*Artificial intelligence in business is the use of AI tools such as machine learning, natural language processing, and computer vision to optimize business functions, boost employee productivity, and drive business value.*

*Artificial intelligence, or the development of computer systems and machine learning to mimic the problem-solving and decision-making capabilities of human intelligence, impacts an array of business processes. Organizations use artificial intelligence (AI) to strengthen data analysis and decision-making, improve customer experiences, generate content, optimize IT operations, sales, marketing and cybersecurity practices and more. As AI technologies improve and evolve, new business applications emerge.*

*Artificial intelligence is used as a tool to support a human workforce in optimizing workflows and making business operations more efficient. These gains are made in various ways, including using AI to automate repetitive tasks, generate information based on machine learning algorithms, quickly process vast amounts of data sets and extract meaningful insights, and predict future outcomes based on data analysis. AI systems power several types of business automation, including enterprise automation and process automation, helping to reduce human error and free up human workforces for higher-level work.*

*According to McKinsey & Company, the use of artificial intelligence in business operations has doubled since 2017.1 This is largely because AI technology can be customized to meet an organization’s unique needs. 63% of McKinsey’s respondents expect their investment in AI technologies to increase over the next three years.2 To use AI in an effective business strategy, an organization must have a clear understanding of its business functions, how AI works and what aspects of the business can be improved through AI implementation.*

*While the use of AI tools to automate repetitive tasks and increase employee productivity remains popular, businesses are also moving beyond these use cases and using AI to assist in higher-level, strategic initiatives that help drive broader business value.*

*Artificial intelligence: A brief overview*

*Artificial intelligence, “the science and engineering of making intelligent machines, especially intelligent computer programs,”3 uses large amounts of data and human knowledge to power computer systems with the ability to categorize data, make predictions, identify errors, have conversations, and analyze information in a similar way to humans.*

*One of the goals of artificial intelligence is to create computer systems that can mimic the critical thinking skills of humans. These systems rely on business data and use technologies like natural language processing (NLP), machine learning (ML), and deep learning to facilitate business operations. Integrating AI into business functions requires a baseline understanding of the following components:*

*Machine learning algorithms*

*These algorithms are a subset of artificial intelligence and are used to make predictions or classifications based on input data. Through training data sets, these algorithms can learn to identify patterns, discover anomalies, or make projections such as future sales revenue. Machine learning algorithms help mine large datasets for key insights that can offer real-world benefits for improved business decisions. Machine learning algorithms benefit from labeled data, which is data that a human expert categorizes before it is processed.*

*Deep learning*

*Deep learning is a subset of machine learning that allows for the automation of tasks without human intervention. Virtual assistants, chatbots, facial recognition and fraud prevention technology all rely on deep learning. By examining data that is related to user behavior, deep learning models can make predictions about future behavior. Compared to general machine learning, deep learning models can more accurately extract information from unstructured data such as text and images and do not require as much human intervention.*

*Natural language processing (NLP)*

*Natural language processing is a branch of AI that “enables computers and digital devices to recognize, understand, and generate text and speech.”4 Customer support chatbots, digital assistants, and voice-operated technologies such as GPS systems are all powered by NLP. Used with machine learning algorithms and deep learning models, NLP allows systems to extract insights from unstructured data that are text- or voice-driven.*

*Computer vision*

*Computer vision is a subset of AI that allows computer systems to extract information from digital images, videos and other visual inputs.5 Computer vision uses both deep learning and machine learning algorithms to learn and identify specific elements of digital imagery. Computer vision is currently applied in several ways, and applications are expanding as the technology progresses. For example, computer vision can be implemented in production lines to detect minor defects during the manufacturing process.*

*Integrating enterprise-grade AI can help free human workforces from repetitive manual tasks, improve data analysis, business strategy and decision-making, and optimize processes organization-wide. To do so, enterprises must have an infrastructure that properly manages data and supports AI technology. Having a strong data governance framework helps keep data available to all relevant stakeholders and secure from data breaches.*

