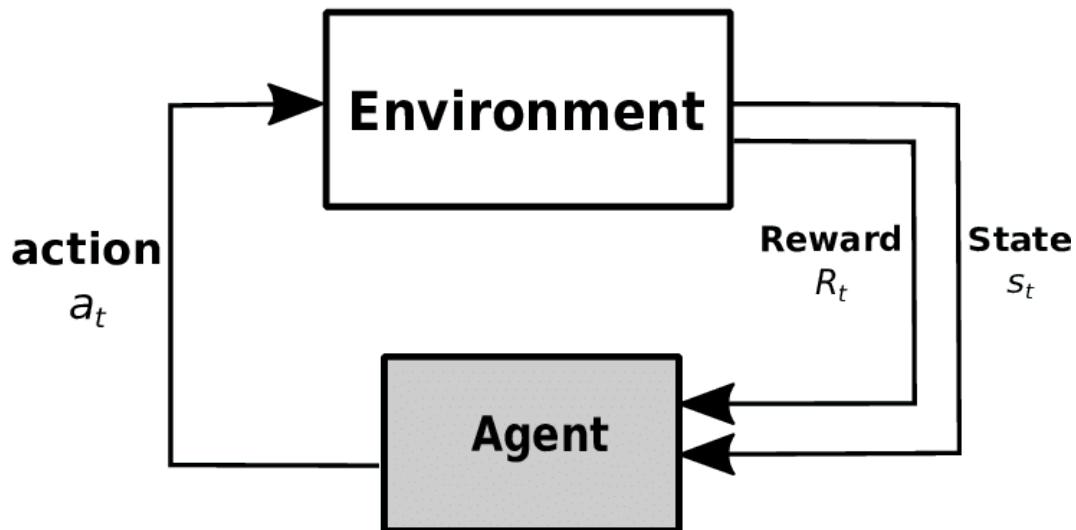


Explainable AI

Reinforcement learning explained



Artificial Intelligence, HBO-ICT

by

Finn Alberts (2062662)
Laurent Dassen (1970133)
Noud Wijngaards (1919954)

15-04-2022

Lectoraat Data Intelligence

supervisor(s)

Shannen Dolls, opdrachtgever,
Koen Steeghs, procesbegeleider

Abstract

Reinforcement learning (RL) is een van de nieuwere technieken op het gebied van artificial intelligence (AI). Deze techniek is nog niet erg bekend bij het grotere publiek. Om het vertrouwen en de kennis van de mogelijkheden van deze techniek te vergroten, is het belangrijk om duidelijkheid te creëren over de werking van RL. Het doel van dit project is dan ook het op een laagdrempelige manier uitleggen van RL middels explainable AI. Explainable AI houdt in dat AI wordt uitgelegt aan de hand van resultaten van demonstrators. Middels een explainer video, waarin het concept van RL wordt uitgelegd met onder andere een Pacman demonstrator, en een demonstrator waarin men zelf met RL aan de slag kan, wordt een bijdrage geleverd aan het verbreden van de kennis over RL.

1

Introductie

Aanleiding

Een onderdeel van de digitale revolutie is de toepassing van "Artificial Intelligence" (AI), het toepassen van intelligentie binnen machines. Door de opkomst van AI kunnen niet alleen 'simpele', gestructureerde taken uitgevoerd worden door machines, maar ook taken welke een menselijke gedachtegang vereisen.

Een manier om artificial intelligence toe te passen is door het gebruik van "reinforcement learning". Reinforcement learning combineert de voordelen van leren, zoals de nieuwsgierigheid en de aanpasbaarheid van de mens, en de voordelen van hoe machines leren, door snelheid en systematiek. Doordat een heel systeem gesimuleerd wordt, kan de zogenaamde "agent" (het AI-systeem) nieuwe acties of aanpakken uitproberen, zich aanpassen wanneer problemen voorkomen en verder bouwen zodra succes wordt geboekt [21]. Op deze wijze kan de agent verschillende soorten activiteiten op efficiënte wijze aanleren, zoals het (uit)spelen van videospellen, problemen gerelateerd aan verkeer oplossen en het verwerken van data. [9]

