sh Ontology	**Paper**				
explanationExperience.rdf	**ExplanationExperience**				
	hasDescription	**Description**			
aimodel.rdf		hasAIModel	**AIModel**		
			trained on	**Dataset**	
				hasDataType	**DataType**
					number of features
					number of instances
			solves	**AITask**	
				hasType	**AITaskType**
				hasGoal	**AITaskGoal = Description**
			utilises	**AIMethod**	
				hasType	**AIMethodType**
aimodelevaluation.rdf			annotated by	**AIModelAssessmentResult**	
				basedOn	**AIModelAssessmentMetric**
				measures	**AIModelAssessmentDimension**
explainer.rdf		hasExplainer (needExplainer?)	**Explainer**		
				hasOutputType	***Explanation***
				hasPortability	**Portability**
				hasConcurrentness	**ExplainerConcurrentness**
				hasPresentation	**InformationContentEntity**
				hasExplanationScope	***Explanation Scope***
				targetType	**Explanation Target**
user.rdf		hasUser	**User**		
			asks	**UserQuestion**	
				hasTarget	**UserQuestionTarget**
				hasType	**QuestionType**
			has intent	**Intent**	
				hasType	IntentType
			has resources	**Technical Facilities**	
				can handle	**ExplanationModality**
			hasGoal	**process**	
			possess	**Domain Knowledge**	
				level of domain knowledge	**KnowledgeLevel**
				level of AI knowledge	**KnowledgeLevel**
behaviour_tree.rdf	hasSolution	**Solution**			
		hasExplainer	**Explainer**		
			utilises	**ExplainabilityTechnique**	
				hasType	**ExplainabilityTechniqueType**
				hasOutputType	**Explanation**
				hasPortability	**Portability**
				hasConcurrentness	**ExplainerConcurrentness**
				hasPresentation	**InformationContentEntity**
				hasExplanationScope	***Explanation Scope***
				targetType	**Explanation Target**
				isCompatiblewithFeatureTypes	**DataType**

			hasComplexity	**ComputationalComplexity**
	hasDone	**UserEvaluation**		
		basedOn	**Metric**	
			measures	**Dimensions**
Code Availability				

Make it personal: a social explanation system applied to group recommendations	*DisCERN: Discovering Counterfactual Explanations using Relevance Features from Neighbourhoods*
PSIE	***DisCERN***
PsieGroupRecommendationExplanationExperience	LungCancerRiskExplanationExperience
Recommend movies to groups based on social knowledge	predict lung cancer risk given clinical data of patients
HappyMovieRecommenderSystem	LungCancerRiskPredictionModel
HappyMovieDataset	LungCancerRiskDataset
Tabular	Tabular
N/A	12
N/A	427
GroupRecommendation	CancerRiskPrediction
Recommendation	Multi-class Classification
Recommend movies to groups based on social knowledge	predict lung cancer risk given clinical data of patients
SocialGroupRecommender	CancerRiskRandomForest
Knowledge based Recommender	RandomForest
3.86, 3.69, 3.56, 3.89	87.94
Likert scale (5 points - being 5 strongly agree)	Accuracy
Usefulness, Decision process, Reusability, Usability	Performance
Content Based Explanation	Counterfactual Explanation
model-specific	model-agnostic
post-hoc	post-hoc
Any	Computational Entity
local	local
prediction	prediction
Moviegoer	Patient/Clinician
Why does the system recommend movie Y for group X?	How can patient X reduce cancer risk Y to lower
System Recommendation	System Recommendation
Why Question	How/What-if Question
Understand system recommendation	Reducing cancer risk
Trust, Satisfaction	Education, Taking Action
ScreenDisplay	ScreenDisplay
Any	Any
User profile	Any/ Clinical knowledge
high	low/high
low	low
	DisCERNCancerRiskExplainer
PsieRuleBasedTechnique	DisCERN
Knowledge Extraction	Feature Relevance+Example based
ContentBasedExplanation	Counterfactual Explanation
model-specific	model-agnostic
post-hoc	post-hoc
Visual/Textual	Computational Entity
local	local
prediction	prediction
Tabular	Tabular

