O	QL	N							□					D,I		PT	OD	O	D	
[CMQ20]	2	2			3	E	MVD	L	QL	W	50											D		PT				
[DSKE20]	3	2	✓		6	E,H	MVD	O	QL	U,W	40			□		■	■					D,I		T	T	D	D	C
[EKC*20]	3				6	R	Text	P	B	W	30	■					□	□				D	✓	T	T	P	T	
[GZL*20]	3	3			4	Use, D, R, E, U	MVD	O	QL	U			□	□	□	□	□	□	□	□	□	D		PT	T	O	D	C
[KBJ*20]	3				5	R	Img	P	QL	N							□					D		T				
[KAS*20]					9	Use,U	MVD	P	QL	N			□	■	□	■						D		T	OD			
[LPH*20]	3				8	T,J,Use	MVD	O	B	T	90	■	□				□		□			D		T	T			
[MXLM20]		3			4	D	MVD	L	QL	W	90											D		PT	OD			
[MXC*20]	3	3			3	R, C	Text	L	QL	W												D	✓	T	T			
[PNKC20]			✓		3	R, U, D	MVD	O	QL	U	60	■	■	□			□			□		I	✓	T	OD			
[SFB*20]			✓	✓	180	D, R	Text	C	B	N	22.6				■	■	■		■	■	■	D	✓	PT	T			
[SSSE20]			✓		9	U, D, R	Img	P	QL	W	60													PT				
[SLC*20]	1	1	✓	✓	8	Use	MVD	L	QN	W	30	□					□	□				D,I		T	T			
[WBL*20]		3			3	D	MVD	L	QL	W	180				□		□					D		DS,DP	T			
[XMT*20]	3				4	E,D,Use	MVD	O	QL	U	90						□					D		PT				
[XCK*20]	2				6	E,U,C	Img	O	QN	N			□	□	■							I		PT	OD	O	D	
[YGLR20]		2	✓		30	E,C,U	MVD	C	B	N	50						■					D	✓	DS, T	OD	D,O	T	

References

- [ARO*17] ABDUL-RAHMAN, A., ROE, G., OLSEN, M., GLADSTONE, C., WHALING, R., CRONK, N., MORRISSEY, R., and CHEN, M. "Constructive Visual Analytics for Text Similarity Detection". en. *Computer Graphics Forum* 36.1 (2017), 237–248. DOI: [10.1111/cgf.12798](https://doi.org/10.1111/cgf.12798) 5.
- [BAL*15] BROOKS, M., AMERSHI, S., LEE, B., DRUCKER, S. M., KAPOOR, A., and SIMARD, P. "FeatureInsight: Visual support for error-driven feature ideation in text classification". *IEEE Conference on Visual Analytics Science and Technology*. 2015, 105–112. DOI: [10.1109/VAST.2015.7347637](https://doi.org/10.1109/VAST.2015.7347637) 5.
- [BHZ*18] BERNARD, J., HUTTER, M., ZEPPELZAUER, M., FELLNER, D., and SEDLMAIR, M. "Comparing Visual-Interactive Labeling with Active Learning: An Experimental Study". en. *IEEE Trans. Visualization and Computer Graphics*. TVCG'18 24.1 (2018), 298–308. DOI: [10.1109/TVCG.2017.2744818](https://doi.org/10.1109/TVCG.2017.2744818) 5.
- [BSP20] BEHRISCH, M., SCHRECK, T., and PFISTER, H. "GUIRO: User-Guided Matrix Reordering". *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 184–194. DOI: [10.1109/TVCG.2019.2934300](https://doi.org/10.1109/TVCG.2019.2934300) 6.
- [CD19] CAVALLO, M. and DEMIRALP, Ç. "Clustrophile 2: Guided Visual Clustering Analysis". *IEEE Transactions on Visualization and Computer Graphics*. TVCG'19 25.1 (2019), 267–276. DOI: [10.1109/TVCG.2018.2864477](https://doi.org/10.1109/TVCG.2018.2864477) 5.
