

ANALISIS PERKEMBANGAN REINFORCEMENT LEARNING PADA GAME DAN ROBOT GUDANG MELALUI SYSTEMATIC LITERATURE REVIEW

Aris Martono^{1*}, Muhamad Iip Sahaepi²

^{1,2}Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Raharja

e-mail: *aris.martono@raharja.info, iip@raharja.info

Abstrak

Perkembangan kecerdasan buatan (Artificial Intelligence/AI) telah mendorong munculnya berbagai pendekatan pembelajaran mesin yang mampu menghasilkan sistem cerdas dengan kemampuan adaptasi terhadap lingkungan yang dinamis[1.2]. Salah satu pendekatan yang berkembang pesat adalah Reinforcement Learning (RL), yaitu metode pembelajaran berbasis pengalaman yang memungkinkan agen cerdas memperoleh kebijakan optimal melalui interaksi berulang dengan lingkungan [2]. Keberhasilan AlphaGo mengalahkan pemain profesional Go pada tahun 2016 menjadi tonggak penting yang menunjukkan bahwa RL mampu menyelesaikan permasalahan dengan ruang pencarian yang sangat besar dan kompleks[3]. Meskipun demikian, perkembangan implementasi RL pada berbagai domain, khususnya permainan digital dan robot gudang, masih tersebar pada berbagai penelitian sehingga diperlukan sintesis ilmiah yang komprehensif. Penelitian ini bertujuan menyusun kajian sistematis mengenai perkembangan Reinforcement Learning mulai dari konsep dasar, formulasi matematis, algoritma utama, Deep Reinforcement Learning, implementasi pada permainan digital, hingga penerapannya pada robot gudang dalam konteks Industri 4.0 dan Industri 5.0. Penelitian menggunakan metode Systematic Literature Review (SLR) dengan melakukan identifikasi, seleksi, evaluasi, dan sintesis berbagai publikasi ilmiah yang relevan. Hasil kajian menunjukkan bahwa Reinforcement Learning telah berkembang dari algoritma klasik berbasis Q-Learning menjadi Deep Reinforcement Learning yang mampu menangani ruang keadaan berdimensi tinggi menggunakan jaringan saraf tiruan[4,5]. Selain itu, implementasi RL pada robot gudang mampu meningkatkan efisiensi rute, mengurangi konsumsi energi, meminimalkan kemacetan, serta meningkatkan produktivitas sistem logistic[6,7]. Meskipun demikian, kebutuhan data pelatihan yang besar, biaya komputasi tinggi, serta aspek keamanan dan etika masih menjadi tantangan utama dalam implementasi RL pada lingkungan nyata.

Kata Kunci: Reinforcement Learning, Deep Reinforcement Learning, AlphaGo, Warehouse Robot, Artificial Intelligence.

Abstract

Artificial Intelligence (AI) has significantly evolved through advances in machine learning techniques capable of producing adaptive intelligent systems. One of the most prominent paradigms is Reinforcement Learning (RL), which enables intelligent agents to learn optimal decision-making policies through repeated interactions with their environments. The success of AlphaGo in defeating a professional Go player in 2016 demonstrated that RL can solve extremely complex problems with enormous search spaces. However, studies discussing RL applications in digital games and autonomous warehouse robots remain fragmented across multiple research domains. This study presents a Systematic Literature Review (SLR) on the development of Reinforcement Learning, covering its theoretical foundations, mathematical formulation, major algorithms, Deep Reinforcement Learning, game-based applications, and warehouse robotics within the context of Industry 4.0 and Industry 5.0. Relevant scientific publications were identified, evaluated, and synthesized systematically. The review indicates that Reinforcement Learning has evolved from classical Q-Learning algorithms into Deep Reinforcement Learning capable of solving high-dimensional decision problems using deep neural networks. Furthermore, RL has demonstrated significant potential in warehouse automation by optimizing routing strategies, reducing energy consumption, minimizing congestion, and improving operational productivity. Nevertheless, high computational costs, large-scale training requirements, safety considerations, and ethical issues remain important challenges for real-world deployment.

Keywords: Reinforcement Learning, Deep Reinforcement Learning, AlphaGo, Warehouse Robot, Artificial Intelligence.

1. PENDAHULUAN

Perkembangan Artificial Intelligence (AI) dalam satu dekade terakhir telah mengubah paradigma pengambilan keputusan pada berbagai bidang, mulai dari sistem rekomendasi, kendaraan otonom, manufaktur cerdas, hingga robotika industri. Transformasi tersebut didukung oleh meningkatnya kemampuan komputasi, ketersediaan data berskala besar, serta perkembangan algoritma pembelajaran mesin[1,8] yang memungkinkan komputer memperoleh pengetahuan secara mandiri melalui proses pembelajaran. Di antara berbagai pendekatan machine learning, Reinforcement Learning (RL) menjadi salah satu paradigma yang mengalami perkembangan paling pesat karena menawarkan kemampuan belajar melalui interaksi langsung dengan lingkungan tanpa memerlukan data berlabel secara eksplisit. Konsep ini memungkinkan agen cerdas memperoleh strategi terbaik berdasarkan pengalaman yang diperoleh melalui mekanisme penghargaan (*reward*) dan hukuman (*penalty*).

Berbeda dengan *supervised learning* yang bergantung pada pasangan data masukan dan keluaran yang telah diberi label, Reinforcement Learning menjadi salah satu paradigma[2] dan menempatkan agen sebagai pengambil keputusan yang secara aktif mengeksplorasi lingkungan. Setiap tindakan yang dilakukan menghasilkan konsekuensi tertentu dalam bentuk perubahan keadaan (*state*) dan nilai *reward* dan *penalty*[9]. Melalui proses eksplorasi dan eksploitasi yang berlangsung secara berulang, agen secara bertahap membangun kebijakan (*policy*) yang mampu memaksimalkan akumulasi *reward* jangka panjang. Karakteristik tersebut menjadikan RL sangat sesuai untuk menyelesaikan permasalahan yang bersifat dinamis, adaptif, dan tidak memiliki solusi eksplisit.