*It also helps promote the use of advanced data analytics. Part of this framework involves a digital transformation and the integration of hybrid cloud and multicloud environments to help manage large volumes of data. Once these systems are in place, an organization can begin mining data for insights and building training models to instruct AI technologies.*

*AI in business use cases*

*As new technologies enter the market, and existing ones improve, the possible applications of artificial intelligence in business grow more numerous. The benefits of AI vary and require the integration of technologies and human workforces to improve operational efficiency and drive business value.*

*Some examples that demonstrate the use of artificial intelligence in business include:*

*IT operations*

*AIOps—artificial intelligence for IT operations—consists of the practice of using AI, machine learning and natural language processing models to streamline IT operations and service management. AIOps allows IT teams to quickly sift through large amounts of data and reduce the amount of time it takes to detect anomalies, troubleshoot errors, and monitor the performance of IT systems. Artificial intelligence helps IT teams achieve greater observability and provides real-time insights into operations.*

*Marketing and sales*

*Customer data helps marketing teams develop marketing strategies by identify trends and spending patterns. Artificial intelligence tools help process these big data sets to forecast future spending trends and conduct competitor analysis. This helps an organization gain a deeper understanding of its place in the market.  
  
AI tools allow for marketing segmentation, a strategy that uses data to tailor marketing campaigns to specific customers based on their interests. Sales teams can use this same data to make product recommendations based on customer analytics.*

*Customer service*

*AI enables businesses to provide 24/7 customer service and faster response times, which help improve the customer experience. AI-powered chatbots can help customers resolve simple queries without requiring a human agent. This ability allows the human customer service workforce to address more complex issues.*

*McKinsey reported savings of USD 80 million for a South American telecommunications company that used conversational AI to prioritize higher-value clients.6Powerful conversational AI tools such as IBM watsonx™ Assistant help chatbots overcome some of the pain points of earlier models, which were unable to handle many customer questions.*

*Content generation*

*Generative AI (GenAI) is a growing field that helps organizations optimize content creation. Tools such as ChatGPT provide content teams with powerful tools to create original content. These tools can generate images or text based on input prompts, and designers, writers, and content leads can use these generative AI outputs to help with brainstorming, outlining, and other project tasks. Gartner estimates that by 2025 generative AI will be used to create 30% of outbound marketing content, up from 2% in 2022.7 Generative tools such as IBM watsonx™ Code Assistant can help developers by generating code.*

*While AI content generation is still largely unregulated, human employees should monitor the use of AI in generating content to prevent copyright infringement, the publication of misinformation, or other unethical business practices.*

*Cybersecurity*

*Artificial intelligence tools can be used to improve network security, anomaly detection, fraud detection, and help prevent data breaches. The increased use of technology in the workplace creates greater opportunities for security breaches; to thwart threats and protect organizational and customer data, organizations must be proactive in detecting anomalies. For example, deep learning models can be used to examine large sets of network traffic data and identify behavior that might signal an attempted attack on the network.*

*Data breaches can be costly and erode customer trust. The IBM Cost of a Data Breach Report 2023 indicates that the average savings for organizations that “use security AI and automation extensively is USD 1.76 million compared to organizations that don’t.”*

*Supply chain management*

*The application of AI in supply chain management comes in the form of predictive analytics, which helps forecast future pricing of shipping and material costs. Predictive analytics also helps organizations maintain appropriate levels of inventory. This reduces bottlenecks, or the overstocking of products.*

*AI technologies are rapidly evolving, and their use is expanding to meet a wider variety of business needs and strategies. New technologies and the innovation of business leaders will dictate the future of AI—understanding how AI fits into your business model is key to maintaining a competitive edge.*

***What is Reinforcement Learning?***

*Say, we have an agent in an unknown environment and this agent can obtain some rewards by interacting with the environment. The agent ought to take actions so as to maximize cumulative rewards. In reality, the scenario could be a bot playing a game to achieve high scores, or a robot trying to complete physical tasks with physical items; and not just limited to these*