- [CGM*17] CENEDA, D., GSCHWANDTNER, T., MAY, T., MIKSCH, S., SCHULZ, H.-J., STREIT, M., and TOMINSKI, C. "Characterizing Guidance in Visual Analytics". *IEEE Trans. Visualization and Computer Graphics* 23.1 (2017), 111–120. DOI: [10.1109/TVCG.2016.2598468](https://doi.org/10.1109/TVCG.2016.2598468) 3.
- [CHH*19] CASHMAN, D., HUMAYOUN, S. R., HEIMERL, F., PARK, K., DAS, S., THOMPSON, J., SAKET, B., MOSCA, A., STASKO, J., ENDERT, A., GLEICHER, M., and CHANG, R. "A User-based Visual Analytics Workflow for Exploratory Model Analysis". en. *Computer Graphics Forum*. EuroVIS'19 38.3 (2019), 185–199. DOI: [10.1111/cgf.13681](https://doi.org/10.1111/cgf.13681) 5.
- [CMQ20] CHENG, F., MING, Y., and QU, H. "DECE: Decision Explorer with Counterfactual Explanations for Machine Learning Models". *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.3030342](https://doi.org/10.1109/TVCG.2020.3030342) 6.
- [CRH*19] CAI, C. J., REIF, E., HEGDE, N., HIPPEL, J., KIM, B., SMILKOV, D., WATTENBERG, M., VIEGAS, F., CORRADO, G. S., STUMPE, M. C., and TERRY, M. "Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making". *Proc. Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: ACM, 2019, 4–14. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300234](https://doi.org/10.1145/3290605.3300234) 5.
- [CVL*18] COPPERS, S., VAN DEN BERGH, J., LUYTEN, K., CONINX, K., van der LEK-CIUDIN, I., VANALLEMEERSCH, T., and VANDEGHINSTE, V. "Intellingo: An Intelligible Translation Environment". en. *Proc. Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, 2018, 1–13. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3174098](https://doi.org/10.1145/3173574.3174098) 5.
- [CWZ*19] CHENG, H.-F., WANG, R., ZHANG, Z., O'CONNELL, F., GRAY, T., HARPER, F. M., and ZHU, H. "Explaining Decision-Making Algorithms Through UI: Strategies to Help Non-Expert Stakeholders". *Proc. Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: ACM, 2019, 559:1–559:12. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300789](https://doi.org/10.1145/3290605.3300789) 5.
- [CYL*20] CHEN, C., YUAN, J., LU, Y., LIU, Y., SU, H., YUAN, S., and LIU, S. "OoDAnalyzer: Interactive Analysis of Out-of-Distribution Samples". *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.2973258](https://doi.org/10.1109/TVCG.2020.2973258) 6.
- [DLW*17] DASGUPTA, A., LEE, J., WILSON, R., LAFRANCE, R. A., CRAMER, N., COOK, K., and PAYNE, S. "Familiarity Vs Trust: A Comparative Study of Domain Scientists' Trust in Visual Analytics and Conventional Analysis Methods". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), 271–280. DOI: [10.1109/TVCG.2016.2598544](https://doi.org/10.1109/TVCG.2016.2598544) 5.
- [dSBD*12] Dos SANTOS AMORIM, E. P., BRAZIL, E. V., DANIELS, J., JOIA, P., NONATO, L. G., and SOUSA, M. C. "iLAMP: Exploring high-dimensional spacing through backward multidimensional projection". *IEEE Conference on Visual Analytics Science and Technology*. Oct. 2012, 53–62. DOI: [10.1109/VAST.2012.6400489](https://doi.org/10.1109/VAST.2012.6400489) 5.
- [DSKE20] DAS, S., SAKET, B., KWON, B. C., and ENDERT, A. "Geono-Cluster: Interactive Visual Cluster Analysis for Biologists". *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.3002166](https://doi.org/10.1109/TVCG.2020.3002166) 6.