Een grote valkuil binnen AI is dat mensen niet de gedachtegang van de zogenaamde "agent" (het AI-systeem) kunnen begrijpen; oftewel waarom bepaalde keuzes worden gemaakt door de agent [24]. Dit kan ertoe leiden dat mensen het systeem niet vertrouwen [25]. Vooral nu AI, waaronder reinforcement learning in het bijzonder, steeds meer gebruikt wordt en steeds populairder wordt, is het van belang om het algemene publiek te voorzien van uitleg over dergelijke concepten en de toepassing ervan [26]. Om hiervoor te zorgen wordt gebruikgemaakt van "explainable AI". Explainable AI is een methode waarmee menselijke gebruikers de resultaten en output van machine learning-algoritmen beter kunnen begrijpen.

Tijdens de open dag op Zuyd Hogeschool in november 2021 zijn enkele voorbeelden getoond over hoe reinforcement learning is toegepast om activiteiten uit te voeren. Door het tentoonstellen van AI-demonstrators worden de mogelijkheden en resultaten van reinforcement learning duidelijker. Bovendien geeft dit de mogelijkheid om meer uitleg te bieden over artificial intelligence, een onderwerp waarbij blijkt uit onderzoek dat er steeds meer interesse voor is, maar waarbij uit de open dag blijkt dat er nog vrij weinig kennis hierover is bij het algemene publiek [8]. Vanuit het lectoraat Data Intelligence van Zuyd Hogeschool is gevraagd om verder te werken aan nieuwe demonstrators om de mogelijkheden van AI te kunnen blijven demonstreren.

Om het concept van artificial intelligence en reinforcement learning duidelijk over te brengen wordt gebruikgemaakt van een herkenbaar spel: Pacman. Het concept en het doel van Pacman is duidelijk te herkennen, waardoor nauwelijks uitleg is vereist met betrekking tot de werking van het spel. Daarnaast is het spel niet erg complex, waardoor het voor een agent makkelijker te leren zou moeten zijn.

Doelstelling

Het doel is om met explainable AI men reinforcement learning op een laagdrempelige manier uit te leggen.

2

Theoretisch kader

Pacman

Pacman is een klassiek computerspel dat is uitgebracht op 22 mei 1980. Het spel is ontwikkeld door Namco. In Pacman wordt er gespeeld als een geel karakter die bolletjes moet opeten om punten te krijgen. Naast deze bolletjes zijn er ook nog een aantal vruchten die opgegeten kunnen worden. Er zitten ook een aantal spookjes in de game. Deze maken het speelveld onveilig en kunnen ervoor zorgen dat Pacman gepakt wordt en hierbij een leven verliest. Pacman heeft een specifiek aantal levens. Het doel van het spel is om alle bolletjes op te eten zonder alle levens te verliezen. [6]

Artificial Intelligence

Artificial Intelligence (AI) houdt zich bezig met het creëren van een artefact dat een vorm van intelligentie vertoont. Dit artefact of agent kan via verschillende methoden slimmer worden. Hierdoor kunnen veel processen geautomatiseerd worden met AI. [3]

Machine Learning

Machine Learning is de studie van computeralgoritmen die zichzelf kunnen verbeteren aan de hand van ervaring en data. Machine learning is onderdeel van artificial intelligence [5]. Door middel van machine learning kan een AI-agent zich "aanpassen aan nieuwe omstandigheden en kunnen patronen gedetecteerd en geëxtrapoleerd worden" [23].