N/A	N/A
	N/A
Questionnaire Usefulness/Helpfulness	
	https://github.com/RGU-Computing/DisCERN-XAI

Evaluating Explainability Methods Intended for Multiple Stakeholders

BTTelecom

BTTelecomRecommenderExplanationExperience
recommend engineering notes to desk support staff to help on-site engineers
EngineerNoteRecommender
BTEngineeringNotes
Text
300 tf-idf features
5352

EngineerNoteRecommendation
Recommendation,classification
Predict next scenario based on description in the engineering notes
EngineerNoteRecommender
Content based Recommender, Machine Learning / term frequency, unsupervised
50.88%, 99.10%(in lab), completeness - 70% (in practice)
Accuracy with and without "No New Action Required" (NNR) class, Automated - top K Accuracy, confidence scorehuman - completeness
Performance, human goal = Improve task performance in their role (i.e. efficiency of scenario organisation).

Neighbourhood Explanation, numerical, textual
model-agnostic
post-hoc
Any
local
prediction

Desk Agent
Why task Y is recommended as next task?
System Recommendation
Why question
why a recommendation has been made?
Transparency, Taking Action, Education
ScreenDisplay
Text, Image

BTNetworkPlannerDomainExpert, BTFieldEngineerDomainExpert, BTDeskAgentDomainNovice
high, high, low
low

BTRecommenderExplainer
BTContentSimilarityBasedTechnique: confidence score, feature-importance, summarisation of sim/difs
Knowledge Extraction + Feature Relevance
Neighbourhood Explanation
model-agnostic
post-hoc
Content + Similarity
local
prediction
Text

N/A

see Notes
Question to get feedback
Usefulness/Educatingness/Efficiency

Directing exploratory search: Reinforcement learning from user interactions with keywords
SciNet
DocumentSearchExplanationExperience
Determine the most related documents given a set of keywords
SciNetSearchEngine
WebOfScienceDataset
Documents
7 things, title, abstract, author names, publication year, publication forum, article, keywords
50million
DocumentRetrieval
InformationRetrieval
Determine the most related documents given a set of keywords
SciNetReinforcementLearning
Reinforcement Learning
0.71
Kappa
agreement between expert and system

Neighbourhood Explanation
model-specific
ante-hoc
visual
local
prediction

Scientist
Why was this result X retrieved for this query Y?
System Recommendation
Why Question
Understand system prediction
Effectiveness, Satisfaction
ScreenDisplay
Any

Search Domain
high
low

SciNetReinforcementLearning
Knowledge Extraction
Neighbourhood Explanation
model-specific
ante-hoc
Interactive visual
local
prediction
Documents

N/A

Questionnaire
Usability/Quality of user experience

Visualizing Recommendations to Support Exploration, Transparency and Controllability

TalkExplorer

TalkExplorerExplanationExperience
Recommend papers based on content and social connections
ConferenceNavigator3RecommenderSystem
ConferenceNavigator3Dataset
Tabular
N/A
N/A
TalkPaperRecommendation
Recommendation
Recommend papers based on content and social connections
CN3ContentBasedRecommender
Tf-idf + kNN
N/A
N/A
N/A

Neighbourhood Explanation
model-specific
ante-hoc
Image
local
prediction

Conference atendee
Why does the system recommend paper Y to user X?
System Recommendation
Why Question
Understand system prediction
Effectiveness, Transparency, Scruitability
ScreenDisplay
Any

Conference Topic Domain
high
low

TalkExplorerKNNTechnique
k-nearest Neighbour
Neighbourhood Explanation
model-specific
ante-hoc
visual
local
prediction
Tabular

N/A

Questions about explanation visualisation knowledge + tasks with TalkExplorer + Likert Scale questions about their needs at a conference and the usefulness of the visualization to address these needs
Think Aloud + Likert Scale
Effectiveness