*The goal of Reinforcement Learning (RL) is to learn a good strategy for the agent from experimental trials and relative simple feedback received. With the optimal strategy, the agent is capable to actively adapt to the environment to maximize future rewards.*

***Key Concepts***

*Now Let’s formally define a set of key concepts in RL.*

*The agent is acting in an****environment****. How the environment reacts to certain actions is defined by a****model****which we may or may not know. The agent can stay in one of many****states****(s∈S) of the environment, and choose to take one of many****actions****(a∈A) to switch from one state to another. Which state the agent will arrive in is decided by transition probabilities between states (P). Once an action is taken, the environment delivers a****reward****(r∈R) as feedback.*

*The model defines the reward function and transition probabilities. We may or may not know how the model works and this differentiate two circumstances:*

* ***Know the model****: planning with perfect information; do model-based RL. When we fully know the environment, we can find the optimal solution by Dynamic Programming (DP). Do you still remember “longest increasing subsequence” or “traveling salesmen problem” from your Algorithms 101 class? LOL. This is not the focus of this post though.*
* ***Does not know the model****: learning with incomplete information; do model-free RL or try to learn the model explicitly as part of the algorithm. Most of the following content serves the scenarios when the model is unknown.*

*The agent’s****policy****π(s) provides the guideline on what is the optimal action to take in a certain state with****the goal to maximize the total rewards****. Each state is associated with a****value****function V(s) predicting the expected amount of future rewards we are able to receive in this state by acting the corresponding policy. In other words, the value function quantifies how good a state is. Both policy and value functions are what we try to learn in reinforcement learning.*

*Fig. 2. Summary of approaches in RL based on whether we want to model the value, policy, or the environment. (Image source: reproduced from David Silver's RL course lecture 1.)*

*The interaction between the agent and the environment involves a sequence of actions and observed rewards in time, t=1,2,…,T. During the process, the agent accumulates the knowledge about the environment, learns the optimal policy, and makes decisions on which action to take next so as to efficiently learn the best policy. Let’s label the state, action, and reward at time step t as St, At, and Rt, respectively. Thus the interaction sequence is fully described by one****episode****(also known as “trial” or “trajectory”) and the sequence ends at the terminal state ST:*

*S1,A1,R2,S2,A2,…,ST*

*Terms you will encounter a lot when diving into different categories of RL algorithms:*

* ***Model-based****: Rely on the model of the environment; either the model is known or the algorithm learns it explicitly.*
* ***Model-free****: No dependency on the model during learning.*
* ***On-policy****: Use the deterministic outcomes or samples from the target policy to train the algorithm.*
* ***Off-policy****: Training on a distribution of transitions or episodes produced by a different behavior policy rather than that produced by the target policy.*

***Model: Transition and Reward***

*The model is a descriptor of the environment. With the model, we can learn or infer how the environment would interact with and provide feedback to the agent. The model has two major parts, transition probability function P and reward function R.*

*Let’s say when we are in state s, we decide to take action a to arrive in the next state s’ and obtain reward r. This is known as one****transition****step, represented by a tuple (s, a, s’, r).*

*The transition function P records the probability of transitioning from state s to s’ after taking action a while obtaining reward r. We use P as a symbol of “probability”.*

*P(s′,r|s,a)=P[St+1=s′,Rt+1=r|St=s,At=a]*

*Thus the state-transition function can be defined as a function of P(s′,r|s,a):*

*Pss′a=P(s′|s,a)=P[St+1=s′|St=s,At=a]=∑r∈RP(s′,r|s,a)*

*The reward function R predicts the next reward triggered by one action:*

*R(s,a)=E[Rt+1|St=s,At=a]=∑r∈Rr∑s′∈SP(s′,r|s,a)*

***Policy***

*Policy, as the agent’s behavior function π, tells us which action to take in state s. It is a mapping from state s to action a and can be either deterministic or stochastic:*