Reinforcement Learning

Reinforcement learning is een van de drie basismethoden die onder machine learning vallen. De andere twee zijn supervised en unsupervised learning. Met reinforcement learning wordt de AI agent in een omgeving gezet waar de agent acties moet uitvoeren. Deze methode werkt met een penalty-reward systeem. Wanneer een goede actie wordt uitgevoerd krijgt de agent een reward. Wanneer een slechte actie wordt uitgevoerd krijgt de agent een penalty. Hiermee kan de agent leren van fouten en ervoor zorgen dat het iedere keer beter wordt. De acties waarvoor de agent een reward of penalty krijgt, worden vastgelegd in een rewardfunctie. Het doel van de agent is om te bepalen wat de optimal policy is om een zo hoog mogelijke reward te behalen. De optimal policy geeft voor iedere situatie aan wat de beste keus is om te maken in die situatie. [13]

Explainable AI

Explainable AI is kunstmatige intelligentie waarbij de resultaten van de oplossing door mensen kunnen worden uitgelegd en begrepen. In andere woorden, het kunnen verwoorden en bewijzen hoe een agent tot bepaalde resultaten komt. [4]

3

Methode

Tijdens de tien weken van dit project is volgens een iteratieve werkwijze gewerkt. Deze werkwijze is afgeleid van agile [7] werken en heeft in combinatie met wekelijkse communicatie met de opdrachtgever geleid tot een iteratief proces. Er is gewerkt met een kanban bord, bestaande uit vijf kolommen: backlog, to do (deze week), doing, done en done (vorige weken). De backlog bevatte de nog uit te voeren taken. De to do-kolom werd wekelijks gevuld vanuit de backlog met de taken voor die week. De doing-kolom gaf de taken weer waar op dat moment aan gewerkt werd en de twee done-kolommen gaven weer welke taken klaar waren in die week en de vorige weken. Aan het einde van iedere week werd geëvalueerd en afgesproken wat de taken voor de komende week werden. Tijdens de wekelijkse bijeenkomsten met de opdrachtgever werd de voortgang gevalideerd. Zo werd gewaarborgd dat de producten in lijn waren met de wensen van de opdrachtgever. In deze bijeenkomsten werden ook extra eisen en wensen van de opdrachtgever opgehaald. Deze werden vervolgens in de backlog geplaatst.

Voor dit project is een system requirements specificatie (SRS) opgesteld. Dit SRS bevat de eisen van de op te leveren producten geprioriteerd met de MoSCoW-methode [18]. Deze eisen zijn gevalideerd door de opdrachtgever. Een van deze producten is de Pacman demonstrator, deze is onderdeel van de explainable AI. Aan de hand van de eisen is een componentendiagram gemaakt voor deze Pacman demonstrator. De demonstrator is gerealiseerd in Python met behulp van baselines3 [10] en OpenAI Gym [20]. Beiden zijn Python-libraries.

Het presteren van de agent in de Pacman demonstrator is getest aan de hand van verschillende reward functies. De resultaten hiervan zijn geplot in een grafiek (over de horizontale as de runs door de tijd heen, over de verticale as de behaalde rewards). Deze grafieken kunnen tevens worden ingezet voor het illustreren van het leerproces van reinforcement learning en kan daarmee dus worden ingezet voor explainable AI.

Om de demonstrator en de testresultaten goed toe te kunnen lichten is een explainer video gerealiseerd. Deze video is gemaakt met de hulp van een expert op het gebied van communicatie & multimedia design. Deze video licht het concept van reinforcement learning toe aan de hand van onder andere de Pacman demonstrator. Daarnaast is er nog een custom rewards demonstrator geschreven zodat men 'hands-on' aan de slag kan met reinforcement learning op IT-gebied. Deze demonstrator is ook gerealiseerd in Python, net zoals de Pacman demonstrator.

Door de combinatie van de explainer video met deze demonstrator wordt ook vanuit educatief perspectief een goede uitleg gegeven over reinforcement learning [22].

