Assessing Explainability in Reinforcement Learning

Amber E. Zelvelder(✉), Marcus Westberg, and Kary Främling

Umeå University, Umeå, Sweden
{amber.zelvelder,marcus.westberg,kary.framling}@umu.se

Abstract. Reinforcement Learning performs well in many different application domains and is starting to receive greater authority and trust from its users. But most people are unfamiliar with how AIs make their decisions and many of them feel anxious about AI decision-making. A result of this is that AI methods suffer from trust issues and this hinders the full-scale adoption of them. In this paper we determine what the main application domains of Reinforcement Learning are, and to what extent research in those domains has explored explainability. This paper reviews examples of the most active application domains for Reinforcement Learning and suggest some guidelines to assess the importance of explainability for these applications. We present some key factors that should be included in evaluating these applications and show how these work with the examples found. By using these assessment criteria to evaluate the explainability needs for Reinforcement Learning, the research field can be guided to increasing transparency and trust through explanations.

Keywords: Reinforcement Learning · Explainable AI · XAI · Interpretable Machine Learning

1 Introduction

One key obstacle hindering the full scale adoption of Machine Learning, including Reinforcement Learning (RL), is its inherent opaqueness. This prevents these 'black box' approaches (i.e.; systems that hide their inner logic from users) from becoming more widespread and receiving greater authority and trust in decisions. Being opaque and having no explanations for why the autonomous agent takes an action or makes a decision can cause both practical and ethical issues [18,19]. With the recent increased calls for transparency in computer-based autonomous decision-making, there has been a surge in research into making Machine Learning algorithms more transparent. RL is often thought to not need further transparency if the reward condition is known, but in RL there is a variety of applications, each with their own set of interactions, recommendations etc. These different applications of RL will also have different needs when it comes to explainability.

D. Calvaresi et al. (Eds.): EXTRAAMAS 2021, LNAI 12688, pp. 223–240, 2021.
https://doi.org/10.1007/978-3-030-82017-6_14

This paper will seek to review how different application domains might affect explainability, and assess applications within these domains to determine the amount and type of explainability. We will start by setting out the background of RL, Explainable AI (XAI) and the state of XAI in RL. In this background section, we will go into detail on the main application areas in which RL is being used, or can potentially be used, followed by an overview of the types of explainability and then which types of XAI have been applied to RL. Then we show our methodology to find examples of RL applications. We will then present possible evaluation criteria and key factors for assessing explanations in RL. After that we will make an assessment of the most notable examples of applications we have identified in RL. Finally, we will conclude with a summary of our findings and present what we believe are the remaining challenges that future work could be based on.

2 Background

In RL, an algorithm learns dynamically from its environment and is driven by either a reward or penalty being given when reaching or being in a specific state or states [24,46]. Because of the way RL algorithms learn, through maximising the final reward, it is particularly suited for problems that require a solution that weighs the short term outcome against the long term outcome [46]. Robotics is the application domain that RL is the most prominent in by far [25], but it has gradually started to see more extensive use in control systems [29] and networking [33]. RL has also been showcased publicly as highly proficient in playing and winning a variety of games, such as GO , checkers and video games [4,12].

With the rise of deep learning, it has been made possible to scale RL to attempt tackling decision-making problems on a larger scale [3]. But with this increase in applications being developed using AI such as RL, there has also been an increase in demand for more transparency [50]. In addition to there being legal calls for transparency, there is also the argument that if autonomous agents can be clear about the reasons for their actions, this would help build rapport, confidence and understanding between the AI agents and human operators, thereby increasing the acceptability of the systems and enhancing end-user satisfaction [2,18].

Just like RL, XAI started gaining increased popularity in the 90s with symbolic reasoning systems, such as MYCIN. The interest in XAI remained mostly academic until the rise of Machine Learning (ML) and its involvement in making increasingly important decisions. The interest in explanations really took off after some concerns regarding bias within Machine Learning. Among the more well known and publicised cases of bias are the Amazon recruitment algorithm that advised against hiring women [38], and Flickr's image tagging algorithm that tagged people of some ethnicities as animals or objects [52]. A more common bias within RL is the possibility of model bias, where the learning environment is too different from the intended target environment [10]. Now the interest level

in XAI is high due to contemporary trust issues and ethics debate in the field of autonomous AI decision-making and the legal debate and requirements that are being imposed as a result. Despite the increased interest in XAI and the widespread implementation of RL in both research and industry, ways to implement explanations into RL have not been thoroughly researched.

One of the reasons lack of active development in RL explanations is because there is an underlying assumption that knowing the reward for RL is explicable enough. Another reason is that RL is often used in more mechanical situations, rather then conversational, in which case there might not be a user to directly interact with or the user is an expert to whom the actions are explicable when combined with observations of the environment.

In the following sections we will present the most prominent Application Domains where RL is used, the types of XAI that are relevant to RL and the current existing XAI techniques implemented for RL.

