Deep Reinforcement Learning

Cindy Waeltermann (C)
Contingent Worker

What Is Deep Reinforcement Learning?

Deep Reinforcement Learning (DRL) is a subfield of Machine Learning (ML) where intelligent agents can learn from past decisions and use that as a basis for future decision-making. While deep reinforcement learning is relatively new, both Reinforcement Learning (RL) and Deep Learning (DL) can be traced back to the 1940s and 1950s when researchers started to explore learning through trial and error by giving feedback in the form of rewards or punishments.

It wasn't until the 1980s that the concept of reinforcement learning became a potential reality. Early algorithms were relatively simple and were limited in their ability to learn complex behaviors.

If you remember the movie War Games, reinforcement learning taught the WOPR supercomputer to run through all possible scenarios for nuclear war based on algorithms that use rewards and punishments for machine learning. In the end, WOPR, having learned from past decisions and the consequences of those decisions, decides that all moves lead to global annihilation and concludes that "The only winning move is not to play." While a bit out of the realm of reality, the movie does make a great analogy.

It is important to understand the origins of the different types of machine learning. The origins of Deep Learning (DL) trace back to the 1940s with the development of Artificial Neural Networks (ANNs). These networks played a crucial role in the development of reinforcement learning as well as deep learning.

A neural network is a deep learning algorithm inspired by the human brain. The human brain consists of interconnected nodes, called neurons, that work together to process and learn from data input. While one neuron can take input values and produce an output value, that output can be connected to the input of another neuron, creating a network of interconnected neurons, much like the human brain. Those interconnected neurons can learn from experience and adjust behavior accordingly, and they can also recognize patterns and make predictions based on those patterns. These networks can range from simple networks with only a few layers to extremely large and complex networks with hundreds of layers and millions of neurons.

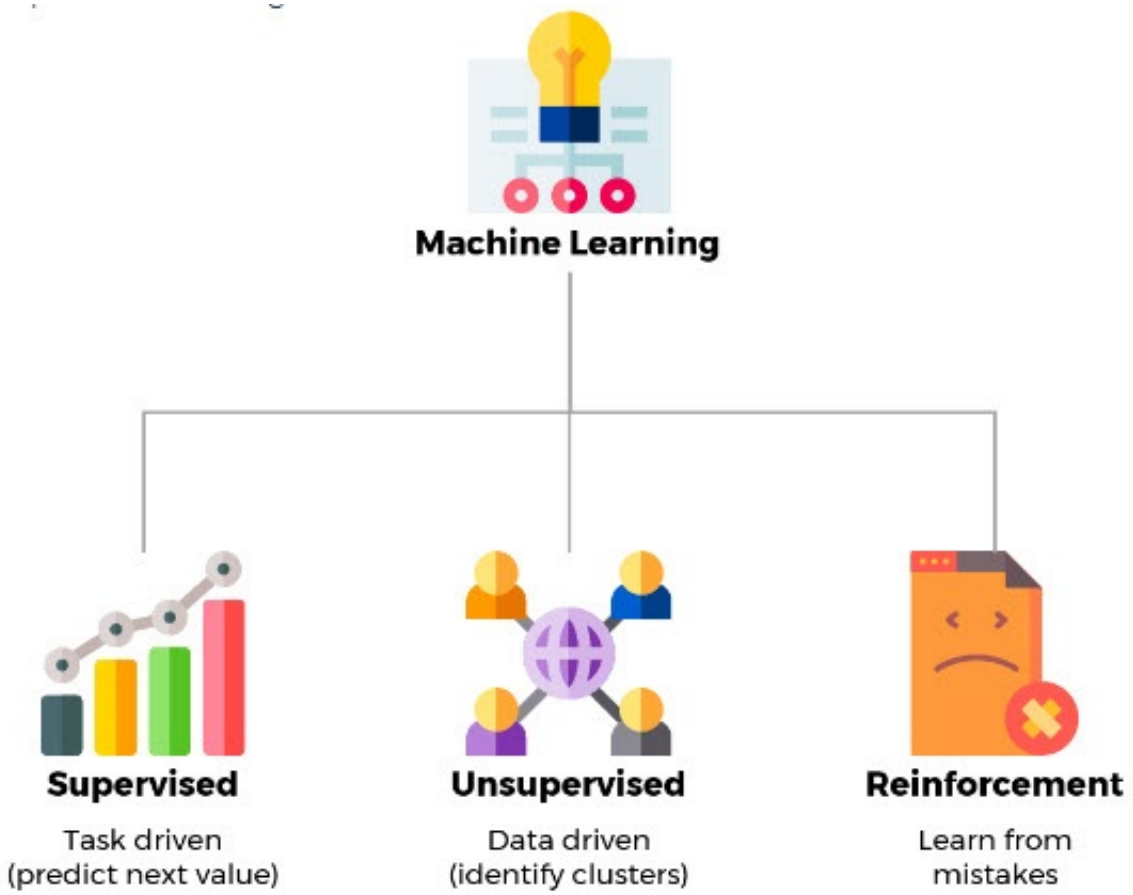
When a human makes a decision, many factors play into the decision-making