4

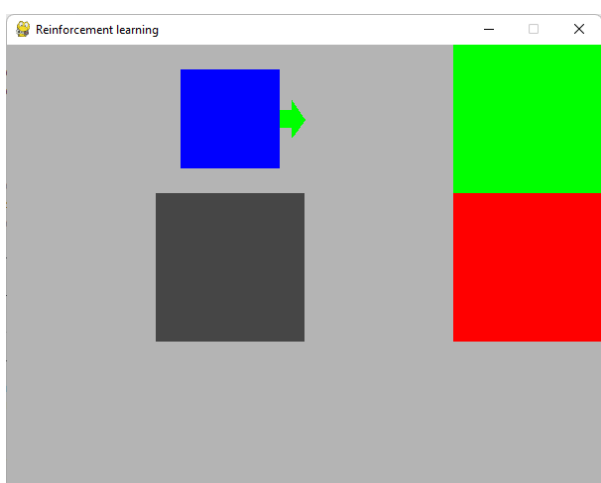
Resultaten

Beschrijving artefact

Het artefact binnen dit project bestaat uit een tweetal onderdelen, welke samen zorgen voor explainable AI. Het eerste onderdeel hiervan is de explainer video, waarin kunstmatige intelligentie, reinforcement learning en de mogelijkheden en valkuilen hiervan worden uitgelegd op een laagdrempelige manier.

Het tweede onderdeel is een applicatie, de custom rewards demonstrator, waarin een gebruiker inspraak heeft op de opbouw van de rewardfunctie voor een simpele simulatie. Aan de hand hiervan wordt de optimal policy gevisualiseerd. Op deze manier wordt de impact van de keuze voor een rewardfunctie duidelijk aan iemand op een wederom laagdrempelige manier. Een schermafbeelding van deze demonstrator is te zien in figuur 4.1.

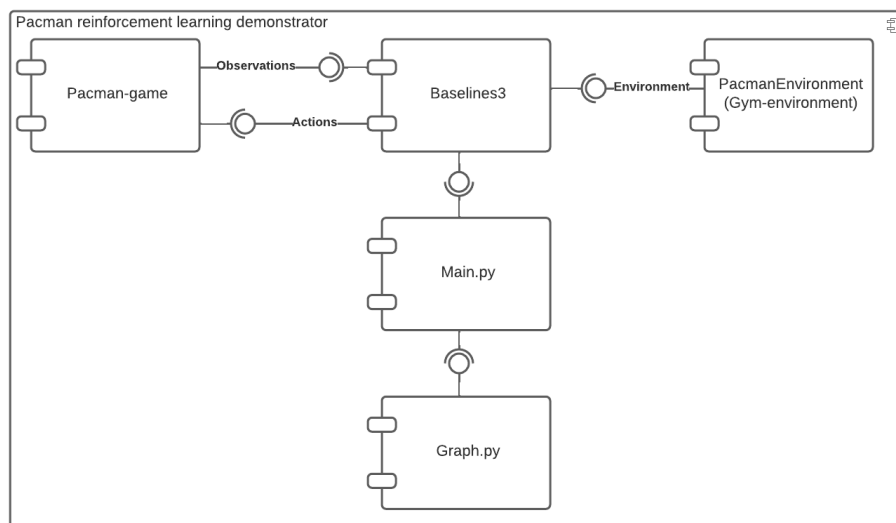
Figure 4.1: Custom rewards demonstrator



Pacman demonstrator

De Pacman demonstrator is gerealiseerd in de programmeertaal Python. Er is gebruik gemaakt van een Python-versie van Pacman [1], daar dit de implementatie vereenvoudigde. Deze versie van Pacman heeft geen verschillen ten opzichte van de originele Pacman, met uitzondering van kleine visuele verschillen. Voor het realiseren van reinforcement learning is gebruik gemaakt van baselines3 [10] en OpenAI Gym [20]. Beiden zijn Python-libraries. Met behulp van deze libraries is een PPO-algoritme (Proximal Policy Optimization) geïmplementeerd, waarmee een agent wordt getraind. Ter visualisatie van de progressie van de agent wordt een grafiek weergegeven tijdens het trainen met het PPO-algoritme. In figuur 4.2 is een overzicht te zien van de verschillende onderdelen van de demonstrator.