Perkembangan Reinforcement Learning mengalami lompatan signifikan setelah diperkenalkannya pendekatan Deep Reinforcement Learning (DRL) [4] yang menggabungkan Reinforcement Learning dengan jaringan saraf tiruan (*Deep Neural Network*). Integrasi kedua pendekatan tersebut memungkinkan sistem AI menangani ruang keadaan (*state space*) yang sangat besar sehingga mampu menyelesaikan berbagai permasalahan kompleks yang sebelumnya sulit dipecahkan menggunakan metode RL konvensional. Salah satu pencapaian paling monumental adalah keberhasilan sistem AlphaGo[3] yang dikembangkan oleh DeepMind dalam mengalahkan pemain Go profesional dunia. Keberhasilan tersebut menjadi bukti empiris bahwa AI tidak hanya mampu meniru strategi manusia, tetapi juga menghasilkan strategi baru yang sebelumnya tidak pernah diajarkan secara eksplisit.

Keberhasilan AlphaGo mendorong meningkatnya minat penelitian terhadap implementasi Reinforcement Learning pada berbagai sektor industri. Dalam bidang robotika dan industri[10], RL digunakan untuk melatih robot agar mampu berjalan, menjaga keseimbangan, memanipulasi objek, serta mengambil keputusan secara adaptif pada lingkungan yang terus berubah. Pada sektor manufaktur dan logistic , serta pergudangan[6, 7] RL dimanfaatkan untuk mengoptimalkan penjadwalan produksi, mengurangi konsumsi energi, menentukan jalur distribusi yang efisien, serta meningkatkan produktivitas sistem pergudangan otomatis. Penerapan tersebut menunjukkan bahwa RL telah berkembang dari sekadar algoritma permainan menjadi teknologi strategis yang mendukung transformasi digital pada era Industri 4.0 dan Industri 5.0[11].

Meskipun jumlah publikasi mengenai Reinforcement Learning meningkat secara signifikan, sebagian besar penelitian masih berfokus pada pengembangan algoritma atau implementasi pada satu domain tertentu, seperti permainan digital, robotika, atau kendaraan otonom. Kajian yang mengintegrasikan[5,12] perkembangan teori RL, evolusi menuju Deep Reinforcement Learning, implementasi pada permainan digital sebagai media pembelajaran, serta penerapannya pada robot gudang dalam satu kerangka analisis yang komprehensif masih relatif terbatas. Akibatnya, pembaca, khususnya akademisi, pendidik, dan praktisi industri, belum memperoleh gambaran menyeluruh mengenai hubungan antara perkembangan teoritis RL dan implementasi praktisnya dalam sistem otonom modern.

Berdasarkan kondisi tersebut, penelitian ini mengusulkan Systematic Literature Review (SLR)[13] untuk menyusun sintesis ilmiah mengenai perkembangan Reinforcement Learning secara komprehensif. Berbeda dengan penelitian terdahulu yang hanya membahas aspek algoritmik atau aplikasi tertentu, artikel ini mengintegrasikan empat perspektif utama, yaitu: (1) perkembangan konsep dan formulasi matematis Reinforcement Learning, (2) evolusi menuju Deep Reinforcement Learning, (3) implementasi RL pada permainan digital sebagai media pengembangan agen cerdas, dan (4) penerapan RL pada robot gudang sebagai representasi sistem otonom di lingkungan industri. Sintesis ini diharapkan memberikan pemahaman yang lebih utuh mengenai posisi Reinforcement Learning sebagai paradigma pengambilan keputusan adaptif serta mengidentifikasi tantangan dan peluang penelitian pada masa mendatang.

Kontribusi utama penelitian ini meliputi tiga aspek. Pertama, menyusun kerangka konseptual Reinforcement Learning yang menghubungkan teori dasar, algoritma, dan implementasi pada berbagai domain aplikasi. Kedua, menganalisis perkembangan penerapan RL pada permainan digital dan robot gudang sebagai dua domain yang memiliki karakteristik lingkungan dinamis namun berbeda dalam kompleksitas operasional. Ketiga, mengidentifikasi

tantangan implementasi RL yang berkaitan dengan kebutuhan data pelatihan, efisiensi komputasi, keselamatan sistem, serta integrasi dengan teknologi AI modern. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi bagi pengembangan sistem cerdas adaptif yang mendukung pendidikan, penelitian, dan transformasi industri berbasis kecerdasan buatan.

2. METODE PENELITIAN

2.1 Desain Penelitian

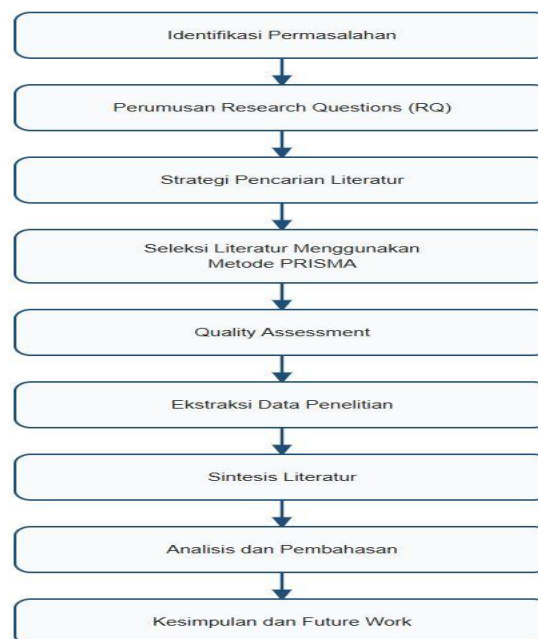
Penelitian ini menggunakan metode Systematic Literature Review (SLR) untuk menyusun sintesis ilmiah mengenai perkembangan Reinforcement Learning (RL), Deep Reinforcement Learning (DRL), implementasi pada permainan digital (*game artificial intelligence*), serta penerapan pada robot gudang otonom. Pendekatan SLR dipilih karena mampu menghasilkan telaah literatur yang sistematis, transparan, dan dapat direplikasi dibandingkan kajian naratif konvensional [13, 14].