Axiomatic Attribution for Deep Networks	Axiomatic Attribution for Deep Networks	Axiomatic Attribution for Deep Networks
IntGradImage	***IntGradRetinopathy***	***IntGradTextClassification***
IGImageClassificationExplanationExperience	DiabeticRetinopathyDetectionExplanationExperience	QuestionCategoryExplanationExperience
predict the category of a given image	predict if a given medical image contains diabetic retinopathy	predict the question category based on question text
IGImageClassificationModel	DiabeticRetinopathyDetectionModel	KimQuestionCategoryPredictionModel
ILSVRC-2014	EyePACS	WikiTableQuestions dataset
Image	Image	text
89401 pixels	1382400 pixels	N/A
456182	128175	22033
ImageClassification	DiabeticRetinopathyDetection	QuestionCategoryPrediction
Multi-class Classification	Binary Classification	Multi-class Classification
predict the category of a given image	predict if a given medical image contains diabetic retinopathy	predict the question category based on question text
GoogleNet	Fine-tubed InceptionV3	KimCNN-multichannel
Convolutional Neural Network	Convolutional Neural Network	Convolutional Neural Network
6.67%	90.3%, 98.1%, 99.1%	N/A
top-5 error	FOCP Sensitivity, FOCP Specificity, AUROC for EyePACS	N/A
Performance	Performance	N/A
Saliency Map Explanation	Saliency Map Explanation	Saliency Map Explanation
model-specific	model-specific	model-specific
post-hoc	post-hoc	post-hoc
Image	Image	Text
local	local	local
prediction	prediction	prediction
Any User	Clinician, Optomologist	Any User
Why does the system predict category Y for image X?	Why does the system predict RDR for image X?	Why does the system predict category Y for question text X?
System Recommendation	System Recommendation	System Recommendation
Why Question	Why Question	Why Question
Understand how system works	Understand how system works	Understand how system works
Transparancy, Trust	Transparancy, Trust, Education	Transparancy, Trust
ScreenDisplay	ScreenDisplay	ScreenDisplay
Any	Any	Any
Public Domain	Clinical Knowledge	Public Domain
Any	High	low
Any	Low	low
IntegratedGradientTechnique	IntegratedGradientTechnique	IntegratedGradientTechnique
IntegratedGradient	IntegratedGradient	IntegratedGradient
Saliency Map Explanation	Saliency Map Explanation	Saliency Map Explanation
model-specific	model-specific	model-specific
post-hoc	post-hoc	post-hoc
Annotated Computational Entity	Annotated Computational Entity	Annotated Computational Entity
local	local	local
prediction	prediction	prediction
Image	Image	Text

N/A	N/A	N/A
N/A	N/A	N/A
https://github.com/ankurtaly/Integrated-Gradients	https://github.com/ankurtaly/Integrated-Gradients	https://github.com/ankurtaly/Integrated-Gradients

Textual Explanations for Self-Driving Vehicles	iBCM: Interactive Bayesian Case Model Empowering Humans via Intuitive Interaction
KimEtAlMethod	***iBCM***
SelfDrivingExplanationExperience	iBCMGradingExplanationExperience
make acceleration or change course decisions in a self-driving car based on video	Cluster student assignment submissions to design grading rubric or to compose feedback
Self-drivingDecisionMakingModel	iBCMClusteringModel
BerkeleyDeepDriveDataset	iBCMAssessmentsDataset
Video	Code
40 seconds (frame rate not known)	N/A
6984	N/A
Self-drivingVehicalControl (acceleration and course)	AssessmentsClustering
Regression	Clustering
make acceleration or change course decisions in a self-driving car based on video	
Deep Neural Networks with Attention	iBCMClusteringMethod
NeuralNetwork	BCM
[2.29, 0.82], [6.06, 0.47]	N/A
[Mean of absolute error, Mean of distance correlation] of Acceleration and Course	N/A
self-drivingVehicalControl performance	N/A
Introspective Explanation	Prototype Explanation
model-agnostic	model-specific
post-hoc	ante-hoc
text	Computational Entity
local	cohort
prediction	prediction
Driver	Lecturer
Why does the vehical system make decision X?	Why does the system assign certian assessments in to one cluster? How does the system assign clusters?
System Recommendation	Model
Why Question	Why/How Question
Understand system decision	Understand system/Understand cohort of predictions
User acceptance, Trust, Understanding and extrapolation of vehicle behavior, Effective communication	Education/Transparency
ScreenDisplay	ScreenDisplay
text	Any
Public Domain	Lecturer Knowledge
high	High
low	Low
LSTMTextGeneratorExplanationTechnique	iBCMTechnique
Data-driven Explanation Generation	Interactive Bayesian Case Model
Introspective Explanation	Prototype Explanation ,Feature Importance Explanation
model-agnostic	model-specific
post-hoc	ante-hoc
text	Visual, textual
local	cohort
prediction	prediction
Any	Tabular,Image,Text