- [DVH*19] DINGEN, D., VEER, M. v., HOUTHUIZEN, P., MESTROM, E. H. J., KORSTEN, E. H. H. M., BOUWMAN, A. R. A., and WIJK, J. v. "Regression-Explorer: Interactive Exploration of Logistic Regression Models with Subgroup Analysis". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 246–255. DOI: [10.1109/TVCG.2018.2865043](https://doi.org/10.1109/TVCG.2018.2865043) 5.
- [EKC*20] EL-ASSADY, M., KEHLBECK, R., COLLINS, C., KEIM, D., and DEUSSEN, O. "Semantic Concept Spaces: Guided Topic Model Refinement using Word-Embedding Projections". *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 1001–1011. DOI: [10.1109/TVCG.2019.2934654](https://doi.org/10.1109/TVCG.2019.2934654) 6.
- [EKS*18] ESLAMI, M., KRISHNA KUMARAN, S. R., SANDVIG, C., and KARAHALIOS, K. "Communicating Algorithmic Process in Online Behavioral Advertising". *Proc. Conference on Human Factors in Computing Systems*. CHI '18. 2018, 432:1–432:13. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3174006](https://doi.org/10.1145/3173574.3174006) 5.
- [ESD*19] EL-ASSADY, M., SPERRLE, F., DEUSSEN, O., KEIM, D., and COLLINS, C. "Visual Analytics for Topic Model Optimization based on User-Steerable Speculative Execution". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 374–384. DOI: [10.1109/TVCG.2018.2864769](https://doi.org/10.1109/TVCG.2018.2864769) 5.
- [ESKC18] EL-ASSADY, M., SEVASTJANOVA, R., KEIM, D., and COLLINS, C. "ThreadReconstructor: Modeling Reply-Chains to Untangle Conversational Text through Visual Analytics". eng. *Computer Graphics Forum* 37.3 (2018), 351–365. DOI: [10.1111/cgf.13425](https://doi.org/10.1111/cgf.13425) 5.
- [ESS*18] EL-ASSADY, M., SEVASTJANOVA, R., SPERRLE, F., KEIM, D., and COLLINS, C. "Progressive Learning of Topic Modeling Parameters: A Visual Analytics Framework". *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), 382–391. DOI: [10.1109/TVCG.2017.2745080](https://doi.org/10.1109/TVCG.2017.2745080) 5.
- [GLC*19] GUZDIAL, M., LIAO, N., CHEN, J., CHEN, S.-Y., SHAH, S., SHAH, V., RENO, J., SMITH, G., and RIEDL, M. O. "Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators". en. *Proc. Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, 2019, 1–13. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300854](https://doi.org/10.1145/3290605.3300854) 5.
- [GZL*20] GOU, L., ZOU, L., LI, N., HOFMANN, M., SHEKAR, A. K., WENDT, A., and REN, L. "VATLD: A Visual Analytics System to Assess, Understand and Improve Traffic Light Detection". *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.3030350](https://doi.org/10.1109/TVCG.2020.3030350) 6.
- [HHC*19] HOHMAN, F., HEAD, A., CARUANA, R., DELINE, R., and DRUCKER, S. M. "Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models". *Proc. Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: ACM, 2019, 579:1–579:13. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300809](https://doi.org/10.1145/3290605.3300809) 5.

- [HKBE12] HEIMERL, F., KOCH, S., BOSCH, H., and ERTL, T. “Visual Classifier Training for Text Document Retrieval”. *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), 2839–2848. DOI: [10.1109/TVCG.2012.2775](https://doi.org/10.1109/TVCG.2012.2775).
- [HOW*19] HITRON, T., ORLEV, Y., WALD, I., SHAMIR, A., EREL, H., and ZUCKERMAN, O. “Can Children Understand Machine Learning Concepts?: The Effect of Uncovering Black Boxes”. en. *Proc. Conference on Human Factors in Computing Systems*. CHI’19. Glasgow Scotland Uk: ACM, 2019, 1–11. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.33006455](https://doi.org/10.1145/3290605.33006455).