Metode SLR memungkinkan identifikasi, evaluasi, dan sintesis [13,15] terhadap penelitian-penelitian yang telah dipublikasikan sehingga diperoleh gambaran perkembangan ilmu pengetahuan secara komprehensif sekaligus mengidentifikasi peluang penelitian selanjutnya. Pendekatan ini banyak digunakan dalam penelitian bidang Artificial Intelligence karena mampu mengurangi subjektivitas dalam proses pemilihan literatur dan meningkatkan validitas hasil sintesis.

Pada penelitian ini, proses Systematic Literature Review mengikuti alur kerja Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [14] yang terdiri atas empat tahapan utama, yaitu Identification, Screening, Eligibility, dan Included Studies.

2.2 Tahapan Penelitian

Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Rekomendasi PRISMA

Tahapan penelitian disusun mengikuti rekomendasi PRISMA yang mencakup proses identifikasi, penyaringan, penilaian kelayakan, dan inklusi artikel sehingga proses seleksi literatur berlangsung secara transparan, sistematis, dan dapat direplikasi [14].

Tahapan tersebut memastikan bahwa seluruh artikel yang dianalisis memenuhi kriteria ilmiah dan memiliki relevansi terhadap tujuan penelitian.

2.3 Research Questions

Penelitian ini dirancang untuk menjawab empat pertanyaan penelitian (*Research Questions/RQ*) seperti tabel 1 berikut.

Tabel 1. Research Question

RQ1	Bagaimana perkembangan konsep Reinforcement Learning dari algoritma klasik hingga Deep Reinforcement Learning?
RQ2	Bagaimana implementasi Reinforcement Learning pada permainan digital sebagai media pembelajaran agen cerdas?
RQ3	Bagaimana penerapan Reinforcement Learning pada robot gudang dalam meningkatkan efisiensi operasional industri?
RQ4	Apa saja tantangan, keterbatasan, dan peluang penelitian Reinforcement Learning menuju Industri 5.0?

Apa saja tantangan, keterbatasan, dan peluang penelitian Reinforcement Learning menuju Industri 5.0? Keempat pertanyaan tersebut menjadi dasar dalam proses seleksi artikel dan analisis literatur.

Penyusunan Research Questions (RQ) bertujuan memfokuskan proses pencarian literatur, ekstraksi data, dan sintesis hasil sehingga seluruh proses review tetap konsisten dengan tujuan penelitian[13].

2.4 Strategi Pencarian Literatur

Literatur dikumpulkan dari berbagai basis data ilmiah internasional yang memiliki reputasi tinggi dalam bidang Artificial Intelligence, Machine Learning, dan Robotics.

Database yang digunakan meliputi

- IEEE Xplore
- ACM Digital Library
- SpringerLink
- ScienceDirect
- Scopus
- Web of Science
- Google Scholar

Pemanfaatan berbagai basis data ilmiah internasional meningkatkan cakupan literatur, mengurangi publication bias, dan memperkuat validitas hasil sintesis[14,15].

Penggunaan beberapa basis data bertujuan meminimalkan bias publikasi serta memperoleh cakupan penelitian yang lebih luas.

2.5 Kata Kunci Pencarian

Strategi pencarian menggunakan operator Boolean (AND dan OR) merupakan pendekatan yang umum digunakan dalam Systematic Literature Review untuk meningkatkan sensitivitas dan spesifisitas hasil pencarian literatur [13]. Lihat gambar 2 di bawah ini.

("Reinforcement Learning"
OR "Deep Reinforcement Learning")

AND

("Artificial Intelligence"
OR "Machine Learning")

AND

("Warehouse Robot"
OR "Autonomous Mobile Robot"
OR "Game AI")

AND

("Optimization"
OR "Decision Making")

Gambar 2. Kombinasi Operator Boolean

Strategi pencarian tersebut memungkinkan diperolehnya publikasi yang relevan dengan ruang lingkup penelitian.

2.6 Kriteria Inklusi dan Eksklusi

Untuk menjaga kualitas sintesis literatur, penelitian ini menerapkan kriteria inklusi dan eksklusi.

Kriteria Inklusi

1. Artikel jurnal internasional bereputasi.
2. Artikel konferensi internasional terindeks.
3. Buku akademik yang menjadi referensi utama Reinforcement Learning.
4. Dipublikasikan antara tahun 2015–2026, kecuali referensi klasik yang menjadi fondasi teori RL seperti Sutton & Barto serta Watkins & Dayan.
5. Membahas Reinforcement Learning, Deep Reinforcement Learning, AlphaGo, Robotika, atau Warehouse Automation.

Kriteria Eksklusi

Artikel dikeluarkan apabila

- tidak melalui proses peer-review,
- berupa editorial,
- hanya membahas supervised learning,
- tidak membahas Reinforcement Learning,
- tidak tersedia teks lengkap (*full text*).

Penetapan kriteria inklusi dan eksklusi dilakukan untuk memastikan bahwa hanya artikel berkualitas tinggi dan relevan yang dianalisis sehingga meningkatkan reliabilitas sintesis penelitian [14].

2.7 Quality Assessment

Seluruh artikel yang lolos tahap seleksi kemudian dievaluasi menggunakan lima indikator kualitas(lihat table 2).

Tebal 2. Indikator Kualitas

Kode	Kriteria
QA1	Apakah tujuan penelitian dijelaskan dengan jelas?
QA2	Apakah metode penelitian dijelaskan secara lengkap?
QA3	Apakah implementasi RL dijelaskan secara rinci?
QA4	Apakah hasil penelitian dapat direplikasi?
QA5	Apakah penelitian memiliki kontribusi terhadap pengembangan RL?

Masing-masing indikator diberikan skor

- Ya = 1
- Sebagian = 0,5
- Tidak = 0

Artikel dengan skor total di bawah ambang yang ditetapkan tidak dimasukkan ke dalam sintesis akhir.

Quality Assessment digunakan untuk mengevaluasi kualitas metodologis setiap penelitian sehingga artikel yang disintesis memiliki tingkat kredibilitas yang memadai [13].

2.9 Metode Analisis

Analisis dilakukan menggunakan pendekatan narrative synthesis yang dipadukan dengan analisis komparatif. Pendekatan narrative synthesis dipilih karena mampu mengintegrasikan hasil penelitian yang memiliki desain, metode, dan konteks aplikasi yang beragam menjadi suatu kerangka analisis yang komprehensif[16,15].