* *Deterministic: π(s)=a.*
* *Stochastic: π(a|s)=Pπ[A=a|S=s].*

***Value Function***

*Value function measures the goodness of a state or how rewarding a state or an action is by a prediction of future reward. The future reward, also known as****return****, is a total sum of discounted rewards going forward. Let’s compute the return Gt starting from time t:*

*Gt=Rt+1+γRt+2+⋯=∑k=0∞γkRt+k+1*

*The discounting factor γ∈[0,1] penalize the rewards in the future, because:*

* *The future rewards may have higher uncertainty; i.e. stock market.*
* *The future rewards do not provide immediate benefits; i.e. As human beings, we might prefer to have fun today rather than 5 years later ;).*
* *Discounting provides mathematical convenience; i.e., we don’t need to track future steps forever to compute return.*
* *We don’t need to worry about the infinite loops in the state transition graph.*

*The****state-value****of a state s is the expected return if we are in this state at time t, St=s:*

*Vπ(s)=Eπ[Gt|St=s]*

*Similarly, we define the****action-value****(“Q-value”; Q as “Quality” I believe?) of a state-action pair as:*

*Qπ(s,a)=Eπ[Gt|St=s,At=a]*

*Additionally, since we follow the target policy π, we can make use of the probility distribution over possible actions and the Q-values to recover the state-value:*

*Vπ(s)=∑a∈AQπ(s,a)π(a|s)*

*The difference between action-value and state-value is the action****advantage****function (“A-value”):*

*Aπ(s,a)=Qπ(s,a)−Vπ(s)*

***Markov Decision Processes***

*In more formal terms, almost all the RL problems can be framed as****Markov Decision Processes****(MDPs). All states in MDP has “Markov” property, referring to the fact that the future only depends on the current state, not the history:*

*P[St+1|St]=P[St+1|S1,…,St]*

*Or in other words, the future and the past are****conditionally independent****given the present, as the current state encapsulates all the statistics we need to decide the future.*

*Fig. 3. The agent-environment interaction in a Markov decision process. (Image source: Sec. 3.1 Sutton & Barto (2017).)*

*A Markov deicison process consists of five elements M=⟨S,A,P,R,γ⟩, where the symbols carry the same meanings as key concepts in the previous section, well aligned with RL problem settings:*

* *S - a set of states;*
* *A - a set of actions;*
* *P - transition probability function;*
* *R - reward function;*
* *γ - discounting factor for future rewards. In an unknown environment, we do not have perfect knowledge about P and R.*

*Fig. 4. A fun example of Markov decision process: a typical work day. (Image source: randomant.net/reinforcement-learning-concepts)*

***Bellman Equations***

*Bellman equations refer to a set of equations that decompose the value function into the immediate reward plus the discounted future values.*

*V(s)=E[Gt|St=s]=E[Rt+1+γRt+2+γ2Rt+3+…|St=s]=E[Rt+1+γ(Rt+2+γRt+3+…)|St=s]=E[Rt+1+γGt+1|St=s]=E[Rt+1+γV(St+1)|St=s]*

*Similarly for Q-value,*

*Q(s,a)=E[Rt+1+γV(St+1)∣St=s,At=a]=E[Rt+1+γEa∼πQ(St+1,a)∣St=s,At=a]*

***Bellman Expectation Equations***

*The recursive update process can be further decomposed to be equations built on both state-value and action-value functions. As we go further in future action steps, we extend V and Q alternatively by following the policy π.*

*Fig. 5. Illustration of how Bellman expection equations update state-value and action-value functions.*

*Vπ(s)=∑a∈Aπ(a|s)Qπ(s,a)Qπ(s,a)=R(s,a)+γ∑s′∈SPss′aVπ(s′)Vπ(s)=∑a∈Aπ(a|s)(R(s,a)+γ∑s′∈SPss′aVπ(s′))Qπ(s,a)=R(s,a)+γ∑s′∈SPss′a∑a′∈Aπ(a′|s′)Qπ(s′,a′)*