- [JSR*19] JI, X., SHEN, H.-W., RITTER, A., MACHIRAJU, R., and YEN, P.-Y. “Visual Exploration of Neural Document Embedding in Information Retrieval: Semantics and Feature Selection”. *IEEE Transactions on Visualization and Computer Graphics* 25.6 (2019), 2181–2192. DOI: [10.1109/TVCG.2019.29039465](https://doi.org/10.1109/TVCG.2019.29039465).
- [KAKC18] KAHNG, M., ANDREWS, P. Y., KALRO, A., and CHAU, D. H. “ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models”. *IEEE Transactions on Visualization and Computer Graphics*. TVCG’18 24.1 (2018), 88–97. DOI: [10.1109/TVCG.2017.27447185](https://doi.org/10.1109/TVCG.2017.27447185).
- [KAS*20] KWON, B. C., ANAND, V., SEVERSON, K. A., GHOSH, S., SUN, Z., FROHNERT, B. I., LUNDGREN, M., and NG, K. “DPVis: Visual Analytics with Hidden Markov Models for Disease Progression Pathways”. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.29856896](https://doi.org/10.1109/TVCG.2020.29856896).
- [KAY*19] KITTLE-DAVIES, J., ALQARA-AWI, A., YANG, R., COSTANZA, E., ROGERS, A., and STEIN, S. “Evaluating the Effect of Feedback from Different Computer Vision Processing Stages: A Comparative Lab Study”. en. *Proc. Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, 2019, 1–12. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.33002735](https://doi.org/10.1145/3290605.33002735).
- [KBJ*20] KRUEGER, R., BEYER, J., JANG, W.-D., KIM, N. W., SOKOLOV, A., SORGER, P. K., and PFISTER, H. “Facetto: Combining Unsupervised and Supervised Learning for Hierarchical Phenotype Analysis in Multi-Channel Image Data”. *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 227–237. DOI: [10.1109/TVCG.2019.29345476](https://doi.org/10.1109/TVCG.2019.29345476).
- [KEV*18] KWON, B. C., EYSENBACH, B., VERMA, J., NG, K., FILIPPI, C. D., STEWART, W. F., and PERER, A. “Clustervision: Visual Supervision of Unsupervised Clustering”. *IEEE Transactions on Visualization and Computer Graphics*. TVCG’17 24.1 (2018), 142–151. DOI: [10.1109/TVCG.2017.27450855](https://doi.org/10.1109/TVCG.2017.27450855).
- [KF14] KAASTRA, L. T. and FISHER, B. “Field Experiment Methodology for Pair Analytics”. *Proc. Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization at VIS*. 2014, 152–159. ISBN: 978-1-4503-3209-5. DOI: [10.1145/2669557.26695722](https://doi.org/10.1145/2669557.26695722).
- [KPN16] KRAUSE, J., PERER, A., and NG, K. “Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models”. *Proc. Conference on Human Factors in Computing Systems*. 2016, 5686–5697. DOI: [10.1145/2858036.28585295](https://doi.org/10.1145/2858036.28585295).
- [LGM*20] LEITE, R. A., GSCHWANDTNER, T., MIKSCH, S., GSTREIN, E., and KUNTNER, J. “NEVA: Visual Analytics to Identify Fraudulent Networks”. en. *Computer Graphics Forum*. EuroVIS’20 39.6 (2020), 344–359. DOI: [10.1111/cgf.140426](https://doi.org/10.1111/cgf.140426).
- [LLL*19] LIU, S., LI, Z., LI, T., SRIKUMAR, V., PASCUCCI, V., and BREMER, P. “NLIZE: A Perturbation-Driven Visual Interrogation Tool for Analyzing and Interpreting Natural Language Inference Models”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 651–660. DOI: [10.1109/TVCG.2018.28652305](https://doi.org/10.1109/TVCG.2018.28652305).