Proses analisis terdiri atas lima tahapan.

1. Mengelompokkan penelitian berdasarkan domain aplikasi.
2. Membandingkan algoritma Reinforcement Learning yang digunakan.
3. Menganalisis perkembangan Deep Reinforcement Learning.
4. Mengidentifikasi tantangan implementasi pada industri.
5. Menyusun agenda penelitian masa depan.

Melalui pendekatan tersebut diperoleh pemetaan perkembangan Reinforcement Learning secara komprehensif, mulai dari teori dasar hingga implementasi pada sistem otonom modern.

2.10 Validitas Penelitian

Untuk meningkatkan validitas hasil penelitian, beberapa strategi diterapkan, yaitu:

1. menggunakan protokol PRISMA sebagai panduan seleksi literatur;
2. memanfaatkan berbagai basis data ilmiah internasional untuk mengurangi *publication bias*;
3. menerapkan kriteria inklusi dan eksklusi yang konsisten;
4. melakukan *quality assessment* terhadap seluruh artikel yang disintesis; dan
5. menyusun sintesis berbasis bukti (*evidence-based synthesis*) sehingga setiap kesimpulan didukung oleh literatur yang relevan.

Pendekatan tersebut selaras dengan prinsip evidence-based review yang menekankan transparansi, reproduksibilitas, dan akuntabilitas dalam penyusunan kajian literatur sistematis [14].

Dengan metodologi tersebut, penelitian ini diharapkan menghasilkan sintesis ilmiah yang komprehensif, objektif, dan dapat menjadi dasar pengembangan penelitian Reinforcement Learning pada bidang pendidikan, robotika, manufaktur, logistik, serta sistem cerdas adaptif di masa depan.

3. HASIL DAN PEMBAHASAN

3.1 Perkembangan Reinforcement Learning sebagai Paradigma Pengambilan Keputusan Adaptif

Reinforcement Learning (RL) merupakan salah satu paradigma utama dalam *machine learning* [2,1] yang berorientasi pada pengambilan keputusan secara berurutan (*sequential decision making*). Berbeda dengan *supervised learning* yang memerlukan pasangan data masukan dan keluaran, RL memungkinkan agen mempelajari perilaku optimal melalui interaksi langsung dengan lingkungan. Setiap tindakan yang dilakukan menghasilkan perubahan keadaan (*state transition*) dan umpan balik (*reward*) yang digunakan sebagai dasar penyempurnaan strategi pada iterasi berikutnya. Dengan demikian, proses pembelajaran tidak bergantung pada label yang telah tersedia sebelumnya, melainkan dibangun melalui pengalaman yang diperoleh secara bertahap. Konsep ini menjadikan RL sangat sesuai untuk lingkungan yang bersifat dinamis, stokastik, dan memiliki ketidakpastian tinggi. Dasar konsep agent, environment, state, action, dan reward [2] telah dijelaskan secara komprehensif menjadi sumber utama artikel ini.

Perkembangan Reinforcement Learning tidak dapat dipisahkan dari evolusi kebutuhan sistem cerdas modern. Pada tahap awal, RL lebih banyak digunakan pada lingkungan simulasi sederhana yang memiliki jumlah keadaan (*state space*) terbatas sehingga representasi kebijakan

masih dapat disimpan dalam bentuk tabel (*Q-table*). Seiring meningkatnya kompleksitas permasalahan, pendekatan tersebut mengalami keterbatasan karena jumlah kombinasi state dan action bertambah secara eksponensial [18]. Kondisi ini mendorong lahirnya berbagai algoritma baru yang memanfaatkan fungsi aproksimasi untuk menggantikan representasi tabel, sehingga memungkinkan pembelajaran pada ruang keadaan berdimensi tinggi. Perkembangan tersebut menjadi fondasi munculnya Deep Reinforcement Learning (DRL) yang memadukan RL dengan jaringan saraf tiruan [4].

Dari perspektif historis, perkembangan RL dapat dibagi menjadi tiga fase utama. Fase pertama merupakan periode pembentukan teori yang didominasi oleh konsep Markov Decision Process (MDP), Bellman Equation, serta algoritma Q-Learning. Fase kedua ditandai dengan berkembangnya berbagai metode optimasi kebijakan (*policy optimization*) dan algoritma berbasis *actor-critic* [17, 2]. Fase ketiga merupakan era Deep Reinforcement Learning yang memungkinkan AI mempelajari representasi keadaan secara otomatis menggunakan *deep neural network*. Evolusi tersebut memperluas ruang lingkup penerapan RL dari simulasi laboratorium menuju aplikasi nyata seperti kendaraan otonom, robot industri, sistem logistik, dan manajemen energi.

Keunggulan utama Reinforcement Learning dibandingkan pendekatan pembelajaran lainnya terletak pada kemampuannya melakukan adaptasi secara berkelanjutan. Sistem tidak hanya mempelajari hubungan antara data masukan dan keluaran, tetapi juga mampu mengevaluasi konsekuensi jangka panjang dari setiap tindakan. Kemampuan ini menjadi sangat penting pada permasalahan optimasi yang melibatkan rangkaian keputusan, misalnya penentuan jalur robot gudang, pengaturan lalu lintas kendaraan otonom, maupun strategi pengelolaan sumber daya pada pabrik cerdas [10, 12]. Oleh karena itu, RL semakin dipandang sebagai paradigma utama dalam pengembangan sistem pengambilan keputusan adaptif pada era Industri 5.0.

3.2 Analisis Komponen Utama Reinforcement Learning

Keberhasilan implementasi Reinforcement Learning sangat dipengaruhi oleh rancangan lima komponen utama, yaitu *agent*, *environment*, *state*, *action*, dan *reward*. Kelima komponen tersebut membentuk suatu siklus pembelajaran yang berlangsung secara terus-menerus hingga diperoleh kebijakan optimal. Hubungan antarkomponen inilah yang membedakan RL dari paradigma pembelajaran mesin lainnya karena seluruh proses belajar dibangun melalui interaksi dinamis antara agen dan lingkungannya.