Figure 4.2: Componentdiagram

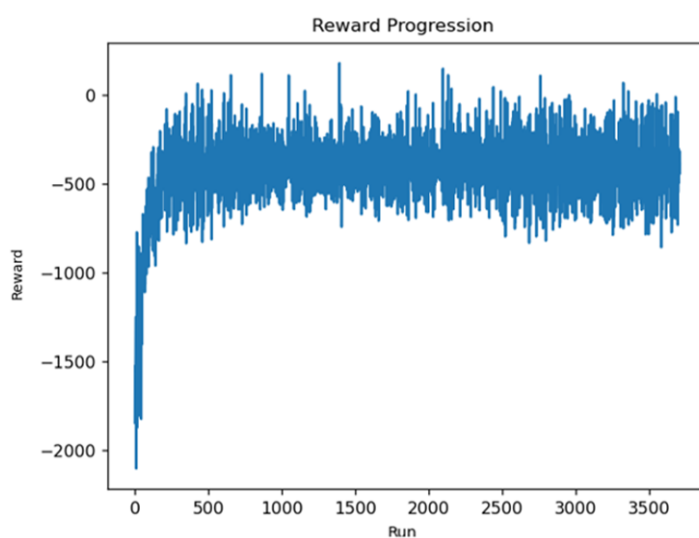


Tijdens het testen werd duidelijk dat Pacman te complex is voor reinforcement learning (zie hoofdstuk Testresultaten). Daarom is ervoor gekozen om de spookjes opgesloten te laten, waardoor het spel minder complex werd. De code van deze demonstrator is publiekelijk beschikbaar op Github [15].

Testresultaten

Uit de testen met verschillende rewardfuncties (met spookjes) werd duidelijk dat een erg snelle stagnatie van de rewards te zien was (zie figuur 4.3). Dit wil zeggen dat de agent zichzelf niet meer verbeterd. De grafiek in figuur 4.3 is het best verkregen resultaat gevisualiseerd (zie bijlage A voor de gebruikte reward functie). Deze stagnatie is kenmerkend voor reinforcement learning [16], maar voor Pacman vond deze al plaats voordat de agent daadwerkelijk voortgang had geboekt in het spel. Dit lijkt erop te wijzen dat reinforcement learning niet geschikt is voor Pacman. Dit wordt ondersteund door de literatuur, van onder andere de Universiteit van Stanford, waarin soortgelijke bevindingen zijn gevonden [12] [11]. Ook een artikel van de website Medium had eenzelfde waarneming [19].

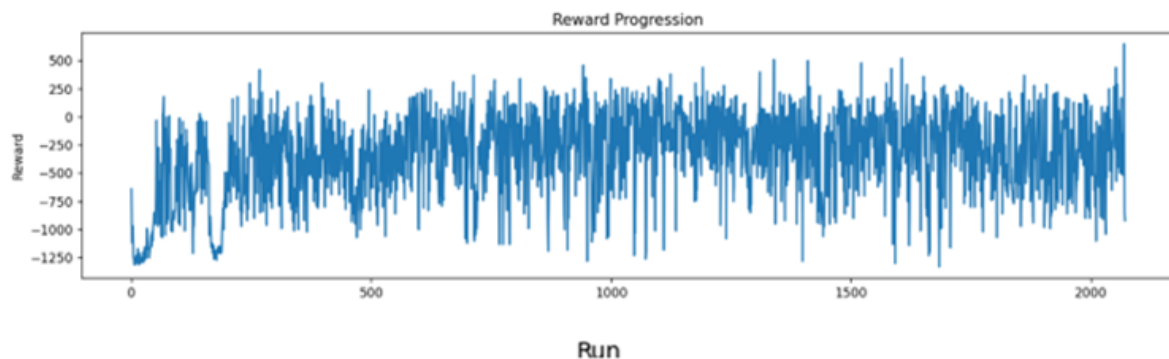
Figure 4.3: Rewardsgrafiek met spookjes



Vanwege deze bevindingen zijn ook tests uitgevoerd met een minder complexe vorm van Pacman, waarbij

de spookjes opgesloten bleven. Zo heeft de Pacman agent kunnen trainen zonder het risico om gepakt te worden door de spookjes. Echter was hier ook sprake van een stagnatie, zie figuur 4.4.