- [LLS*18] LIU, M., LIU, S., SU, H., CAO, K., and ZHU, J. “Analyzing the Noise Robustness of Deep Neural Networks”. *IEEE Conference on Visual Analytics Science and Technology*. VAST’18. 2018, 60–71. DOI: [10.1109/VAST.2018.88025095](https://doi.org/10.1109/VAST.2018.88025095).
- [LPH*20] LEKSCHAS, F., PETERSON, B., HAEHN, D., MA, E., GEHLENBORG, N., and PFISTER, H. “PEAX: Interactive Visual Pattern Search in Sequential Data Using Unsupervised Deep Representation Learning”. en. *Computer Graphics Forum* 39.3 (2020), 167–179. DOI: [10.1111/cgf.139716](https://doi.org/10.1111/cgf.139716).
- [LSC*18] LIU, M., SHI, J., CAO, K., ZHU, J., and LIU, S. “Analyzing the Training Processes of Deep Generative Models”. *IEEE Transactions on Visualization and Computer Graphics*. TVCG’18 24.1 (2018), 77–87. DOI: [10.1109/TVCG.2017.27449385](https://doi.org/10.1109/TVCG.2017.27449385).
- [LSL*17] LIU, M., SHI, J., LI, Z., LI, C., ZHU, J., and LIU, S. “Towards Better Analysis of Deep Convolutional Neural Networks”. *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), 91–100. DOI: [10.1109/TVCG.2016.25988315](https://doi.org/10.1109/TVCG.2016.25988315).
- [LXL*18] LIU, S., XIAO, J., LIU, J., WANG, X., WU, J., and ZHU, J. “Visual Diagnosis of Tree Boosting Methods”. *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), 163–173. DOI: [10.1109/TVCG.2017.27443785](https://doi.org/10.1109/TVCG.2017.27443785).
- [MCZ*17] MING, Y., CAO, S., ZHANG, R., LI, Z., CHEN, Y., SONG, Y., and QU, H. “Understanding Hidden Memories of Recurrent Neural Networks”. *IEEE Conference on Visual Analytics Science and Technology*. 2017, 13–24. DOI: [10.1109/VAST.2017.85857215](https://doi.org/10.1109/VAST.2017.85857215).
- [MLMP18] MÜHLBACHER, T., LINHARDT, L., MÖLLER, T., and PIRINGER, H. “TreePOD: Sensitivity-Aware Selection of Pareto-Optimal Decision Trees”. *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), 174–183. DOI: [10.1109/TVCG.2017.27451585](https://doi.org/10.1109/TVCG.2017.27451585).
- [MP13] MÜHLBACHER, T. and PIRINGER, H. “A Partition-Based Framework for Building and Validating Regression Models”. *IEEE Transactions on Visualization and Computer Graphics*. TVCG’13 19.12 (2013), 1962–1971. DOI: [10.1109/TVCG.2013.1255](https://doi.org/10.1109/TVCG.2013.1255).
- [MQB19] MING, Y., QU, H., and BERTINI, E. “RuleMatrix: Visualizing and Understanding Classifiers with Rules”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 342–352. DOI: [10.1109/TVCG.2018.28648125](https://doi.org/10.1109/TVCG.2018.28648125).
- [MXC*20] MING, Y., XU, P., CHENG, F., QU, H., and REN, L. “ProtoSteer: Steering Deep Sequence Model with Prototypes”. *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 238–248. DOI: [10.1109/TVCG.2019.29342676](https://doi.org/10.1109/TVCG.2019.29342676).
- [MXLM20] MA, Y., XIE, T., LI, J., and MACIEJEWSKI, R. “Explaining Vulnerabilities to Adversarial Machine Learning through Visual Analytics”. *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 1075–1085. DOI: [10.1109/TVCG.2019.29346316](https://doi.org/10.1109/TVCG.2019.29346316).
- [PNKC20] PARK, H., NAM, Y., KIM, J.-H., and CHOO, J. “HyperTendrill: Visual Analytics for User-Driven Hyperparameter Optimization of Deep Neural Networks”. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.30303806](https://doi.org/10.1109/TVCG.2020.30303806).