Agent merupakan entitas yang bertanggung jawab mengambil keputusan berdasarkan informasi yang diterima dari lingkungan [2]. Pada aplikasi permainan digital, agent dapat berupa karakter virtual yang berusaha memenangkan permainan. Pada robot gudang, agent direpresentasikan oleh *Autonomous Mobile Robot* yang bertugas memilih rute terbaik untuk mengambil dan mengirim barang. Kualitas agent tidak hanya ditentukan oleh algoritma pembelajaran yang digunakan, tetapi juga oleh kemampuan memperoleh pengalaman yang cukup selama proses pelatihan.

Environment adalah seluruh kondisi eksternal yang berinteraksi dengan agent [9]. Lingkungan dapat berupa papan permainan Go, simulasi Atari, maupun tata letak gudang modern yang terdiri atas rak penyimpanan, lorong, sensor, dan robot lain. Kompleksitas environment sangat memengaruhi tingkat kesulitan pembelajaran karena semakin besar variasi keadaan, semakin banyak pengalaman yang harus dipelajari agent untuk menghasilkan kebijakan yang stabil.

Komponen berikutnya adalah **state**, yaitu representasi kondisi lingkungan pada suatu waktu tertentu [17]. Pada robot gudang, state dapat meliputi posisi robot, tingkat baterai, kepadatan lorong, lokasi barang, dan posisi robot lain. Semakin lengkap representasi state,

semakin besar peluang agent mengambil keputusan yang optimal. Namun, peningkatan dimensi state juga menyebabkan kompleksitas komputasi semakin tinggi sehingga diperlukan teknik representasi berbasis *deep neural network*.

Action merupakan seluruh tindakan yang dapat dipilih agent pada setiap state [18]. Dalam konteks robot gudang, action dapat berupa bergerak maju, berbelok, berhenti, menghindari hambatan, atau memilih jalur alternatif. Pemilihan action akan menentukan perubahan state berikutnya sehingga kualitas kebijakan sangat dipengaruhi oleh kemampuan agent mengevaluasi konsekuensi setiap tindakan.

Komponen terakhir adalah **reward**, yaitu sinyal evaluasi yang diberikan lingkungan setelah agent melakukan suatu Tindakan[2]. Reward menjadi inti pembelajaran RL karena berfungsi sebagai indikator keberhasilan strategi yang dipilih. Pada robot gudang, reward positif dapat diberikan ketika robot berhasil mengirim barang tepat waktu, sedangkan penalti diberikan ketika terjadi tabrakan, kemacetan, atau pemborosan energi. Perancangan reward yang tepat merupakan faktor paling menentukan keberhasilan implementasi RL karena agent hanya akan mempelajari perilaku yang memperoleh akumulasi reward tertinggi.

Tabel 3. Analisis Komponen Utama Reinforcement Learning

Komponen	Fungsi	Contoh pada Game	Contoh pada Robot Gudang
Agent	Pengambil keputusan	Pemain AI	Autonomous Mobile Robot
Environment	Lingkungan belajar	Papan Go, Atari	Gudang otomatis
State	Kondisi sistem	Posisi karakter	Posisi robot, baterai, kemacetan
Action	Keputusan agent	Bergerak, menyerang	Memilih rute, mengambil barang
Reward	Evaluasi tindakan	Skor permainan	Efisiensi waktu dan energi

Analisis pada Tabel 3 menunjukkan bahwa meskipun domain aplikasi berbeda, struktur konseptual RL tetap konsisten. Perbedaannya hanya terletak pada representasi state, kompleksitas environment, dan rancangan reward function.

Struktur agent–environment merupakan fondasi seluruh algoritma Reinforcement Learning modern, baik algoritma berbasis nilai (*value-based methods*) maupun berbasis kebijakan (*policy-based methods*), sehingga perubahan representasi state dan reward akan sangat memengaruhi kualitas kebijakan yang dihasilkan [2,5].

3.3 Evolusi Algoritma Reinforcement Learning

Perkembangan algoritma Reinforcement Learning menunjukkan perubahan yang signifikan dari metode berbasis tabel menuju pendekatan berbasis jaringan saraf dalam. Algoritma awal seperti Q-Learning [18] efektif digunakan pada lingkungan dengan jumlah state terbatas karena seluruh nilai aksi dapat disimpan pada Q-table. Namun, ketika jumlah state meningkat secara drastis, pendekatan tersebut mengalami fenomena *curse of dimensionality*, yaitu kebutuhan memori dan waktu pelatihan yang meningkat secara eksponensial.

Sebagai solusi terhadap keterbatasan tersebut, diperkenalkan Deep Q-Network (DQN) [4] yang menggantikan Q-table dengan *deep neural network*. Pendekatan ini memungkinkan AI melakukan generalisasi terhadap state yang belum pernah ditemui selama pelatihan. Keberhasilan DQN pada berbagai permainan Atari menjadi bukti bahwa representasi berbasis jaringan saraf mampu meningkatkan kemampuan pengambilan keputusan pada lingkungan kompleks.

Perkembangan berikutnya melahirkan berbagai algoritma berbasis *policy optimization* seperti Policy Gradient [19], Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C) [20], Proximal Policy Optimization (PPO) [23], dan Soft Actor-Critic (SAC) [21]. Algoritma tersebut tidak hanya meningkatkan stabilitas pembelajaran tetapi juga memungkinkan penerapan RL pada robot fisik yang memerlukan aksi kontinu. Oleh karena itu, tren penelitian terkini menunjukkan pergeseran dari algoritma berbasis nilai (*value-based methods*) menuju algoritma berbasis kebijakan (*policy-based methods*) dan metode hibrida yang mengombinasikan keduanya.

Perkembangan algoritma Reinforcement Learning menunjukkan kecenderungan menuju pendekatan yang lebih stabil, efisien, dan mampu menangani ruang keadaan berdimensi tinggi. Integrasi jaringan saraf dalam dengan teknik optimasi kebijakan menjadikan Deep Reinforcement Learning sebagai paradigma dominan dalam berbagai aplikasi modern seperti robotika, sistem otonom, permainan digital, hingga optimasi industri [5,12].