Figure 4.4: Rewardsgrafiek zonder spookjes



De reden waarom reinforcement learning niet geschikt is voor Pacman, is niet duidelijk. Mogelijk is de omgeving van Pacman te dynamisch, doordat de bewegingen van de spookjes voor een deel afhangen van de acties van de speler. Dit verklaart echter niet waarom de agent slecht presteert, wanneer de spookjes geen rol spelen. Een mogelijke verklaring hiervoor is dat de grootte van de omgeving (grid van 31 bij 28) het spel te complex maakt voor de agent.

Andere technieken

Het is voor een agent mogelijk om Pacman succesvol te spelen met andere technieken dan reinforcement learning. Zo is in de literatuur een implementatie te vinden met neurale netwerken [2] en heeft Microsoft Maluuba middels een techniek genaamd HRA (Hybrid Reward Architecture) een succesvolle AI kunnen bouwen waarmee een maximale score binnen het spel "Ms. Pac-Man" is behaald [17]. Deze technieken zijn niet verder onderzocht.

Explainer video

In de gerealiseerde explainer video worden een aantal zaken uitgelegd op een laagdrempelige manier. Er wordt allereerst uitgelegd hoe reinforcement learning werkt. Vervolgens wordt toegelicht dat deze AI-techniek niet geschikt is voor ieder probleem. Ter illustratie van dit laatste wordt gebruik gemaakt van de Pacman demonstrator. Hierbij wordt tevens een korte vergelijking gemaakt, waarin te zien is dat de oplossingen met HRA en neurale netwerken wel succesvol zijn voor Pacman. Verder wordt toegelicht dat er andere problemen zijn, waarvoor reinforcement learning wel geschikt is, zoals Super Mario Bros. en Super Mario Kart.

Custom rewards demonstrator

De gerealiseerde custom rewards demonstrator is een applicatie waarin een gebruiker zelf kan experimenteren met verschillende rewardfuncties, om zo een gevoel te krijgen voor de impact van deze rewardfuncties op het gedrag van de agent. De gebruiker krijgt na een keuze voor een reward gemaakt te hebben, een simulatie te zien van het gedrag van de agent bij een optimale policy (horende bij de gekozen rewardfunctie). De code van deze demonstrator is beschikbaar via Github [14].

System requirements specificatie

Vanwege de iteratieve werkwijze en de wekelijkse bijeenkomsten met de opdrachtgever is het SRS een aantal keer bijgewerkt. Hierbij zijn eisen toegevoegd, verwijderd of aangepast. Het definitieve SRS is te zien in bijlage B

5

Discussie

Uit de tussenresultaten van dit project zou een conclusie kunnen worden getrokken dat reinforcement learning geen geschikte techniek is voor Pacman. Hierover kan echter geen uitsluitel worden gegeven. Het is mogelijk dat in de toekomst nieuwe reinforcement learning technieken het wel mogelijk zouden kunnen maken om een agent Pacman te leren spelen. Daarnaast is niet uit te sluiten dat met een oneindig lange trainingstijd de agent Pacman wel zou leren. De grafieken van de rewards suggereren echter wel dat de agent zich niet meer verbeteren.

De ontwikkelde explainer video is gemaakt zonder eerdere ervaring op dit gebied. Vanwege het gebrek aan deze ervaring is via de opdrachtgever contact opgenomen met een expert (afgestuurd in communicatie & multimedia design). Hij heeft ons middels feedback ondersteunt in het maken van de video. Desalniettemin blijft het ontwikkelen van een explainer video buiten onze expertise liggen en kan de kwaliteit van deze video dus niet worden vergeleken met video's van experts op dit gebied.

Tot slot was een van de eisen van de opdrachtgever (prioriteit could) dat de applicatie webbased deploybaar was. In overleg met de opdrachtgever is voor beide demonstrators echter anders besloten. Voor de Pacman demonstrator had dit als reden dat de realisatie van Pacman in Python voor webbased te complex was voor de gegeven tijd. Python was daarentegen wel zeer gewenst vanwege de beschikbare reinforcement learning libraries. De implementatie in Python had de voorkeur, omdat door de opdrachtgever materiaal ter beschikking in gesteld van vorige projecten die een vergelijkbare werkwijze hadden. De reden dat bij de custom rewards demonstrator voor Python was gekozen (wat wederom niet geschikt is voor webbased), is de korte resterende tijd voor oplevering.