2.1 RL Main Application Domains

RL can be considered to be in its early stages of development when it comes to applications. The majority of works in the literature use simple test scenarios that are not always representative of real-life needs, but as some of these could be considered simplified versions of general applications, we will use some simplified examples to illustrate the potential use of RL in each application domain cluster outlined. We will rely on several examples from the recent edition of Sutton and Barto's book [46] on RL in order to identify the main clusters of RL applications. We will also include references to more recent works in literature where appropriate. These will be primarily used to illustrate the potential of RL within specific domain clusters.

Physics. Numerous examples of applied RL involve simplified Physics tasks. This is because we have easy access to mathematical models of the laws of physics, which allow us to build simulated environments in which the algorithm is tested. The simplified and well-known examples of this category (e.g.; cart-pole [46]) can also be moved to a more applied domain. Another example is the mountain-car task, where an under-powered vehicle surrounded by two hills needs to get up a hill but doesn't have the power to ascend it without gathering momentum by backing up the other way first [13,46]. This example could serve in the optimization of fuel use in actual cars in that situation if further developed.

Robotics. Due to it's similarity to the natural learning process, RL is well-suited to train the movement of robots, particularly the optimization of reaching movement goals. Because of this, robotics is an application area where RL has been thoroughly trialled and successfully applied for a variety of purposes [27]. Robotics usually uses a delayed or continuous reward, as it is the set of actions they seek to reward, rather than individual actions. One common application is to use it for path optimization, where the reward is given if the destination is

reached (and sometimes a bigger reward if it didn't take long). Many examples of successful applications of RL in Robotics can be found in the survey done by Kober *et al.* [25].

Games. Playing games is one of the ways RL is most well-known to people outside of the Computer Science expertise. RL has been shown to have great potential at learning games, because most games have a clear reward structure of winning the game. There have been cases of RL-trained algorithms being able to beat the (human) masters of the games Go and Chess [4]. There has been further research in making algorithms that can tackle multiple games [45]. Research into RL algorithms that perform video games on a human level (e.g.; With the same tools and at the same or higher skill level as a person) is also being performed, and has had some success [4,12]. Deep RL is also being used in newer research into using machine learning to play games and videogames, including videogames with or against other human players, with promising preliminary results [3].

Autonomous Vehicles and Transport. Although most autonomous vehicles use supervised learning to make sure they learn the correct rules, there has been research into using Reinforcement Learning instead [43]. Also, in the transport sector there are related systems that are using machine learning. For instance, transport systems use route optimization in a way similar to robotics, and there are also urban development applications that use machine learning to manage traffic control [30,35].

Healthcare. The healthcare sector is increasingly utilising ML in their systems, and this includes RL [20]. A lot of the current work on bringing machine learning to the healthcare sector is still in the early stages, but ML algorithms that enhance Computer Vision are already used commercially in diagnostics, medical imaging and surgery, to supplement the medical personnel [17,28]. With ML starting to influence the sector, and the motivation to improve this sector being high, new research is performed continuously to increase the successful use of ML. While ML within healthcare currently mostly supports the experts and professionals, research is also taking more interest in developing algorithms that interact directly with the (potential) patients, to either help them know what doctor they need, whether they are at risk of special ailments, or how to manage their health, either generally or with a specific condition [39].

Finance Predictions. Some areas of finance are doing research into how RL and other machine learning methods could be used to predict developments in aspects of the market [40]. As the finance sector already extensively uses rule- and trend-based models to try to stay ahead of the curve, machine learning is a natural next step for finance. Most applications within finance follow current trends and make estimations based on historical data [6].

Other. There are a few other domains in which RL development is prominent, which will not be as obviously relevant to this paper, but still are worth mentioning as possible domains to look into at a later point. Reinforcement learning is used in industry to try to alert when machines need preventative maintenance [9,15]. It is also used to assist with elevator scheduling [7,53] using a continuous-time Markov chain [16]. Other ways RL is used is in various types of optimization, examples including network communication optimization and general network management [16], optimising/minimising resource consumption [29] and optimising memory control [4]. RL is also being used to automatically optimise the web data shown to users by advertisers [4].

2.2 XAI Types

Sheh [44] proposes a way of categorising the types of explanations that are most commonly required for AI in different contexts. Further research into this has been done by Anjomshoae et al. [2], leading to a distinction of several types of explanations that are currently used in AI.

Teaching explanations aim to teach humans (General users, domain experts, AI experts) about the concepts that the AI has learned. These explanations don't always need to be accompanied by a decision, as the teaching is what is at the core of the explanation. These explanations can take the form of hypotheticals, for example as answers to follow-up questions regarding a previous explanation (i.e. "if parameter X was different, how would this have affected the decision?").

A possible subtype of teaching explanations are **contrastive explanations**, where the hypotheticals involve showing the user why the decision is better by contrasting it with the poorer choice(s). There has already been research into how contrastive explanations can be used in RL [2,49]. In this research, the possible consequences of other actions were generated to show them in contrast to what decision was made [36].

There are **introspective explanations** in the form of **tracing explanations** and **informative explanations**. The former is a trace of internal events and actions taken by the AI, the purpose of which is to provide a complete account (of desired granularity) of the decision process to track down faults or causes behind incorrect decisions. The latter type involves explaining discrepancies between agent decisions and user expectation by looking at the process behind the given decision, the purpose here being to improve human-robot and human-system interaction by either pointing out where an error may have occurred or convince the user that the agent is correct. What both have in common is that they draw directly from the underlying models and decision-making processes of the agent. Due to the nature of ML, these kinds of introspective explanations are often not compatible with such techniques, and in the few cases that they might be, they will not be complete [44].