N/A

N/A

https://github.com/pair-code/saliency

N/A

Questionnaire
Usefulness/Efficiency

A case-based reasoning system for aiding detection and classification of nosocomial infections	Explaining Models by Propagating Shapley Values
InNoCBR	***DeepSHAPGlobal***
InnoInfectionExplanationExperience	MortalityPredictionExplanationExperience
Predict patient's infection based on a clinical, laboratory, and medico administrative based data	Predict patient mortality base on clinical, nutritional and behaviorial factors
InnoInfectionDiagnosisModel	MLPMortalityPredictionModel
InnHostpitalDataset	NHANESDataset
Tabular	Tabular
6 tuple {S, B, V, S, C, E}	79
5385	14407
InfectionDiagnosis	MortalityPrediction
Multi-class Classification	Binary Classification
Predict patient's infection based on a clinical, laboratory, and medico administrative based data	Predict patient mortality base on clinical, nutritional and behaviorial factors
InnoHybridMethod	MoralityMLP
Rules+PARTRules+NLP(NB)	Neural Network
70.21%, 0.62, 55.75%, 19.18%	82.56%
Accuracy, Kappa, false-positive rate (before and after modified data)	Accuracy
performance	Performance
Reasoning Path Explanation	Feature Importance Explanation
model-specific	model-agnostic
ante-hoc	post-hoc
text	Any
local	global
prediction	model
Spanish NHS Doctor	Clinician
Why does the system predict infection Y for patient X?	What/How features contributed to predicting mortality Y for patient X?
System Recommendation	Model
Why Question	What/How Question
Validate system prediction	Understand how model make decisions
Trust/Transparency	Transparency
ScreenDisplay	ScreenDisplay
Any	Any
Infection Detection/Prevention Knowledge	Clinical Knowledge
high	high
low	low
InnoDecisionPathTechnique	DeepSHAPExplanationTechnique
Decision tree	SHAP/Game-theory
Reasoning Path Explanation	Feature Importance Explanation
model-specific	model-agnostic
anti-hoc	post-hoc
Text	violin plot chart
local	global
prediction	model
Tabular	Tabular

N/A	N/A
N/A	N/A
	https://github.com/lrjball/shap

Explaining Models by Propagating Shapley Values
DeepSHAPLocal
MortalityExplanationExperience
Predict patient mortality base on clinical, nutritional and behaviorial factors
MLPMortalityPredictionModel
NHANESDataset
Tabular
79
14407
MortalityPrediction
Binary Classification
Predict patient mortality base on clinical, nutritional and behaviorial factors
MoralityMLP
Neural Network
82.56%
Accuracy
Performance

Feature Importance Explanation
model-agnostic
post-hoc
Any
local
prediction

Clinician
What/How features contributed to predicting mortality Y for patient X?
System Recommendation
What/How Question
Understand why model made a decision
Transparency/Education
ScreenDisplay
Any

Clinical Knowledge
high
low

DeepSHAPExplanationTechnique
SHAP/Game-theory
Feature Importance Explanation
model-agnostic
post-hoc
bar chart
local
prediction
Tabular

N/A

N/A

https://github.com/lrjball/shap