process. Not only do past experiences come into play, but also ambiguous factors such as emotion, reasoning, intuition, and memory, to name a few. These factors can influence decision-making just as much as past experiences in humans. Sometimes more.

In a machine learning model that uses deep reinforcement learning, a dataset is used to train the AI component to make decisions based on how past decisions were handled. With AI and deep reinforcement learning, past experience is based on a set of data. Past experience is limited to only to the size of the data that was input, so the larger the data set, the more learning can be applied to the model.

Types of Machine Learning

Because deep reinforcement learning combines the principles of both reinforcement learning and deep learning, it is important to understand the different types of machine learning and the fact that they have different goals and approaches in problem-solving. There are several types of learning techniques applied to models in ML and AI.



Supervised Learning

In Supervised Learning (SL), an algorithm is trained on labeled data and creates mapping between input and output variables. Mapping between input and output variables refers to the relationship between the input data and the corresponding output data that the machine learning algorithm is designed to learn. Supervised Learning feeds historical input and output data in machine learning algorithms, with processing in between each input/output pair that helps the model to predict output on the basis of prior experiences. Supervised learning can be used for image classification, Natural Language Processing (NLP), or in models that use regression to predict values.

Unsupervised Learning

In Unsupervised Learning, models identify patterns and structure within an

unlabeled dataset. The ML algorithm dataset has no predefined output values and creates organization within the data. Unsupervised learning is useful when labeled data is unavailable or where the goal is pattern identification. The most common applications of unsupervised learning are clustering and association problems. Clustering produces a model that groups objects based on certain properties, such as color. Association takes those clusters and identifies rules that exist between them.

Reinforcement Learning

Reinforcement learning (RL) is based on rewarding desired behaviors or punishing undesired ones. Instead of one input producing one output, the algorithm produces a variety of outputs and is trained to select the right one based on certain variables. So, for example, a computer program could be trained to win a video game by identifying patterns in the actions that lead to it scoring more points than the other players.