By contrast, **post-hoc explanations** provide rationalisations of the decision-making process without true introspection. These explanations are helpful in models where tracing underlying processes is not an option, such as with

black-box models. These explanations can be derived from a simulation of what the underlying processes might be like, sometimes working from a parallel model that attempts to sufficiently approximate the hidden model.

Execution explanations are the simplest form of explanations, presenting the action or set of actions that the AI agent undertook. Similar to tracing explanations, this type of explanation provides a history of events, but does so by listing the explicit operations undertaken.

Post-hoc and execution explanations, together with teaching explanations, form the three types of explanations most compatible with RL, though their usefulness varies. Post-hoc and execution explanations provide methods of tracing and forming narratives regarding decision-making without having to "look inside the box". In turn, teaching explanations can help users understand the model of the application better.

2.3 Explainable Reinforcement Learning

In RL there is not necessarily a reason for the algorithm to take an action or make a suggestion. The actions or decisions made by the algorithm are often based on the experience that algorithm builds. Because of this there has been a particular interest in contrastive, post-hoc and tracing explanations for RL. Since gaming applications have been a very popular domain for RL, and games are generally a low-risk activity, it is a very popular domain to test new techniques in. Videogames have been used to test the understanding and clarity of saliency maps as explanations [1,21]. A numerical explanation has also been studied using videogames, and these studies also include user feedback to assess the quality of the explanations [11,21,42]. Another method of creating explanations for RL has been to amend the RL algorithm, to maintain an amount of "memory" [8], which provides a form of tracing explanation. This has also been implemented in a way to provide a visual representation of the internal memory of an agent by Jaunet et al.[23], which can be used a combination between a teaching and a tracing explanation. There is also some research looking into transparent or interpretable models, such as PIRL, hierarchical policies and LMUT [41].

3 Methodology

In order to make sure that our proposed evaluation criteria are thorough and applicable in existing application domains, we performed a literature survey of RL survey papers and books to extract as many examples as possible of RL being used in practice, or having a potential use in practice. We have assessed 19 pieces of literature, of which 14 contained viable examples. In total we noted 91 examples of possible applications. The levels of detail of each of these applications varied and some refer to the same application. In Fig. 1 we show a breakdown of the number of examples found in the more prominent sources.

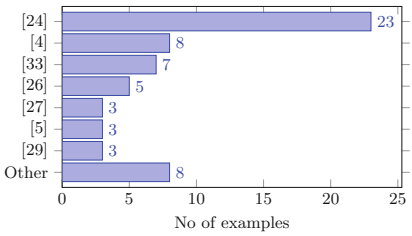

Fig. 1. Number of examples by source

For each survey paper we assessed whether the methods and applications found were applied to any scenario that could be used in practice. For each of these we made note of the survey papers they were mentioned in, their general application domain, a description of the application and, if mentioned, the RL method and the reward setup. After we had listed these, we assessed similarities between different listings and merged them in our data where appropriate, preserving the multiple sources. We were then left with 50 examples that could be evaluated. After these examples were collected, we found that they were in 8 different categories (Fig. 2). In this we found 16 were in the robotics domain, 8 each in the games, control systems and networking domains, 3 in autonomous vehicles and 2 in both finance and healthcare.

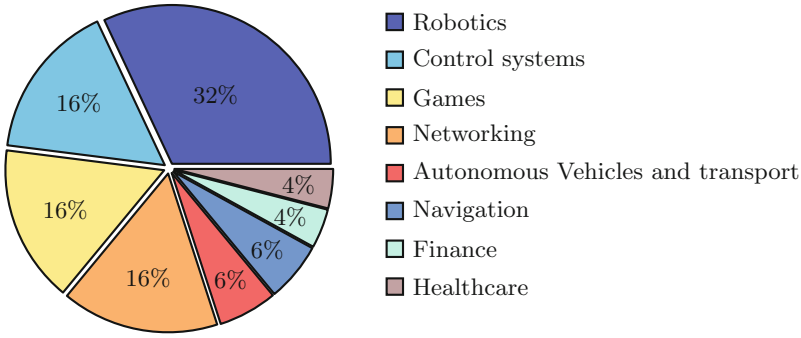

Fig. 2. Application domains of Practical examples

3.1 Analysis of Application Domains

Many of the examples in **robotics** show that RL can be used for a robotic agent to master its understanding of physics by achieving balance of itself or another object, or making some other type of adjustments based on gravity and other laws of physics [4,13,24,26,27]. Another type of robotics that frequently uses RL is robots that move around and perform a task, such as moving objects [4,24,26] or finding an exit [4,31]. There are also robots that use RL to play ball games