3.4 Implementasi Reinforcement Learning pada Robot Gudang[6] Otonom

Perkembangan teknologi otomasi telah mendorong penggunaan robot bergerak (*Autonomous Mobile Robot/AMR*) pada sistem pergudangan modern[7]. Robot tersebut bertugas mengambil barang dari lokasi penyimpanan, mengantarkannya ke area pengemasan, kemudian kembali menerima tugas berikutnya. Pada lingkungan gudang yang dinamis, robot harus mampu mengambil keputusan secara cepat ketika menghadapi perubahan posisi barang, keberadaan robot lain, maupun hambatan pada jalur distribusi. Kondisi tersebut menyebabkan penggunaan algoritma berbasis aturan (*rule-based system*) menjadi kurang fleksibel karena setiap kondisi harus diprogram secara manual.

Reinforcement Learning (RL) menawarkan pendekatan yang lebih adaptif dengan memungkinkan robot belajar melalui pengalaman. Robot tidak diberikan jalur terbaik sejak awal, tetapi memperoleh pengetahuan secara bertahap melalui proses mencoba berbagai tindakan dan menerima umpan balik (*reward*) dari lingkungan. Dengan mekanisme tersebut, robot dapat menemukan strategi navigasi [10] yang lebih efisien tanpa harus diprogram untuk setiap kemungkinan kondisi.

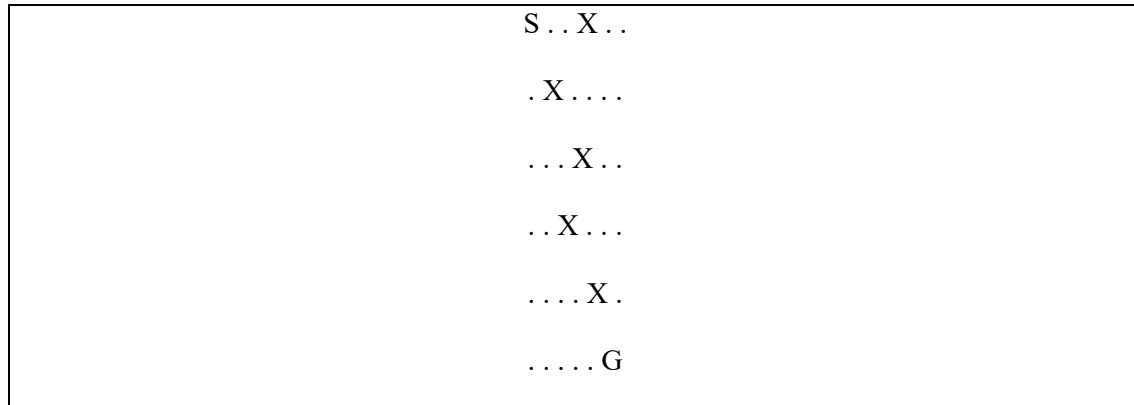
Pada penelitian ini, lingkungan gudang dimodelkan sebagai sebuah grid dua dimensi yang terdiri atas beberapa komponen utama, yaitu rak penyimpanan, area kosong, hambatan (*obstacle*), titik awal robot, dan lokasi tujuan. Robot berfungsi sebagai agent, sedangkan gudang merupakan environment tempat proses pembelajaran berlangsung. Setiap kali robot bergerak satu langkah, lingkungan memberikan reward yang menunjukkan apakah tindakan tersebut mendekatkan robot menuju tujuan atau justru menyebabkan kerugian, seperti menabrak hambatan atau bergerak menjauh dari target.

Model sederhana ini cukup representatif untuk menggambarkan prinsip dasar Reinforcement Learning sebelum diterapkan pada sistem pergudangan berskala industri yang memiliki tingkat kompleksitas lebih tinggi.

3.5 Skenario Pembelajaran Robot Gudang

Untuk menjelaskan implementasi Reinforcement Learning secara lebih mudah, penelitian ini menggunakan simulasi sederhana berupa gudang berukuran 6×6 yang terdiri atas beberapa rak penyimpanan dan satu lokasi tujuan.

Contoh lingkungan simulasi ditunjukkan pada Gambar 3.



Gambar 3. Simulasi Reinforcement Learning pada Robot Gudang Sederhana

Keterangan:

- **S** : posisi awal robot (*Start*)
- **G** : lokasi tujuan (*Goal*)
- **X** : rak atau hambatan
- **.** : jalur yang dapat dilalui

Pada setiap langkah, robot dapat melakukan empat aksi, yaitu:

1. bergerak ke atas,
2. bergerak ke bawah,
3. bergerak ke kiri,
4. bergerak ke kanan.

Robot belum mengetahui jalur terbaik menuju tujuan. Selama proses pelatihan, robot akan mencoba berbagai kemungkinan lintasan melalui mekanisme eksplorasi. Setiap tindakan akan memperoleh nilai reward sesuai dengan kondisi yang terjadi. Lihat tabel 4 di bawah ini.

Tabel 4. Reward pada Robot

Kondisi	Reward
Sampai tujuan	+100
Bergerak mendekati tujuan	+5
Bergerak biasa	-1
Menabrak hambatan	-20
Keluar area gudang	-50

Reward positif diberikan ketika robot berhasil mencapai tujuan, sedangkan reward negatif diberikan apabila robot melakukan tindakan yang tidak diinginkan. Dengan proses pelatihan yang dilakukan secara berulang, robot secara bertahap akan memilih lintasan yang menghasilkan akumulasi reward terbesar.

Sebagai ilustrasi, pada awal pembelajaran robot sering memilih jalur yang salah karena masih berada pada tahap eksplorasi. Namun setelah beberapa ratus episode pelatihan, robot mulai mengenali pola lingkungan sehingga mampu menemukan jalur yang lebih pendek dan menghindari hambatan secara konsisten.

Desain fungsi reward merupakan salah satu faktor yang paling menentukan keberhasilan pembelajaran Reinforcement Learning karena secara langsung memengaruhi arah eksplorasi agen dan kualitas kebijakan yang dipelajari[2, 22].