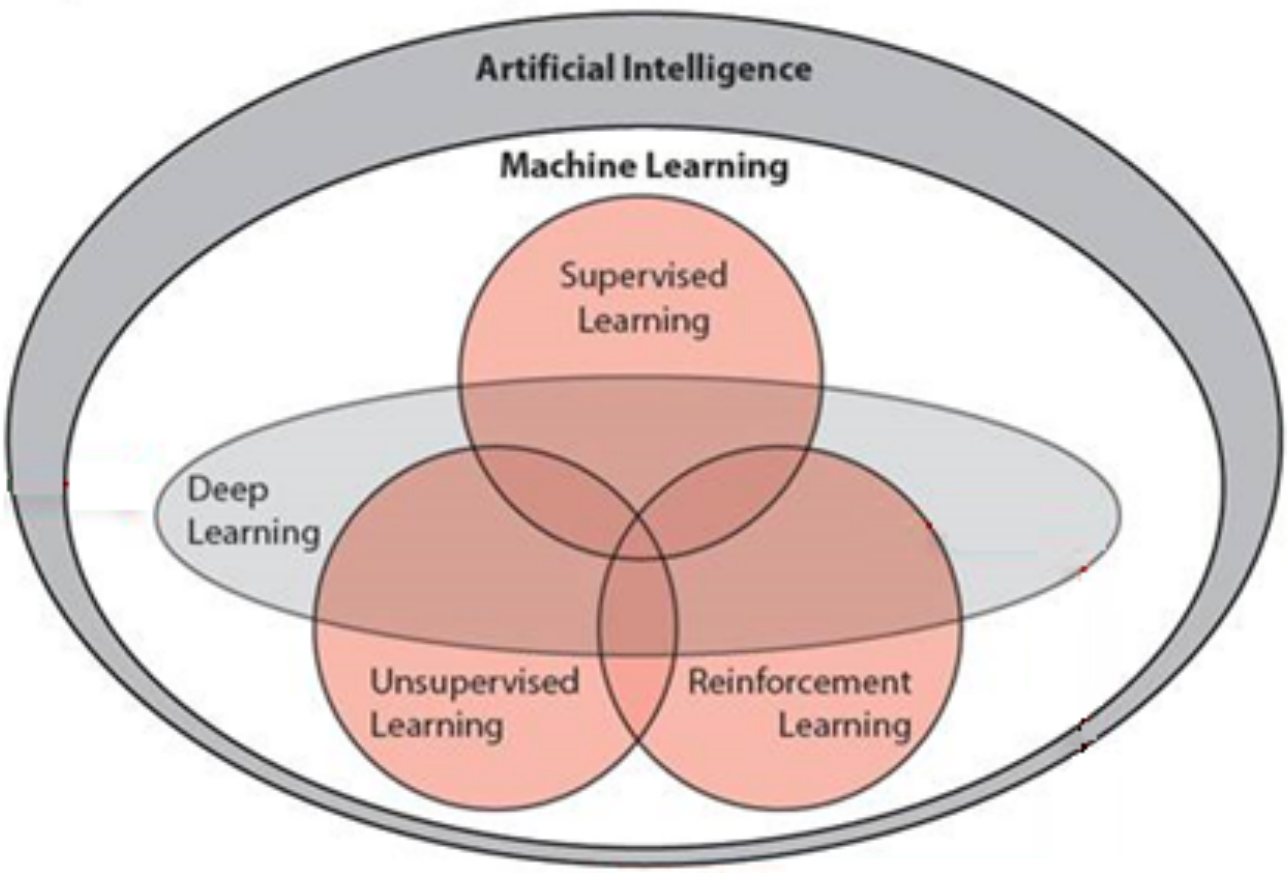
Deep Learning

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain, allowing it to “learn” from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Deep Reinforcement Learning

Deep Reinforcement Learning is the combination of reinforcement learning (RL) and deep learning (DL) that enables machines to solve complex decision-making tasks. Deep Reinforcement Learning introduces deep neural networks to solve Reinforcement Learning problems — hence the name "deep".