[5,26,31] and one example uses RL to do a peg-hole insertion [26]. Another important point to make about robotics is that a lot of RL applications in other domains can also be connected to robotics. For instance, in healthcare one of the applications is gesture reading and replication [32], which would be a combination of computer vision and robotics. **Games** have often been used to showcase how well a machine learning algorithm can perform, as with TD-Gammon and Samuel's Checker player beating human masters of the games of backgammon and checkers [4,24]. Most applications in this domain are one versus one board games. There are also applications that use RL for playing videogames [4,22]. The examples within the **control systems** domain found were primarily in optimization [4,24], resource allocation [4,24,29] and task scheduling [4,24,29]. A lot of these applications are located in factory settings, but there are also examples in smart homes and Internet-of-Things settings. Most of the **networking** applications are related to access control, caching and connection preservation [4,33]. There was one example relating to personalising web services [4] and network security [33]. In **Autonomous vehicles and Transport**, the applications are primarily about avoiding collision and interpreting other traffic [5,26,33,43]. **Navigation** is very similar to the Autonomous vehicles, with the main difference being that traffic is not necessarily a problem [4,14,24]. In **finance** we found two examples, one involved creating economic models [4] and the other was an automated trading application [5]. In **healthcare**, we found the application mentioned before that reads and replicates surgical gestures [32] and an application that detects and maps a person's bloodvessels [54].

4 Key Criteria

There are several factors that contribute to the need for explainability, a large amount of which relate to Human-Computer Interaction (HCI), but there are several other types of factors that contribute. In this section we explain the different criteria that drive the evaluation and assessment of explanations, the need for explanations and the types of explanations required. We start by describing four key factors (Fig. 3) which we will use for assessments, we then have a few sections to highlight notable contributing factors that assist in the evaluation of the key factors and in what way they contribute to the key factors.

Out of the key factors, two are closely related to HCI, User Expectancy and User Expertise, and two of them more related to the consequences of the application, Urgency of the output and Legality. The first key factor, and perhaps the best starting point for assessing the explanations, is the **User Expectancy**. The amount of explanation a user expects is a direct influence on the amount of explanation required, but it is also influenced by the users expectations of the applications actions or outputs [37]. If a user expects little or no explanation, and the application gives the expected output, there is no need for an explanation. But if the user expects an explanation, or if the application is behaving in a way that is outside of the user expectations, a more expansive explanation would be required. Because of the nuances of this factor, it also has a strong

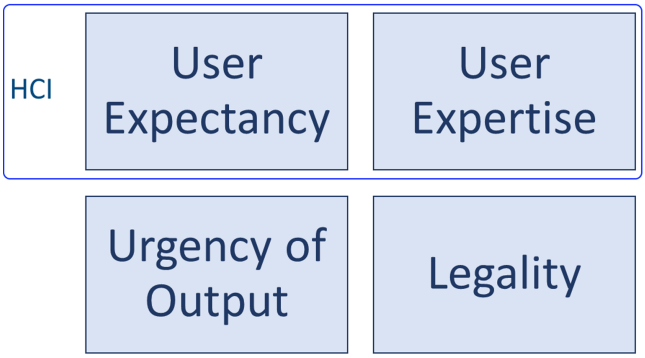

Fig. 3. Four key factors that influence the need for explainability

influence on the type of explanation that would be the most suitable for the application. The other user-connected key factor is, the **User Expertise**. The level of expertise the intended user has varies between every application. This factor mostly dictates the type of explanation to be made available [47]. If the user can be anyone, the explanations will need to be more informative, using general language. If the target users are experts in the application domain, but not on this specific application, an informative explanation that is specific to the domain can be presented. Finally, if the user is an expert on the application and/or any devices the application controls, an explanation that helps the user trace down faults within the system, and explanations of executed actions would be preferred.

A key factor that gives a limitation on the detail and type of explanation given is the **Urgency of the Output**. Longer and detailed explanations also take longer to produce and also take longer for the user to interact with. This is a time-bound scale, which means that it determines if there is time to produce any form of explanation before a decision is taken. If an action needs to be taken immediately, like in an autonomous vehicle, there might not be enough time to explain it to the user, so in this case an explanation can not be expected until later. If there is no urgency, the application can produce a full report and even request the user to approve the decision or action based on the explanation.

Finally, a factor that has gotten more important recently, is the **Legality**. This factor is driven by the laws affecting the application domain, AI in general or the specific case of the application. In this case it matters if it affects or evaluates an individual or group of individuals. If it can do neither there is no legal need for explanations. If it evaluates, then there should be the possibility to produce an explanation of the evaluation, but this can be retrospective. If it affects, it is preferred that an immediate explanation is given, as well as a retrospective explanation being made available.

These key factors were chosen on the grounds that they cover and represent the core areas of XAI concern. Differences in user expertise has shown to affect expectations of output and explanatory content [47], and the context of the

audience has great impact on how explanations are to be tailored [37]. In this way, expertise and expectancy have a fair amount of overlap, but in this paper we treat expectancy separately to acknowledge other factors that can also impact expectations, and if explanations are to be expected at all. Urgency of output is chosen because it has a defining impact on the nature of explanation to be provided and the context in which it is delivered. Finally, legality is of great concern for XAI researchers both on the basis of ethical and economic concerns, thus placing it as a very important key factor.

