Supplementary Material

# Full Boolean Search Queries

Below is the complete set of search strings used. Searches were conducted on the ACM, IEEExplorer, Science Direct, and Springer Link digital libraries.

* ("data visualization" OR "information visualization" OR "knowledge visualization") AND ("reinforcement learning")
* ("data visualization" OR "information visualization" OR "knowledge visualization") AND ("reinforcement learning") AND ("explanation")
* ("data visualization" OR "information visualization" OR "knowledge visualization") AND ("reinforcement learning") AND ("explainable ai")
* ("data visualization" OR "information visualization" OR "knowledge visualization") AND ("reinforcement learning") AND ("XAI")
* ("data visualization" OR "information visualization" OR "knowledge visualization") AND ("reinforcement learning") AND ("black box")
* ("reinforcement learning") AND ("explainable")
* ("reinforcement learning") AND ("explanation")
* ("reinforcement learning") AND ("explainable ai")
* ("reinforcement learning") AND ("XAI")
* ("reinforcement learning") AND ("visual analytics")
* ("reinforcement learning") AND ("hybrid analytics")
* ("reinforcement learning") AND ("human-in-the-loop")

# Summary Table of Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Author** | **Title** | **Purpose** | **Theory or Method** | **Domain / Example** | **Publication Type** | **User Study (n)** |
| Amir et al. (2019) | Summarizing agent strategies. | Policy SummarizationHuman CollaborationQuery-based explanations | Theory | Robotics | Journal | - |
| Dao et al. (2018) | Deep Reinforcement Learning Monitor for Snapshot Recording | Visualization | Method | GamesGridworldPongPacman | Conference | - |
| Dethise et al. (2019) | Cracking Open the Black Box: What Observations Can Tell Us About Reinforcement Learning Agents | VisualizationVerification | Method | NetworkingVideo bitrate | Conference | - |
| Fukuchi et al. (2017a) | Application of Instruction-Based Behavior Explanation to a Reinforcement Learning Agent with Changing Policy | Policy Summarization | Method | Games | Conference | - |
| Fukuchi et al. (2017b) | Autonomous Self-Explanation of Behavior for Interactive Reinforcement Learning Agents | Policy Summarization | Method | Games | Conference | - |
| Hayes and Shah (2017) | Improving Robot Controller Transparency Through Autonomous Policy Explanation | Policy SummarizationHuman CollaborationQuery-based explanations | Method | RoboticsCartPoleGridworld | Conference | - |
| Huang et al. (2019) | Enabling robots to communicate their objectives. | Human Collaboration | Theory | Autonomous Vehicles | Journal | Mechanical Turk (*n* = 191) |
| Iyer et al. (2018) | Transparency and Explanation in Deep Reinforcement Learning Neural Networks | Visualization | Method | GamesPacman | Conference | University Students (*n*= 40) |
| Joo and Kim (2019) | Visualization of Deep Reinforcement Learning using Grad-CAM: How AI Plays Atari Games? | Visualization | Method | Games | Conference | - |
| Kazak et al. (2019) | Verifying Deep-RL-Driven Systems | VerificationQuery-based explanations | Method | NetworkingVideo bitrate | Conference | - |
| Lage et al. (2019) | Toward Robust Policy Summarization | Policy Summarization | Theory | GamesPacman | Conference | - |
| Madumal et al. (2019) | Explainable reinforcement learning through a causal lens | Policy SummarizationQuery-based explanations | Theory | GamesStarcraft II | arXiv | Mechanical Turk (*n*= 120) |
| Mishra et al. (2018) | Visual Sparse Bayesian Reinforcement Learning: A Framework for Interpreting What an Agent Has Learned | Visualization | Method | Gridworld | Conference | - |
| Pan et al. (2019) | Semantic Predictive Control for Explainable and Efficient Policy Learning | Visualization | Method | Autonomous VehiclesGamesGTA VFlappy Bird | Conference | - |
| Pynadath et al. (2018). | Transparency Communication for Machine Learning in Human-Automation Interaction | Human Collaboration | Theory | Defence / Military | Book Chapter | Yes(no details) |
| Sridharan and Meadows (2019) | Towards a Theory of Explanations for Human–Robot Collaboration | Policy Summarization | Theory | Robotics | Journal | - |
| Query-based explanations |
| Stamper and Moore (2019) | Exploring Teachable Humans and Teachable Agents | Policy Summarization | Method | GamesSpace InvadersConnect 4 | Conference | - |
| Tabrez et al. (2019) | Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning | Human Collaboration | Method | RoboticsGamesSoduku | Conference | Yes (no details) |
| Tabrez and Hayes (2019) | Improving Human-Robot Interaction Through Explainable Reinforcement Learning | Human Collaboration | Method | Gridworld | Conference | Yes(no details) |
| Wang et al. (2019) | DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks. | Visualization | Method | GamesBreakout | Conference | Experts(*n* = 3) |
| Ehsan et al. (2019) \* | Automated rationale generation: a technique for explainable AI and its effects on human perceptions | Human Collaboration | Method | GamesFrogger | Conference | Turk Prime (*n* = 60; *n* = 65) |
| Greydanus et al. (2017) \* | Visualizing and understanding atari agents | Visualization | Method | GamesBreakoutMs PacmanFrostbiteEnduro | arXiv | University Students(*n* = 31) |
| Lyu, et al. (2019) \* | Sdrl: Interpretable and data-efficient deep reinforcement learning leveraging symbolic planning | Policy Summarization | Method | Taxi DomainGridworldGamesMontezuma's Revenge | Conference | - |
| Verma et al. (2018) \* | Programmatically interpretable reinforcement learning. | Policy Summarization | Method | GamesTorcs Car Racing | arXiv | - |
| Yang et al. (2018) \* | Learn to interpret atari agents. | Visualization | Method | GamesBeam RiderEnduroFrostbiteMs PacmanPongSpace Invaders | arXiv | - |

\* indicates paper found through snowball sampling method