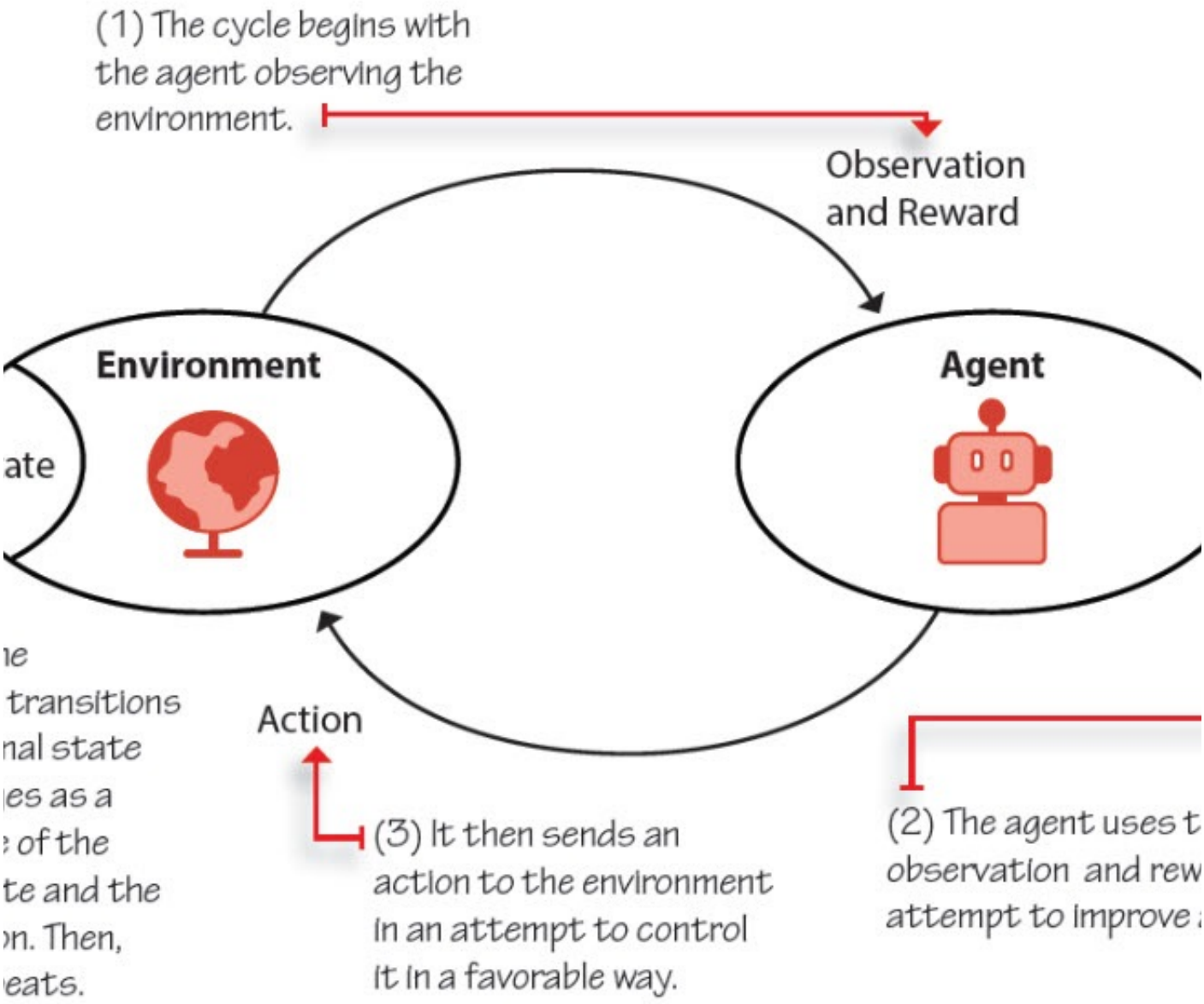
Deep reinforcement learning involves the utilization neural networks. It is a collection of techniques and methods for using neural networks to solve ML tasks, whether SL, UL, or RL. DRL is simply the use of DL to solve RL tasks.

This is the task of deciding, from experience, the sequence of actions to perform in an uncertain environment in order to achieve some goals. The main idea is that

an artificial agent may learn by interacting with its environment, similarly to a biological agent. Using the experience gathered, the artificial agent should be able to optimize some objectives given in the form of cumulative rewards. This approach applies in principle to any type of sequential decision-making problem relying on past experience.

The reinforcement process is a loop that outputs a sequence of state, action, reward, and next state. There are several fundamental elements in reinforcement learning models:

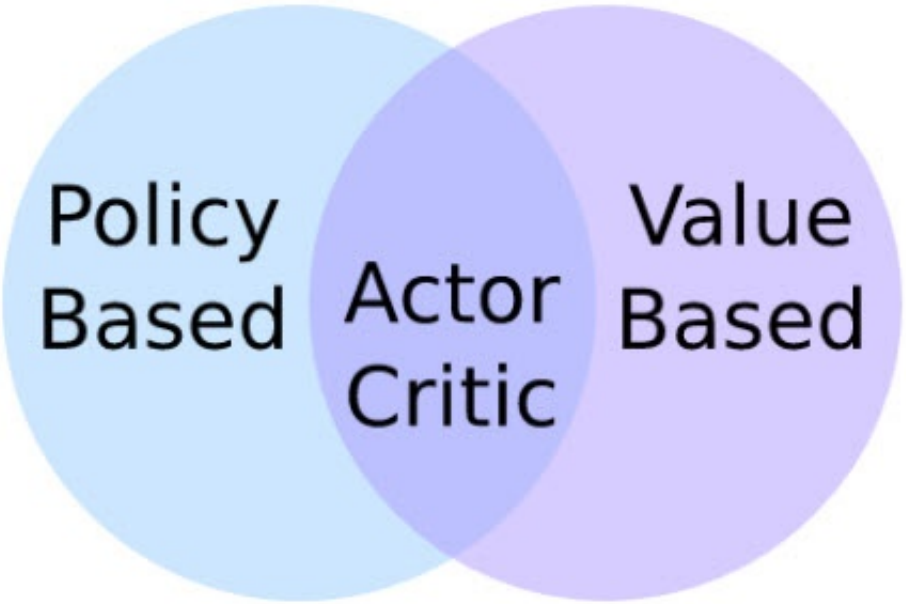
1. The agent receives an observation from the environment.
2. The agent uses observation and reward to improve the task.
3. The agent sends an action to the environment.
4. The environment changes as a consequence of the previous state and the agent's action.
5. The cycle repeats again from Step 1.