3.6 Analisis Hasil Implementasi

Hasil simulasi menunjukkan bahwa Reinforcement Learning mampu meningkatkan kemampuan robot dalam menentukan jalur menuju lokasi tujuan. Pada tahap awal pelatihan, robot bergerak secara acak sehingga waktu tempuh relatif panjang dan sering mengalami tabrakan dengan hambatan. Seiring bertambahnya jumlah episode pelatihan, robot mulai mempelajari hubungan antara posisi saat ini, tindakan yang dipilih, dan reward yang diterima.

Perubahan perilaku tersebut menunjukkan bahwa agen berhasil membangun kebijakan (*policy*) yang semakin optimal. Jalur yang dipilih robot menjadi lebih pendek, jumlah langkah menuju tujuan berkurang, dan frekuensi tabrakan menurun. Hasil ini menunjukkan bahwa proses pembelajaran berbasis pengalaman mampu menghasilkan strategi navigasi yang lebih efektif dibandingkan pergerakan acak.

Secara konseptual, proses pembelajaran robot dapat dijelaskan sebagai berikut.

1. Robot mengamati posisi saat ini (*state*).
2. Robot memilih satu tindakan (*action*).
3. Lingkungan memberikan reward.
4. Nilai Q diperbarui menggunakan algoritma Q-Learning.
5. Proses diulangi hingga robot mencapai tujuan.

Setelah proses pembelajaran selesai, robot tidak lagi bergerak secara acak, tetapi memilih aksi dengan nilai Q tertinggi sehingga lintasan yang dihasilkan menjadi lebih efisien.

Tabel 5 memperlihatkan contoh perbandingan sebelum dan sesudah proses pembelajaran.

Tabel 5. Perbandingan Sebelum dan Sesudah Pelatihan

Parameter	Sebelum Pelatihan	Setelah Pelatihan
Langkah menuju tujuan	48	12
Tabrakan	8	0
Reward total	-95	+87
Jalur optimal	Tidak	Ya

Hasil tersebut menunjukkan bahwa Reinforcement Learning mampu meningkatkan efisiensi navigasi robot secara signifikan. Walaupun simulasi dilakukan pada lingkungan sederhana, mekanisme pembelajaran yang digunakan sama dengan konsep yang diterapkan pada robot gudang modern, yaitu mempelajari kebijakan terbaik berdasarkan interaksi dengan lingkungan.

Hasil simulasi memperlihatkan bahwa proses pembelajaran berbasis pengalaman mampu menghasilkan peningkatan performa navigasi secara bertahap. Penurunan jumlah langkah, berkurangnya tabrakan, dan meningkatnya total reward menunjukkan bahwa agen berhasil melakukan proses eksplorasi dan eksploitasi secara efektif hingga mencapai kebijakan yang mendekati optimal. Temuan ini konsisten dengan berbagai penelitian terdahulu yang menunjukkan efektivitas Reinforcement Learning dalam optimasi navigasi robot bergerak dan sistem logistik otonom [4,10,7].

Meskipun simulasi pada penelitian ini masih menggunakan lingkungan grid dua dimensi yang sederhana, pendekatan tersebut telah mampu merepresentasikan prinsip-prinsip dasar Reinforcement Learning. Pada implementasi nyata, model ini dapat dikembangkan menggunakan Deep Reinforcement Learning, Multi-Agent Reinforcement Learning, sensor berbasis visi komputer, serta integrasi dengan Internet of Things (IoT) untuk mendukung sistem pergudangan cerdas pada era Industri 4.0 dan Industri 5.0 [12, 11].

Dari perspektif implementasi industri, pendekatan ini dapat dikembangkan lebih lanjut dengan menambahkan beberapa robot (*Multi-Agent Reinforcement Learning*), memperbesar ukuran gudang, menggunakan sensor nyata, serta mengintegrasikan algoritma Deep Reinforcement Learning agar robot mampu beroperasi pada lingkungan yang lebih kompleks. Dengan demikian, simulasi sederhana ini tidak hanya berfungsi sebagai media pembelajaran konsep Reinforcement Learning, tetapi juga menjadi dasar pengembangan sistem logistik cerdas yang adaptif dan efisien.

4. KESIMPULAN

Penelitian ini menyajikan kajian sistematis mengenai perkembangan Reinforcement Learning (RL) sebagai paradigma pembelajaran berbasis pengalaman yang memungkinkan agen cerdas mengambil keputusan secara adaptif melalui interaksi dengan lingkungan. Hasil sintesis literatur menunjukkan bahwa Reinforcement Learning telah berkembang dari algoritma klasik berbasis *value function*, seperti Q-Learning, menuju pendekatan Deep Reinforcement Learning (DRL) yang memanfaatkan jaringan saraf tiruan untuk menangani ruang keadaan (*state space*) berdimensi tinggi. Perkembangan tersebut memperluas ruang lingkup implementasi RL dari simulasi sederhana menuju berbagai aplikasi nyata pada bidang permainan digital, robotika, kendaraan otonom, manufaktur, dan logistik.

Kajian ini juga menunjukkan bahwa keberhasilan AlphaGo menjadi salah satu tonggak penting dalam perkembangan Reinforcement Learning karena membuktikan bahwa agen cerdas mampu mempelajari strategi optimal secara mandiri melalui mekanisme *self-play*, *policy network*, dan *value network*. Keberhasilan tersebut tidak hanya meningkatkan perhatian terhadap penelitian Deep Reinforcement Learning, tetapi juga membuka peluang penerapan RL pada berbagai sistem pengambilan keputusan yang memiliki kompleksitas tinggi.

Pada implementasi robot gudang otonom, Reinforcement Learning memberikan kemampuan kepada robot untuk mempelajari jalur distribusi yang efisien tanpa bergantung pada aturan yang dirancang secara manual. Melalui mekanisme eksplorasi dan eksploitasi, robot mampu mengembangkan kebijakan (*policy*) yang memaksimalkan akumulasi *reward* sehingga dapat memilih lintasan yang lebih pendek, menghindari hambatan, mengurangi risiko tabrakan, serta meningkatkan efisiensi penggunaan energi. Simulasi sederhana yang disajikan pada penelitian ini memperlihatkan bahwa proses pembelajaran berbasis pengalaman mampu menghasilkan peningkatan kualitas navigasi dibandingkan pendekatan pergerakan acak.