***Bellman Optimality Equations***

*If we are only interested in the optimal values, rather than computing the expectation following a policy, we could jump right into the maximum returns during the alternative updates without using a policy. RECAP: the optimal values V∗ and Q∗ are the best returns we can obtain, defined here.*

*V∗(s)=maxa∈AQ∗(s,a)Q∗(s,a)=R(s,a)+γ∑s′∈SPss′aV∗(s′)V∗(s)=maxa∈A(R(s,a)+γ∑s′∈SPss′aV∗(s′))Q∗(s,a)=R(s,a)+γ∑s′∈SPss′amaxa′∈AQ∗(s′,a′)*

*Unsurprisingly they look very similar to Bellman expectation equations.*

*If we have complete information of the environment, this turns into a planning problem, solvable by DP. Unfortunately, in most scenarios, we do not know Pss′a or R(s,a), so we cannot solve MDPs by directly applying Bellmen equations, but it lays the theoretical foundation for many RL algorithms.*

***Common Approaches***

*Now it is the time to go through the major approaches and classic algorithms for solving RL problems. In future posts, I plan to dive into each approach further.*

***Dynamic Programming***

*When the model is fully known, following Bellman equations, we can use Dynamic Programming (DP) to iteratively evaluate value functions and improve policy.*

***Temporal-Difference Learning***

*Similar to Monte-Carlo methods, Temporal-Difference (TD) Learning is model-free and learns from episodes of experience. However, TD learning can learn from****incomplete****episodes and hence we don’t need to track the episode up to termination. TD learning is so important that Sutton & Barto (2017) in their RL book describes it as “one idea … central and novel to reinforcement learning”.*

***Known Problems***

***Exploration-Exploitation Dilemma***

*The problem of exploration vs exploitation dilemma has been discussed in my previous post. When the RL problem faces an unknown environment, this issue is especially a key to finding a good solution: without enough exploration, we cannot learn the environment well enough; without enough exploitation, we cannot complete our reward optimization task.*

*Different RL algorithms balance between exploration and exploitation in different ways. In MC methods, Q-learning or many on-policy algorithms, the exploration is commonly implemented by ε-greedy; In ES, the exploration is captured by the policy parameter perturbation. Please keep this into consideration when developing a new RL algorithm.*

***Deadly Triad Issue***

*We do seek the efficiency and flexibility of TD methods that involve bootstrapping. However, when off-policy, nonlinear function approximation, and bootstrapping are combined in one RL algorithm, the training could be unstable and hard to converge. This issue is known as the****deadly triad****(Sutton & Barto, 2017). Many architectures using deep learning models were proposed to resolve the problem, including DQN to stabilize the training with experience replay and occasionally frozen target network.*

***Case Study: AlphaGo Zero***

*The game of Go has been an extremely hard problem in the field of Artificial Intelligence for decades until recent years. AlphaGo and AlphaGo Zero are two programs developed by a team at DeepMind. Both involve deep Convolutional Neural Networks (CNN) and Monte Carlo Tree Search (MCTS) and both have been approved to achieve the level of professional human Go players. Different from AlphaGo that relied on supervised learning from expert human moves, AlphaGo Zero used only reinforcement learning and self-play without human knowledge beyond the basic rules.*

*Fig. 10. The board of Go. Two players play black and white stones alternatively on the vacant intersections of a board with 19 x 19 lines. A group of stones must have at least one open point (an intersection, called a "liberty") to remain on the board and must have at least two or more enclosed liberties (called "eyes") to stay "alive". No stone shall repeat a previous position.*

*With all the knowledge of RL above, let’s take a look at how AlphaGo Zero works. The main component is a deep CNN over the game board configuration (precisely, a ResNet with batch normalization and ReLU). This network outputs two values:*