To calculate the expected cumulative reward (expected return), the rewards that come sooner (at the beginning of the game) are more probable to happen since they are more predictable than the long-term future reward.

An optimal policy is the "brain" of the agent, which determines the action to take (based on a state) and produces the actions that maximize the expected return. There are three main approaches for using neural networks to learn optimal policies: value-based, policy-based, and actor-critic methods.

- **Value-based**



methods focus on estimating the value of each possible action in each possible state and selecting the action with the highest expected value. This involves estimating the "value function" for each state, which is the expected cumulative reward that can be obtained from that state by following a particular policy. Value-based methods are often used in reinforcement learning, where an agent learns to make decisions by trial and error through interacting with an environment.

- **Policy-based methods** focus on directly optimizing the policy itself, without estimating the value function. This involves searching for the policy

that maximizes the expected cumulative reward over time. Policy-based methods are often used in optimization problems where the state space is continuous and the action space is large, such as in robotics or control systems.

- **Actor-critic methods** combine both value-based and policy-based methods. In actor-critic methods, the agent maintains two separate models: an actor that selects actions based on the current state, and a critic that estimates the value of each state-action pair. The actor selects actions based on a policy that is updated based on the estimated value of each state-action pair provided by the critic. The critic, in turn, learns by minimizing the difference between its estimated values and the actual observed rewards received by the agent. The actor-critic methods directly optimize the policy.

Algorithms used in Deep Reinforcement Learning

In deep reinforcement learning, the Q-value and function are important concepts that relate to the prediction of an expected future reward for a given action in a particular state.

- **Q-value** is a function that estimates the expected future reward that an intelligent agent will receive by taking a particular action in a given state. It takes in the current state and the action, and outputs the expected future reward for that action in that state. The Q-value function is often represented by a neural network, known as a Q-network.
- The **value function** estimates the expected future reward for a given state. The value function takes in the current state and outputs the expected future reward for that state. The value function is also often represented by a neural network, known as a value network.

The Q-value and value functions are related because the Q-value of a particular action in a given state can be determined by adding the expected reward for that action to the expected future reward for the resulting state. In other words, the Q-value of an action is the sum of the immediate reward plus the value of the

resulting state.

In deep reinforcement learning, algorithms are used to help intelligent agents learn from their interactions with an environment and to make decisions and take actions that maximize its reward over a period of time. The algorithms check out the environment, learn what it can from it, generalize that learning and apply it, and receive a reward. (Rewards may not be immediate but may occur after a sequence of events.) In other words, they help intelligent agents to learn and adapt to an environment in order to make decisions and take actions that maximize its cumulative reward, even in complex and changing environments.

Deep reinforcement learning typically uses a combination of algorithms such as:

- **Q-learning** – updates the Q-value of each state-action pair based on the reward obtained and the maximum Q-value of the next state. (Value Based)
- **Deep Q-Network (DQN)** – uses deep neural networks to approximate Q-function. (Value Based)
- **Policy Gradient** – sets the policy or actions taken in each state to maximize the expected cumulative reward. (Policy Based)
- **Actor-Critic** – hybrid algorithm that combines policy gradient and value-based methods with separate modules for policy and value estimation. (Actor
- **Proximal Policy Optimization** – policy gradient algorithm that uses a trust region optimization approach. (Policy Based)

Implementing deep reinforcement learning is a complex process that requires expertise in machine learning, data science, computer science, and domain-specific knowledge. It is typically implemented with high performance platforms like GPUs or cloud because of the resources it requires.