In the following sections we explain how some important criteria feed into these key factors.

4.1 The Intended User

The starting point for any application should be the intended user. This is very important for explanations, as there has to be an individual or group of individuals that the explanation would be intended for. The user affects both HCI-based aspects of the key factors and it also influences both the amount of explanation that should be given, as well as the type of explanations (see Sect. 2.2). The development of an application and explanations for the applications should be started from the user expectations, requirements and expertise.

4.2 Means of Interaction

Any application that includes direct interaction with some kind of user will need some explanation, but this can be limited by the means that the application or the user have to interact. With many RL implementations, there is some limitation to what kind of interaction there can be, and this can limit the possibilities of their communication. For instance, in robotics the application can usually interact by movement or other non-verbal communication, which humans can sometimes intuitively interpret [51]. In applications located in factories or other industrial devices, the means might just be a digital display or a blinking light. In other RL applications the interaction can be done via a monitor, or using audio. This contributing factor has no great influence on the key factors, but is used as a limitation to the type of explanations that can be implemented for an application.

4.3 Industry Sector

The type of industry sector that an application is developed in and for is key in determining the need and nature of the explainability of machine learning and AI systems. The industry often dictates who the intended user is, and what aspects of the system needs automation. For instance, an AI in healthcare can either interact with the medical professionals, or with the patients themselves. But in the manufacturing industry, the user is far more likely to be a person who has expert knowledge on the topic of the algorithm. Although explanations

are preferred in most industry sectors, the need for explanations is greater in some than in others. For instance, in the healthcare sector it is very important for AI to explain themselves, because the decision or advice of the RL algorithm could affect the health of an individual. This is the case for all users of AI within healthcare, and it has been shown that people are more likely to trust the health advice of an AI if an explanation and/or motivation is given by the AI that pertains to the patient personally [39]. This is different from the finance sector, where financial predictions are often assumed to be estimates, so a motivation or detailed explanation is not needed as much, but a chart or list of rules as a form of introspective informative explanation to support the prediction would be preferred. This contributing factor strongly influences the HCI aspects of the key factors, as well as the other key factors to a lesser extend, depending on the specific application.

4.4 Urgency/Time-Restraint

The urgency by which an algorithm needs to make a decision has some influence on how much explanation can be expected and/or is needed. If it is, for instance, an AI that drives a car, making the decision to do an emergency brake for a suddenly crossing pedestrian is more important than explaining it. The same goes for a machine stopping production if there is a misalignment in the system. Both of these could be followed up by a longer explanation, after the urgency is reduced. In other examples, such as an algorithm that makes mortgage agreement decisions, time is of less importance than an explanation regarding how a decision was made.

4.5 Legal

This factor is emphasized by recent legislation across the world. As the EU has recently passed a law known as the GDPR [48], which dictates that if an algorithm is fully or partially responsible for any decisions made regarding a person, that person has the right to know the reasons behind the algorithm's conclusion.

4.6 Responsibility

Whether the user, the creator of the application, the manufacturer of a device using the application, or some other individual or organisation has the responsibility for action or decisions chosen by the application is also a driver for wanting or needing explanations of differing types. This contributing factor is tied to the legality key factor.

5 Assessment

Although some of the key factors can be subjective, in Fig. 4 we show an overview of how the unique examples found in Sect. 3 could be classified on a scale of 1–5

on the key factors presented in Sect. 4. This gives an overview of the need for explanations in different sectors and shows how much variability exists in the domains. For a specific application, the details of how the specific key factors are relevant are more important then just their value. Therefore, in this section we will evaluate four of the scenarios we extracted from the examples reviewed in Sect. 3. We are using the key factors from Sect. 4 to perform this evaluation. The examples in this section were chosen because they are from varied application domains and have very different users. They are therefore expected to have very different needs when it comes to explanations.