- [PZDD19] PARVINZAMIR, F., ZHAO, Y., DENG, Z., and DONG, F. “MyEvents: A Personal Visual Analytics Approach for Mining Key Events and Knowledge Discovery in Support of Personal Reminiscence”. en. *Computer Graphics Forum* 38.1 (2019), 647–662. DOI: [10.1111/cgf.135965](https://doi.org/10.1111/cgf.135965).
- [RAL*17] REN, D., AMERSHI, S., LEE, B., SUH, J., and WILLIAMS, J. D. “Squares: Supporting Interactive Performance Analysis for Multiclass Classifiers”. *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), 61–70. DOI: [10.1109/TVCG.2016.25988285](https://doi.org/10.1109/TVCG.2016.25988285).

- [SDMT16] STAHNKE, J., DÖRK, M., MÜLLER, B., and THOM, A. “Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions”. *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2016), 629–638. DOI: [10.1109/TVCG.2015.2467717](https://doi.org/10.1109/TVCG.2015.2467717) 5.
- [SFB*20] SMITH-RENNER, A., FAN, R., BIRCHFIELD, M., WU, T., BOYD-GRABER, J., WELD, D. S., and FINDLATER, L. “No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML”. en. *Proc. Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, 2020, 1–13. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376624](https://doi.org/10.1145/3313831.3376624) 6.
- [SKB*18] SACHA, D., KRAUS, M., BERNARD, J., BEHRISCH, M., SCHRECK, T., ASANO, Y., and KEIM, D. A. “SOMFlow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance”. *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), 120–130. DOI: [10.1109/TVCG.2017.2744805](https://doi.org/10.1109/TVCG.2017.2744805) 5.
- [SLC*20] SUN, J., LI, Y., CHEN, C., LEE, J., LIU, X., ZHANG, Z., HUANG, L., SHI, L., and XU, W. “FDHelper: Assist Unsupervised Fraud Detection Experts with Interactive Feature Selection and Evaluation”. en. *Proc. Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, 2020, 1–12. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376140](https://doi.org/10.1145/3313831.3376140) 6.
- [SMD*16] SARKAR, A., MORRISON, C., DORN, J. F., BEDI, R., STEINHEIMER, S., BOISVERT, J., BURGGRAAFF, J., D’SOUZA, M., KONTSCHIEDER, P., ROTA BULÒ, S., WALSH, L., KAMM, C. P., ZAYKOV, Y., SELLEN, A., and LINDLEY, S. “Setwise Comparison: Consistent, Scalable, Continuum Labels for Computer Vision”. en. *Proc. Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, 2016, 261–271. ISBN: 978-1-4503-3362-7. DOI: [10.1145/2858036.2858199](https://doi.org/10.1145/2858036.2858199) 5.
- [SSBC19] SULTANUM, N., SINGH, D., BRUDNO, M., and CHEVALIER, F. “Doccurate: A Curation-Based Approach for Clinical Text Visualization”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 142–151. DOI: [10.1109/TVCG.2018.2864905](https://doi.org/10.1109/TVCG.2018.2864905) 5.
- [SSKE19] SPERRLE, F., SEVASTIANOVA, R., KEHLBECK, R., and EL-ASSADY, M. “VIANA: Visual Interactive Annotation of Argumentation”. *IEEE Conference on Visual Analytics Science and Technology*. 2019, 11–22. DOI: [10.1109/VAST47406.2019.8986917](https://doi.org/10.1109/VAST47406.2019.8986917) 5.
- [SSSE20] SPINNER, T., SCHLEGEL, U., SCHÄFER, H., and EL-ASSADY, M. “explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning”. *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), 1064–1074. DOI: [10.1109/TVCG.2019.2934629](https://doi.org/10.1109/TVCG.2019.2934629) 6.