How is deep reinforcement learning being used?

Deep Reinforcement Learning is currently being used to develop intelligent

agents that can make decisions based on trial and error. The intelligent agent interacts with an environment by selecting actions based on the current state of the environment. The environment provides feedback to the agent in the form of rewards – which are used to update the agent's decision-making strategy by learning policies that map states to actions so that the expected cumulative reward is maximized over time.

One of the most famous deep reinforcement learning use cases was the AlphaGo AI program from DeepMind that beat a human world champion in 2016 in the game of Go. Go is considered to be both simpler than chess (fewer rules but higher number of potential moves)- 10360 for Go versus 10123 for chess.

In 1997, Deep Blue from IBM beat Gary Kasparov at the game of chess. Chess is known as the most challenging classical game for artificial intelligence because of its complexity. It was taught the rules of chess along with millions of possible board configurations on deep neural networks. These networks take a description of the board as input and process it through a number of different network layers containing millions of neuron-like connections.

One neural network, the "policy network", selects the next move to play. The other neural network, the "value network", predicts the winner of the game. AlphaGo was introduced to numerous amateur games to help it develop an understanding of reasonable human play. Then it was taught to play against different versions of itself thousands of times, each time learning from its

mistakes. Over time, AlphaGo improved and became increasingly stronger and better at learning and decision-making, and eventually beat the world chess champion. It was a major milestone in AI, demonstrating that machine learning algorithms could master a complex game. So much for playing that nice game of chess, eh?

The following companies use deep reinforcement learning for consumer products:

DeepMind	DeepMind is a leader in the field of DRL and has example is their AlphaGo program, which used D
cognitivecare	Cognitive Care is a healthcare product that uses course of action for a patient based on individua generated by healthcare systems, Cognitive Car optimize clinical workflow.
tickeron	Tickeron is a financial analytics platform that us reinforcement learning (DRL), to provide investo
Watson	Watson is an AI system developed by IBM. It was top of Advanced Natural Language Processing a humans in America's popular quiz game show n game and won a prize worth $1 million.

	Meta uses deep reinforcement learning to train a for optimizing content delivery on social media p reinforcement learning to adjust the graphics qu of the device. Meta's smart camera uses DRL to

Deep reinforcement learning tools are used in a wide range of software applications. As a rapidly evolving field, new software and applications that offer tools for deep reinforcement learning are being developed on a constant basis.

Popular deep reinforcement learning tools for developers:

	TensorFlow is an open-source machine learning learning techniques, including DRL. TensorFlow Network (DQN) algorithm.
	PyTorch is another popular open-source machin DRL models and has become quite popular amo
	OpenAI Gym is a toolkit for developing and comp environments and benchmarks for evaluating DF
	RLlib is an open-source library for reinforcement tools for building DRL models, including support

	Unity ML-Agents is a toolkit for developing and t supports DRL algorithms and enables developer environments.

How does Deep Reinforcement Learning benefit businesses?

There are multiple use cases for deep reinforcement learning across a broad spectrum of industries:

- **Robotics** – deep reinforcement learning is used to train robots to perform tasks that require multiple steps, such as assembling a product. The robot can learn the sequence of actions required to complete the task and adjust its actions based on the feedback it receives.
- **Finance** – deep reinforcement learning is being used to predict market trends, and to develop trading strategies for stocks and other financial use cases. (You can read more about Algorithmic Trading here.)
- **Healthcare** – deep reinforcement learning is being used to develop personalized treatment plans based on medical history and current health conditions. In fact, there is a company, PathAI, that has already developed several products for diagnostic and clinical development.
- **Self-driving cars** – deep reinforcement learning is used to train autonomous vehicles to navigate through complex traffic scenarios.
- **Natural Language Processing (NLP)** – deep reinforcement learning is being used to develop chatbots and virtual assistants.
- **Industrial Automation** – deep reinforcement learning is being used to optimize production processes. Startups like Covariant, Ocado's Kindred and Bright Machines are using deep reinforcement learning to change how machines are controlled in factories and warehouses, solving inordinately difficult challenges such as getting robots to detect and pick

up objects of various sizes and shapes out of bins, among others.
- **Marketing** – deep reinforcement learning is being used to personalize marketing campaigns by learning from customer behavior and interactions. For example, deep reinforcement learning is used to develop personalization engines that help marketers deliver hyper-personalized content. Check out this article in the Harvard Business Review to learn more about how AI is influencing the marketing and sales sector.