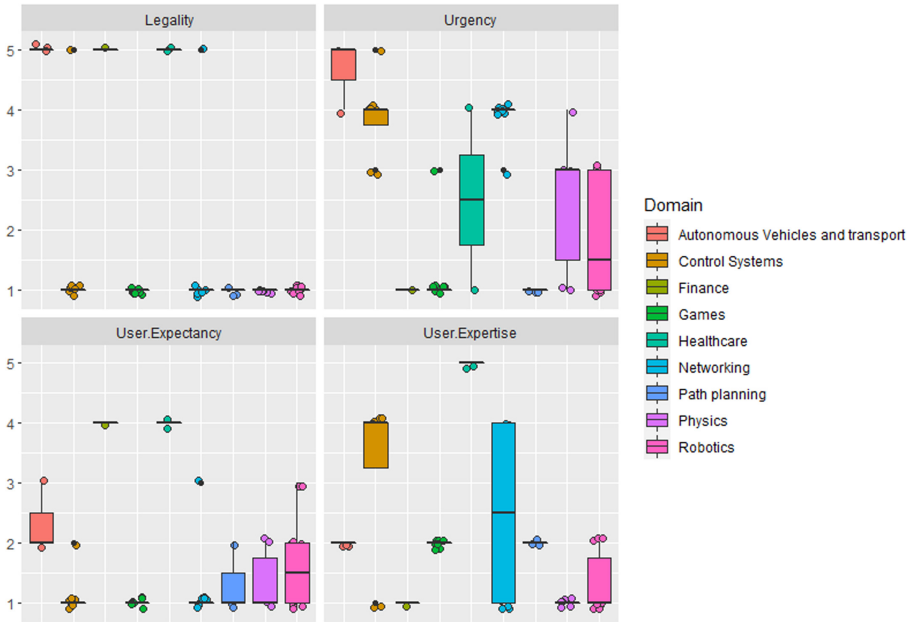

Fig. 4. Figure for classification of known examples on a scale of 1–5

5.1 A Box-Moving Robot

This example was chosen for evaluation because the functionality of the robot is simple and can be kept within the robotics domain, but has the potential of being used in many other domains with minor changes. We will be assuming the robot is an industry-ready adaptation and has therefore the capability to find a (specific) box, move the box, and recover from a position in which the robot itself is stuck in a corner [4,24,26,34]. This application does not have a specific intended user, which presents two options as to who to regard as a user; the person who has given the robot a task to do, or anybody who might be in the area where the robot operates. This implies that the user expertise and expectation of explanations will be low, assuming the robot performs as

intended. The urgency in the case of this robot is mostly not time-critical as the robot can wait to make a decision if there is no threat of any kind present. This means that the robot has time to display an indicator or play a sound to facilitate human understanding. Specific design decisions may limit these capacities. For example, the OBELIX robot [34] currently has no means to play sounds or show any kind of display, which means its means of communications are left to physical movements. The robot does not make decisions about people, but has a chance to encounter people within its work space. This means that the only legal requirement for the robot to give an explanation is when it interferes with an individuals actions in any way, such as by bumping into them or obstructing their path with the boxes.

5.2 Cloud Computing Resource Allocation

This example was picked because cloud computing is increasingly popular, and the optimisation of all its procedures is critical to its success. In the cloud computing resource allocation problem, there is a server cluster with a certain amount of physical servers and each of these physical servers can provide a limited number of resources [29]. A job will be processed when enough resources are available, the algorithm is employed to optimise how and when the jobs are allocated and to which machines, to optimise the processing time and minimise the power consumption. In this scenario, a user will be the person who submits a job to the cloud computing. The user will generally have some expertise in cloud computing, but the amount of expertise will vary. A user will typically have little to no expectation of explanations, unless the system has issues with the performance. As one of the goals of the algorithm is to optimise the time, there is no time to explain actions before they are taken. As individuals are not evaluated, there is no legal requirement for explanations. The recommended type of explanation would be a tracing explanation or execution explanation to track down the cause of a fault within the system.

5.3 Frogger Videogame

We are evaluating the frogger videogame as an example from the gaming domain. This example was selected because there already exist an experimental study into explanations [11,42] for this game and because of its iconic reputation. In the game of Frogger you need to guide a frog from one side of a map to the other. In the first part, the frog needs to avoid being hit by a car, and at the second part the frog needs to jump between moving logs to reach the other side of a river. Since this is a game being played by the RL algorithm, there is no user, only observers. We can assume the user knows the rules of the videogame, but has no further expertise. As the objects other than the frog move in real-time, that is the time sensitivity required for the reactions. The game doesn't make decisions on individuals, so there is no legal requirement for explanations. Because of academic interest, the explanations currently being developed for Frogger are to indicate, either numerically or by description, what observations of the AI are

being most relevant in its decision at any point during the run. This application is therefore often used to perform benchmarking of explanation techniques and user studies.

5.4 Surgical Robot

The example of a surgical robot has been chosen because it has a strong contrast with the other examples so far. The specific example we use is of a robot performing a suturing task [32]. In the referenced paper this is being done in a simulated environment, but since the goal is to let it perform in a medical environment, we shall evaluate it as such. The robot is performing a medical operation (suture) on a person (the patient), and is being monitored by another person (the doctor). The intended user is the doctor, who will have a high amount of medical expertise to assess that the robot does the correct procedures, and will also be capable of spotting any mistakes made. In the application, a display of certain parameters was included, so this display could be considered a starting point for an explanation. The display in the example indicated how accurate the expected kinaesthetic response was compared to the actual, which can indicate a possible problem if the accuracy is too far off. Since the performance of this robot directly affects the health of an individual, this means that legally there is a strict requirement for explanations to be available prior to use on a patient.

6 Discussion

There are still issues that prevent RL from more widespread application in general, with the core issues being centered around the inability to adapt, lack of correlation, application complexity, increasingly larger and more complex data and the narrow focus of current XAI techniques [27,51]. In RL there is a reward state that can always help clarify the goal that the algorithms are working towards, so one of the important steps towards explainability is to make sure the reward states can be viewed by the user in a way the typical user can understand. As RL algorithms can sometimes use very complicated reasoning, which might be beyond what a human understands, it can be hard for an algorithm to be accompanied by an explanation that a human can easily follow. This is further complicated if the system has to make several actions in succession that each require a longer explanation than a human would be able to keep up with. In deep RL this becomes even more of a problem as the input and the parameters are very expansive, making it harder to tie specific outcomes to specific parameters. Another challenge is that an explanation must be good enough that a human can help keep it accountable. Related to this is the problem of who can be held responsible for any wrong decision that is made as a consequence of an explanation being insufficient. Because XAI is still in early stages of implementation within machine learning and even more so when it comes to RL, due to RL having a more limited application area, there is very little existing research in XAI for RL. In RL most current research into explainability is focusing on

'recommender' systems. However, as presented earlier in this paper, RL is used or will soon be used in many other critical areas where explanations might be required in various degrees.