6

Conclusie

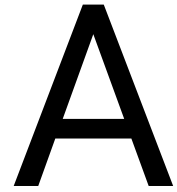
Aan de hand van explainable AI is een bijdrage geleverd aan het verbreden van kennis over reinforcement learning bij het grotere publiek. Deze explainable AI is opgebouwd uit een explainer video en een custom rewards demonstrator.

De explainer video geeft een algemene uitleg over reinforcement learning, gevolgd door de mogelijkheden en beperkingen ervan. De beperkingen ervan worden onder andere geïllustreerd aan de hand van de Pacman demonstrator, waarin te zien is dat de agent niet in staat is het spel succesvol te spelen, zelfs wanneer zonder spookjes wordt gespeeld. De explainer video laat daarnaast ook zien dat er, voor Pacman in het bijzonder, andere mogelijkheden zijn waarbij de agent wel succesvol is. Verder maakt de video aan de hand van Super Mario Bros. en Super Mario Kart ook duidelijk dat reinforcement learning voor andere problemen wel een geschikte oplossing kan zijn.

Ook is ter aanvulling op deze explainer video een custom rewards demonstrator gerealiseerd waardoor iemand zonder kennis op het gebied van AI (of IT in het algemeen) toch aan de slag kan met reinforcement learning. De combinatie van deze twee zorgt voor explainable AI wat ook vanuit een educatief perspectief een goede uitleg van reinforcement learning biedt. Hiermee is de doelstelling, het op een laagdrempelige manier uitleggen van reinforcement learning met behulp van explainable AI, bereikt.

Dit project houdt de deur open voor een aantal vervolgprojecten. Een van deze projecten zou een onderzoek kunnen zijn, waarin wordt onderzocht waarom reinforcement learning niet geschikt is voor ieder probleem (waaronder Pacman). Een ander vervolgproject kan een verder onderzoek naar HRA zijn (een techniek die nog wat minder bekend is) en hoe deze techniek breder kan worden ingezet. Ook het realiseren van een webbased demonstrator kan in een vervolgwerk worden uitgewerkt. Tot slot is geen field test uitgevoerd met de gerealiseerde explainable AI om te kunnen zien of deze ook daadwerkelijk geschikt is voor het begrip explainable AI. Deze field test wordt zeker aanbevolen.

Appendices



Bijlage A

Table A.1: Rewards en penalties in Pacman demonstrator

Actie	Reward / Penalty
Het behalen van punten	Reward: het aantal behaalde punten sinds het vorige frame
Het behalen van level 1	Reward: 10 000 punten
Het verstrijken van tijd (snellere runs zijn beter)	Penalty: 0,5 punten
Het indrukken van een knop	Penalty: 5 punten
Gepakt worden door een spookje	Penalty: 500 - de behaalde score * 0,338

B

Bijlage B

Table B.1: System requirements specificatie

Nummer	Eis	Prioriteit
1	De demonstrator kan Pacman spelen.	Must
2	De demonstrator werkt met reinforcement learning. Hierbij probeert de demonstrator een zo hoog mogelijke score te behalen.	Must
3	De demonstrator kan op een website worden weergegeven of in werking gebracht worden.	Could
4	De werking van de demonstrator wordt uitgelegd middels een video.	Should
4a	De gebruikte termen in de video zijn te begrijpen zonder voorkennis over het onderwerp te hebben.	Should
4b	De demonstrator wordt in de video vergeleken met Pacman AIs die gebruik maken van andere technologieën.	Should
4c	Het gebruikte materiaal in de video is royalty-free.	Must
5	De werking van de demonstrator wordt gedocumenteerd in een ontwerpdocument.	Must
5a	De gemaakte keuzes gedurende het project zijn onderbouwd in het ontwerpdocument.	Must
6	De deployment van de demonstrator wordt gedocumenteerd in een overdrachtsdocument.	Must
7	De uitprobeer-applicatie geeft een stimulatie van de optimale policy aan de hand van de ingegeven rewards.	Must
8	De gebruiker kan in de uitprobeer-applicatie zelf de rewards bepalen voor een rewardfunctie van een simpele stimulatie.	Must