Drawbacks of Deep Reinforcement Learning algorithms

There are several drawbacks to using Deep Reinforcement Learning in AI software products. Deep reinforcement learning algorithms are incredibly complex and require large amounts of data and computing resources, including vast neural networks and GPUs. They also have limited ability to generalize learned tasks and the models can be difficult to interpret.

Startup activity

New AI companies that use deep reinforcement learning are popping up at a fever pitch. Since the release of the GPT-3 model, it appears that all industries are moving full steam ahead with AI innovation.

	Osaro is a San Francisco-based startup tha industries such as logistics and manufactu

	OspreyData is a startup that uses DRL to o platform can help oil and gas companies in
	Viz.ai is a startup that uses DRL to build int platform can help doctors diagnose and tre
	Nauto is a startup that uses DRL to build in powered platform can help fleet operators

Competitor activity

Avalara

While Avalara does use AI and ML techniques in some of their products, these techniques are generally focused on automating tax calculations, identifying

potential exemptions, and managing tax compliance processes and do not utilize deep reinforcement learning.

Sovos

Sovos, like Avalara, uses AI for tax compliance products. They do not have any deep reinforcement learning products at this time.

Thomson Reuters

Like Avalara and Sovos, Thomson Reuters has tax solutions that utilize AI and ML for compliance purposes. They do not seem to have any AI products that utilize deep reinforcement learning.

Stripe (TaxJar)

Stripe has several AI-based fraud detection features, such as Radar for Fraud Teams, which uses machine learning algorithms to analyze transactions and identify suspicious patterns. Additionally, Stripe has developed an AI-powered tool called Sigma that can analyze a business's financial data to help identify opportunities for growth or cost savings.

Wolters Kluwer

Wolters Kluwer Tax & Accounting has developed some AI-powered solutions such as CCH Tagetik, which is a financial planning and analysis software that uses machine learning algorithms to make better financial decisions and forecasts. However, it is unclear whether or not CCH Tagetik uses deep reinforcement learning.

Potential impact for Vertex

Deep Reinforcement Learning is at the center of many artificial intelligence initiatives because of its vast potential to transform the way we live. From autonomous vehicles to robotics to health care – the potential for impact is

profound.

Conclusion

As Vertex becomes more experienced with defining value functions related to complex issues like tax compliance, Deep Reinforcement Learning may be able to play a role. For example, if we can track notices from jurisdictions indicating errors and score this against the transactions and tax returns that led to the errors, DRL might be used to monitor and uncover patterns of errors to help analysts improve compliance over time.

References

- What is a Neural Network (IBM)
- Understanding the 3 Key Types of Machine Learning
- An Introduction to Deep Reinforcement Learning
- Hugging Face AI – The Reinforcement Learning Network
- Introduction to RL and Deep Q Networks
- Policy-Based Methods
- Policy Gradients in a Nutshell
- Understanding Actor-critic Methods
- Diagnostic Development – Path AI
- Clinical Development – Path AI
- Covariant
- Ocado's Kindred
- Bright Machines
- DRL in Industrial automation
- Harvard Business Review
- https://arturo.ai/
- Neuromation
- https://deep6.ai/
- Osaro
- OspreyData
- RLlib

- Unity ML-Agents
- DeepMind
- TensorFlow
- PyTorch
- OpenAI Gym
- RLlib
- https://www.nauto.com/
- Unity ML-Agents
- Future of AI & ML in TaxTech space (Avalara)
- Avalara Managed Tax Category Classification Simplifies Product Classifications and Taxability Determinations for Businesses (Avalara)
- https://livebook.manning.com/book/grokking-deep-reinforcement-learning/chapter-1/v-14/23