7 Conclusion

This paper has shown which application domains RL is most used in, and why explainability is important in RL. It has also presented guidelines that can be used to evaluate the explainability needs for specific applications. The guidelines are centered around the HCI aspects of user expectations and expertise as well as the urgency of the output and the legal requirements where applicable. We have assessed various notable applications that use RL algorithms using the guidelines provided. As we continue to work towards Explainable RL, the guidelines set out in this paper will help identify the need for explainability in new RL applications.

Acknowledgement. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. We would also like to thank Amro Najjar, for providing guidance on starting this paper.

References

1. Anderson, A., et al.: Explaining reinforcement learning to mere mortals: An empirical study. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, August 2019. http://dx.doi.org/10.24963/ijcai.2019/184
2. Anjomshoae, S., Najjar, A., Calvaresi, D., Främling, K.: Explainable agents and robots: Results from a systematic literature review. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. p. 1078–1088. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019)
3. Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A.: A Brief Survey of Deep Reinforcement Learning. IEEE Signal Processing Magazine, Special Issue on Deep Learning for Image Understanding p. 16 (aug 2017)
4. Barto, A., Thomas, P., Sutton, R.: Some Recent Applications of Reinforcement Learning. Workshop on Adaptive and Learning Systems (2017)
5. Busoniu, L., Cluj-napoca, U.T., Babuska, R., Schutter, B.D.: Innovations in Multi-Agent Systems and Applications - 1, vol. 310. Springer Nature (2010)
6. Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A.: Reinforcement learning and savings behavior. The Journal of Finance **64**(6), 2515–2534 (2009)
7. Crites, R.H., Barto, A.G.: Elevator group control using multiple reinforcement learning agents. Machine Learning **33**(2), 235–262 (1998)
8. Cruz, F., Dazeley, R., Vamplew, P.: Memory-based explainable reinforcement learning. In: Liu, J., Bailey, J. (eds.) AI 2019. LNCS (LNAI), vol. 11919, pp. 66–77. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35288-2_6
9. Das, T.K., Gosavi, A., Mahadevan, S., Marchalleck, N.: Solving semi-Markov decision problems using average reward reinforcement learning. Manage. Sci. **45**(4), 560–574 (1999)

10. Deisenroth, M., Rasmussen, C.: Reducing model bias in reinforcement learning (12 2010)
11. Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., Riedl, M.: Automated rationale generation: a technique for explainable AI and its effects on human perceptions, pp. 263–274 (03 2019). https://doi.org/10.1145/3301275.3302316
12. Erev, B.I., Roth, A.E.: Predicting how people play games?: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev. **88**(4), 848–881 (1998)
13. Främling, K.: Light-weight reinforcement learning with function approximation for real-life control tasks. Proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics, Intelligent Control Systems and Optimization (ICINCO-ICSO) (2008)
14. Garcia, J., Fernandez, F.: A comprehensive survey on safe reinforcement learning. J. Mach. Learn. Res. **16**, 1437–1480 (2015)
15. Gosavi, A.: Reinforcement learning for long-run average cost. Eur. J. Oper. Res. **155**(3), 654–674 (2004). Traffic and Transportation Systems Analysis
16. Gosavi, A.: Reinforcement learning: a tutorial survey and recent advances. INFORMS Journal of Computing **21**, 178–192 (2018)
17. Gupta, M., Konar, D., Bhattacharyya, S., Biswas, S. (eds.): Computer Vision and Machine Intelligence in Medical Image Analysis. AISC, vol. 992. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-8798-2
18. Hellström, T., Bensch, S.: Understandable robots-what, why, and how. Paladyn J. Behav. Robot. **9**(1), 110–123 (2018)
19. Hendricks, L.A., Akata, Z., Rohrbach, M., Schiele, B., Darrell, T.: Generating Visual Explanations
20. Hinto, G.: Deep learning - a technology with the potential to transform healthcare. JAMA **320**, 1101–1102 (2018)
21. Huber, T., Limmer, B., André, E.: Benchmarking perturbation-based saliency maps for explaining deep reinforcement learning agents. arXiv preprint arXiv:2101.07312 (2021)
22. Jaderberg, M., et al.: Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science **364**(6443), 859–865 (2019)
23. Jaunet, T., Vuillemot, R., Wolf, C.: DRLViz: understanding decisions and memory in deep reinforcement learning. In: Computer Graphics Forum, vol. 39 (2020)
24. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: a survey. J. Artif. Intell. Res. **4**, 237–285 (1996)
25. Kober, J., Bagnell, A.J., Peters, J.: Reinforcement learning in robotics: a survey. Reinforcement Learn. **32**, 1238–1274 (2012)
26. Kober, J., Bagnell, J.A., Peters, J.: Reinforcement learning in robotics: a survey. Int. J. Robot. Res. **32**, 1238–1274 (2013)
27. Kormushev, P., Calinon, S., Caldwell, D.: Reinforcement learning in robotics: applications and real-world challenges. Robotics **2**(3), 122–148 (2013)
28. Law, H., Ghani, K., Deng, J.: Surgeon technical skill assessment using computer vision based analysis. In: Doshi-Velez, F., Fackler, J., Kale, D., Ranganath, R., Wallace, B., Wiens, J. (eds.) Proceedings of the 2nd Machine Learning for Healthcare Conference. Proceedings of Machine Learning Research, vol. 68, pp. 88–99. PMLR, Boston, Massachusetts, 18–19 August 2017
29. Li, H., Wei, T., Ren, A., Zhu, Q., Wang, Y.: Deep reinforcement learning: Framework, applications, and embedded implementations: invited paper. In: IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers, ICCAD 2017-November, pp. 847–854 (2017)