Bibliography

- [1] Pacmancode, augustus 2021.
- [2] Code Bullet. Ai learns to play pacman using neat, april 2018.
- [3] Wikipedia Community. Kunstmatige intelligentie, januari 2021.
- [4] Wikipedia Community. Explainable artificial intelligence, april 2022.
- [5] Wikipedia Community. Machine learning, februari 2022.
- [6] Wikipedia Community. Pac-man, februari 2022.
- [7] Scrum Company. Wat is agile?
- [8] Wesley Gomes de Sousa, Elis Regina Pereira de Melo, Paulo Henrique De Souza Bermejo, Rafael Araújo Sousa Farias, and Adalmir Oliveira Gomes. How and where is artificial intelligence in the public sector going? a literature review and research agenda. *Government Information Quarterly*, 36(4): 101392, 2019. ISSN 0740-624X. doi: <https://doi.org/10.1016/j.giq.2019.07.004>. URL <https://www.sciencedirect.com/science/article/pii/S0740624X18303113>.
- [9] Ben Dickson. Reinforcement learning frustrates humans in teamplay, mit study finds, november 2021.
- [10] Antonin Raffin en Ashley Hill en Adam Gleave en Anssi Kanervisto en Maximilian Ernestus en Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, pages 1–8, 2021.
- [11] Amelia Christensen en Bernardo Ramos en Kushal Ranjan. Recurrent deep q-learning for pac-man. 2016.
- [12] Abeynaya Gnanasekaran en Jordi Feliu Faba en Jing An. Reinforcement learning in pacman.
- [13] Błażej Osiński en Konrad Budek. What is reinforcement learning. juli 2018.
- [14] Finn Alberts en Laurent Dassen en Noud Wijngaards. Finnalberts/custom-rewards-reinforcement-learning, april 2022.
- [15] Finn Alberts en Laurent Dassen en Noud Wijngaards. Finnalberts/pacman-reinforcement-learning, april 2022.
- [16] Joeran Beel Keith Tunstead. Combating stagnation in reinforcement learning through ‘guided learning’ with ‘taught-response memory’. 2019.
- [17] Microsoft Maluuba. Hybrid reward architecture (hra) achieving super-human performance on ms. pacman, juni 2017.
- [18] Project Management. Moscow.
- [19] Medium. How to train ms-pacman with reinforcement learning, mei 2021.
- [20] OpenAI. Gym.
- [21] Gary Saarevirta. Why reinforcement learning will deliver what ‘ai’ promised. URL [https://www.daisyintelligence.com/blog/reinforcement-learning-ai#:~:text=Reinforcement%20learning%20delivers%20decisions.,successes%20\(or%20positive%20reinforcement\)](https://www.daisyintelligence.com/blog/reinforcement-learning-ai#:~:text=Reinforcement%20learning%20delivers%20decisions.,successes%20(or%20positive%20reinforcement)).
- [22] Ilse Sistermans. Video in education in the netherlands. juni 2017.

- [23] Peter Norvig Stuart Russell. *Artificial Intelligence: A Modern Approach*. Pearson Education Limited, 2010.
- [24] Sandhya Subramanyan. Everything you want to know about transparent and explainable ai. URL <https://www.smartkarrot.com/resources/blog/transparent-explainable-ai/>.
- [25] Violet Turri. Everything you want to know about transparent and explainable ai. URL <https://insights.sei.cmu.edu/blog/what-is-explainable-ai/>.
- [26] Nelson Vithayathil Varghese and Qusay H. Mahmoud. A survey of multi-task deep reinforcement learning. *Electronics*, 9(9), 2020. ISSN 2079-9292. doi: 10.3390/electronics9091363. URL <https://www.mdpi.com/2079-9292/9/9/1363>.