- [WBL*20] WANG, X., BRYAN, C. J., LI, Y., PAN, R., LIU, Y., CHEN, W., and MA, K.-L. “Umbra: A Visual Analysis Approach for Defense Construction Against Inference Attacks on Sensitive Information”. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. DOI: [10.1109/TVCG.2020.3037670](https://doi.org/10.1109/TVCG.2020.3037670) 6.
- [WGSY19] WANG, J., GOU, L., SHEN, H., and YANG, H. “DQNviz: A Visual Analytics Approach to Understand Deep Q-Networks”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 288–298. DOI: [10.1109/TVCG.2018.2864504](https://doi.org/10.1109/TVCG.2018.2864504) 5.
- [WGY18] WANG, J., GOU, L., YANG, H., and SHEN, H. “GANviz: A Visual Analytics Approach to Understand the Adversarial Game”. *IEEE Transactions on Visualization and Computer Graphics* 24.6 (2018), 1905–1917. DOI: [10.1109/TVCG.2018.2816223](https://doi.org/10.1109/TVCG.2018.2816223) 5.
- [WGZ*19] WANG, J., GOU, L., ZHANG, W., YANG, H., and SHEN, H. “DeepVID: Deep Visual Interpretation and Diagnosis for Image Classifiers via Knowledge Distillation”. *IEEE Transactions on Visualization and Computer Graphics* 25.6 (2019), 2168–2180. DOI: [10.1109/TVCG.2019.2903943](https://doi.org/10.1109/TVCG.2019.2903943) 6.
- [WMJ*19] WANG, Q., MING, Y., JIN, Z., SHEN, Q., LIU, D., SMITH, M. J., VEERAMACHANENI, K., and QU, H. “ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning”. *Proc. Conference on Human Factors in Computing Systems*. CHI ’19. New York, NY, USA: ACM, 2019, 681:1–681:12. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300911](https://doi.org/10.1145/3290605.3300911) 6.
- [WSW*18] WONGSUPHASAWAT, K., SMILKOV, D., WEXLER, J., WILSON, J., MANÉ, D., FRITZ, D., KRISHNAN, D., VIÉGAS, F. B., and WATTENBERG, M. “Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow”. *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), 1–12. DOI: [10.1109/TVCG.2017.2744878](https://doi.org/10.1109/TVCG.2017.2744878) 5.
- [XCK*20] XIE, Y., CHEN, M., KAO, D., GAO, G., and CHEN, X. ’. “CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis”. en. *Proc. Conference on Human Factors in Computing Systems*. CHI’20. Honolulu HI USA: ACM, 2020, 1–13. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376807](https://doi.org/10.1145/3313831.3376807) 6.
- [XMT*20] XIE, T., MA, Y., TONG, H., THAI, M. T., and MACIEJEWSKI, R. “Auditing the Sensitivity of Graph-based Ranking with Visual Analytics”. *IEEE Transactions on Visualization and Computer Graphics*. TVCG’20 (2020), 1–1. DOI: [10.1109/TVCG.2020.3028958](https://doi.org/10.1109/TVCG.2020.3028958) 6.
- [XXM*19] XU, K., XIA, M., MU, X., WANG, Y., and CAO, N. “EnsembleLens: Ensemble-based Visual Exploration of Anomaly Detection Algorithms with Multidimensional Data”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 109–119. DOI: [10.1109/TVCG.2018.2864825](https://doi.org/10.1109/TVCG.2018.2864825) 6.
- [YGLR20] YAN, J. N., GU, Z., LIN, H., and RZESZOTARSKI, J. M. “Silva: Interactively Assessing Machine Learning Fairness Using Causality”. en. *Proc. Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, 2020, 1–13. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376447](https://doi.org/10.1145/3313831.3376447) 6.
- [ZWLC19] ZHAO, X., WU, Y., LEE, D. L., and CUI, W. “iForest: Interpreting Random Forests via Visual Analytics”. *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 407–416. DOI: [10.1109/TVCG.2018.2864475](https://doi.org/10.1109/TVCG.2018.2864475) 6.