30. Liang, X., Du, X., Wang, G., Han, Z.: Deep Reinforcement Learning for Traffic Light Control in Vehicular Networks. arXiv e-prints, March 2018
31. Littman, M.L.: Markov games as a framework for multi-agent reinforcement learning. Mach. Learn. Proc. **1994**, 157–163 (1994)
32. Liu, D., Jiang, T.: Deep reinforcement learning for surgical gesture segmentation and classification. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 247–255. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_29
33. Luong, N.C., Hoang, D.T., Gong, S., Niyato, D., Wang, P., Liang, Y.C., Kim, D.I.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutorials **21**(4), 3133–3174 (2019)
34. Mahadevan, S., Connell, J.: Automatic programming of behavior-based robots using reinforcement learning. Artif. Intell. **55**(2), 311–365 (1992)
35. Mannion, P., Duggan, J., Howley, E.: Parallel reinforcement learning for traffic signal control. Procedia Comput. Sci. **52**, 956–961 (2015). The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015)
36. Miller, T.: Contrastive explanation: a structural-model approach. CoRR abs/1811.03163 (2018)
37. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019)
38. Mujtaba, D.F., Mahapatra, N.R.: Ethical considerations in AI-based recruitment. In: 2019 IEEE International Symposium on Technology and Society (ISTAS), pp. 1–7. IEEE (2019)
39. Neerincx, M.A., van der Waa, J., Kaptein, F., van Diggelen, J.: Using perceptual and cognitive explanations for enhanced human-agent team performance. In: Harris, D. (ed.) EPCE 2018. LNCS (LNAI), vol. 10906, pp. 204–214. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91122-9_18
40. Nevmyvaka, Y., Feng, Y., Kearns, M.: Reinforcement learning for optimized trade execution. In: Proceedings of the 23rd International Conference on Machine Learning, ICML 2006, pp. 673–680. ACM, New York (2006)
41. Puiutta, E., Veith, E.M.S.P.: Explainable reinforcement learning: a survey. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2020. LNCS, vol. 12279, pp. 77–95. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57321-8_5
42. Sequeira, P., Gervasio, M.: Interestingness elements for explainable reinforcement learning: understanding agents' capabilities and limitations. Artif. Intell. **288**, 103367 (2020)
43. Shalev-Shwartz, S., Shammah, S., Shashua, A.: Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving (2016)
44. Sheh, R.: Different XAI for different HRI. In: AAAI Fall Symposium Technical Report, pp. 114–117 (2017)
45. Silver, D., et al.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science **362**, 1140–1144 (2018)
46. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction, 2nd edn. The MIT Press, Cambridge (2018)
47. Szymanski, M., Millecamp, M., Verbert, K.: Visual, Textual or Hybrid: The Effect of User Expertise on Different Explanations, pp. 109–119. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3397481.3450662

48. Voigt, P., von dem Bussche, A.: The EU General Data Protection Regulation (GDPR). Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57959-7
49. van der Waa, J., van Diggelen, J., van den Bosch, K., Neerincx, M.A.: Contrastive explanations for reinforcement learning in terms of expected consequences. CoRR abs/1807.08706 (2018)
50. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv. JL & Tech. **31**, 841 (2017)
51. Westberg, M., Zelvelder, A., Najjar, A.: A historical perspective on cognitive science and its influence on XAI research. In: Calvaresi, D., Najjar, A., Schumacher, M., Främling, K. (eds.) EXTRAAMAS 2019. LNCS (LNAI), vol. 11763, pp. 205–219. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30391-4_12
52. Yapo, A., Weiss, J.: Ethical implications of bias in machine learning. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018)
53. Yuan, X., Buşoniu, L., Babuška, R.: Reinforcement learning for elevator control. IFAC Proc. Vol. **41**(2), 2212–2217 (2008). 17th IFAC World Congress
54. Zhang, P., Wang, F., Zheng, Y.: Deep reinforcement learning for vessel centerline tracing in multi-modality 3D volumes. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 755–763. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_86