Meskipun demikian, hasil kajian juga mengidentifikasi beberapa tantangan utama dalam implementasi Reinforcement Learning pada lingkungan nyata. Tantangan tersebut meliputi kebutuhan data pelatihan yang besar, waktu pelatihan yang relatif panjang, tingginya kebutuhan sumber daya komputasi, serta kompleksitas koordinasi pada sistem yang melibatkan banyak agen (*multi-agent system*). Selain itu, perancangan *reward function* yang kurang tepat dapat menyebabkan agen mempelajari perilaku yang tidak sesuai dengan tujuan sistem sehingga proses pembelajaran menjadi kurang optimal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa Reinforcement Learning merupakan salah satu pendekatan yang memiliki potensi besar dalam mendukung pengembangan

sistem cerdas adaptif pada era Industri 5.0. Kemampuan belajar dari pengalaman, melakukan adaptasi terhadap perubahan lingkungan, serta mengoptimalkan pengambilan keputusan menjadikan Reinforcement Learning sebagai teknologi yang relevan untuk berbagai aplikasi, khususnya pada sistem logistik cerdas, robotika, manufaktur, dan otomasi industri.

5. SARAN

Berdasarkan hasil kajian yang telah dilakukan, terdapat beberapa arah penelitian yang berpotensi meningkatkan efektivitas implementasi Reinforcement Learning pada robot gudang otonom maupun sistem cerdas lainnya.

Pertama, penelitian selanjutnya dapat mengembangkan implementasi Deep Reinforcement Learning sebagai pengganti algoritma Q-Learning konvensional. Pendekatan ini memungkinkan agen menangani lingkungan dengan ruang keadaan yang jauh lebih besar melalui representasi fitur menggunakan jaringan saraf tiruan. Penggunaan algoritma seperti Deep Q-Network (DQN), Proximal Policy Optimization (PPO), atau Soft Actor-Critic (SAC) diperkirakan mampu meningkatkan kemampuan adaptasi robot terhadap kondisi gudang yang lebih kompleks.

Kedua, pengembangan Multi-Agent Reinforcement Learning (MARL) menjadi arah penelitian yang penting mengingat sebagian besar gudang modern menggunakan banyak robot yang bekerja secara bersamaan. Melalui pendekatan ini, setiap robot tidak hanya belajar secara individual, tetapi juga mampu berkoordinasi dengan robot lain untuk mengurangi kemacetan, menghindari konflik jalur, serta meningkatkan efisiensi distribusi barang.

Ketiga, penelitian mendatang dapat mengintegrasikan Reinforcement Learning dengan Digital Twin sehingga proses pelatihan dilakukan pada lingkungan virtual yang merepresentasikan kondisi gudang sebenarnya. Pendekatan ini memungkinkan evaluasi berbagai skenario operasional tanpa mengganggu aktivitas produksi serta mengurangi risiko kerusakan peralatan selama proses pembelajaran.

Keempat, integrasi Reinforcement Learning dengan teknologi Internet of Things (IoT), sensor cerdas, dan sistem pemantauan berbasis *real-time* perlu dieksplorasi lebih lanjut. Data yang diperoleh dari berbagai sensor dapat dimanfaatkan sebagai representasi *state* yang lebih akurat sehingga keputusan yang dihasilkan robot menjadi lebih responsif terhadap perubahan kondisi lingkungan.

Kelima, penelitian selanjutnya dapat mengkaji penerapan Explainable Reinforcement Learning (XRL) untuk meningkatkan transparansi proses pengambilan keputusan oleh agen cerdas. Kemampuan menjelaskan alasan di balik setiap keputusan menjadi aspek penting dalam implementasi sistem AI pada lingkungan industri yang memerlukan tingkat kepercayaan dan keselamatan yang tinggi.

Terakhir, diperlukan penelitian eksperimental yang membandingkan kinerja berbagai algoritma Reinforcement Learning pada lingkungan gudang yang sama menggunakan indikator evaluasi yang terukur, seperti waktu penyelesaian tugas, konsumsi energi, tingkat keberhasilan pengiriman, jumlah tabrakan, utilisasi robot, serta efisiensi operasional. Hasil penelitian tersebut diharapkan dapat memberikan rekomendasi empiris mengenai algoritma yang paling sesuai untuk berbagai skenario pergudangan modern.

Dengan mempertimbangkan berbagai peluang tersebut, Reinforcement Learning diperkirakan akan menjadi salah satu teknologi kunci dalam pengembangan smart warehouse, autonomous logistics, dan cyber-physical systems yang mendukung transformasi industri menuju ekosistem manufaktur yang lebih cerdas, adaptif, dan berkelanjutan.

DAFTAR PUSTAKA

- [1] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- [2] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- [3] Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- [4] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.
- [5] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). *Deep reinforcement learning: A brief survey*. IEEE Signal Processing Magazine, 34(6), 26–38.
- [6] Wurman, P. R., D'Andrea, R., & Mountz, M. (2008). Coordinating hundreds of cooperative, autonomous vehicles in warehouses. *AI Magazine*, 29(1), 9–20.
- [7] Gu, T., Dolan, J. M., & Lee, J. (2021). Reinforcement learning for warehouse robot path planning: A survey. *Robotics and Autonomous Systems*, 140, 103748.
- [8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [9] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237–285.
- [10] Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, 32(11), 1238–1274.
- [11] Nahavandi, S. (2019). Industry 5.0—A human-centric solution. *Sustainability*, 11(16), 4371.
- [12] Li, Y. (2023). Deep reinforcement learning: Recent advances and applications. *Artificial Intelligence Review*, 56(4), 3151–3184.
- [13] Kitchenham, B., & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Keele University.
- [14] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*. *BMJ*, 372, n71.
- [15] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339.
- [16] Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. ESRC Methods Programme.

- [17] Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.
- [18] Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3–4), 279–292.
- [19] Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. *NeurIPS*.
- [20] Mnih, V., Badia, A. P., Mirza, M., et al. (2016). *Asynchronous Methods for Deep Reinforcement Learning*. ICML.
- [21] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor*. ICML.
- [22] Ng, A. Y., Harada, D., & Russell, S. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. ICML.
- [23] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal Policy Optimization Algorithms*. arXiv:1707.06347.