Opponent Modelling dalam Interaksi Strategis Berulang: Asumsi, Metode, dan Praktik Evaluasi

Faqih Mahardika
Department of Computer Science and Electronics
Gadjah Mada University
Yogyakarta, Indonesia
fm.faqihmahardika@gmail.com

***Abstract*—Kerja sama dan konflik merupakan karakteristik fundamental dari interaksi dalam masyarakat manusia, sistem biologis, dan lingkungan agen-artifisial, di mana agen secara berulang menghadapi pertukaran strategis antara kepentingan diri jangka pendek dan hasil kolektif jangka panjang. *Opponent Modelling* memainkan peran sentral dalam konteks tersebut dengan memungkinkan agen untuk menginferensi, mengantisipasi, dan beradaptasi terhadap perilaku pihak lain, dengan aplikasi yang mencakup negosiasi, interaksi pasar, hingga sistem kecerdasan buatan multi-agen. Tinjauan literatur ini mensintesis penelitian terdahulu mengenai *Opponent Modelling* dalam permainan strategis berulang, dengan menggunakan *Iterated Prisoner's Dilemma* sebagai instansiasi kanonik dari dilema sosial berulang, alih-alih sebagai domain yang membatasi. Tinjauan ini menganalisis pendekatan yang ada berdasarkan tiga dimensi: asumsi terhadap perilaku lawan, metodologi *opponent modelling*, dan praktik evaluasi. Literatur yang disurvei menunjukkan keberagaman yang signifikan dalam asumsi perilaku lawan, termasuk lawan yang stasioner, reaktif, berbasis pembelajaran, dan dimediasi oleh populasi, yang sering kali tersemat secara implisit dalam rancangan eksperimen. Secara metodologis, pendekatan yang digunakan mencakup pembelajaran berbasis gradien, *deep reinforcement learning*, penalaran kepercayaan rekursif, dinamika evolusioner, dan *system identification*. Praktik evaluasi didominasi oleh metrik berbasis hasil, seperti tingkat kerja sama, *average payoff*, dan ukuran terkait ekuilibrium. Meskipun informatif untuk kinerja jangka panjang, metrik-metrik tersebut umumnya diterapkan pada horizon interaksi yang tidak dibatasi, sehingga membatasi pemahaman mengenai efisiensi dan ketepatan waktu dalam proses identifikasi serta adaptasi terhadap lawan. Tinjauan ini menyoroti adanya kesenjangan struktural dalam praktik evaluasi yang ada dan menekankan perlunya kerangka penilaian yang sadar-horizon (*horizon-aware*) agar lebih merefleksikan keterbatasan interaksi berulang di dunia nyata.**

***Index Terms*—Opponent Modelling, Iterated Prisoner's Dilemma, Online Adaptation, Repeated Games**

I. Introduction

Kerja sama dan konflik merupakan karakteristik fundamental dari interaksi dalam masyarakat manusia, sistem biologis, dan lingkungan agen-artifisial. Teori permainan menyediakan kerangka analitis yang ketat untuk mempelajari saling ketergantungan strategis semacam ini, dan karya-karya fundamental mengenai evolusi kerja sama menunjukkan bagaimana pola perilaku dapat muncul ketika agen secara berulang menghadapi dilema yang menuntut keseimbangan antara kepentingan diri dan manfaat bersama [2], [3]. Di antara berbagai model tersebut, *Prisoner's Dilemma* (PD) telah menjadi representasi kanonik dari situasi di mana rasionalitas individual bertentangan dengan kesejahteraan kolektif. Dalam PD satu kali permainan (*one-shot PD*), kerja sama mutual menghasilkan *joint payoff* tertinggi, namun struktur insentifnya mendorong pemain rasional untuk melakukan defeksi, sehingga menyingkap ketegangan inti yang berulang dalam berbagai konteks sosial, ekonomi, dan politik.

Ketika permainan tersebut diulang dalam bentuk *Iterated Prisoner's Dilemma* (IPD), lanskap strategisnya berubah secara signifikan. Pengulangan memungkinkan agen untuk mengondisikan tindakannya berdasarkan interaksi sebelumnya, sehingga memunculkan perilaku seperti resiprositas, pembalasan, pemaafan, dan pembangunan kepercayaan. Turnamen IPD yang berpengaruh dari Axelrod menunjukkan bahwa strategi kontingen sederhana terutama *Tit-for-Tat* dapat menumbuhkan kerja sama yang stabil bahkan dalam lingkungan yang kompetitif, sekaligus mengilustrasikan bagaimana interaksi jangka panjang dan memori berkontribusi pada munculnya norma kerja sama tanpa penegakan terpusat [2], [3]. Temuan-temuan ini menegaskan IPD sebagai alat fundamental untuk menganalisis pengambilan keputusan adaptif yang bergantung pada sejarah interaksi.

Relevansi dinamika IPD meluas ke berbagai domain. Dalam biologi, teori *reciprocal altruism* menjelaskan bagaimana pertemuan berulang mendukung kerja sama kontingen, serta merumuskan kondisi di mana perilaku menolong dapat berevolusi di antara organisme yang berorientasi pada kepentingan diri [4]. *Evolutionary game theory* memformalkan mekanisme tersebut melalui konsep *evolutionarily stable strategies*, yang menunjukkan bagaimana aturan perilaku tertentu dapat bertahan di bawah tekanan seleksi [7]. Dalam ilmu sosial, kerangka permainan berulang digunakan untuk menjelaskan kerja sama manusia, penegakan norma, dan sistem hukuman dalam konteks *public goods* [5]. Dalam ekonomi dan pengelolaan sumber daya, permasalahan seperti perikanan komersial menunjukkan bagaimana saling ketergantungan strategis dalam lingkungan bersama mencerminkan struktur dilema berulang [6]. Di berbagai bidang tersebut, IPD berfungsi sebagai lensa yang fleksibel untuk memahami bagaimana kerja sama dapat muncul, menjadi stabil, atau runtuh, bergantung pada insentif, struktur informasi, dan kondisi institusional.

Meskipun analisis klasik terhadap IPD memberikan wawasan yang esensial, asumsi penyederhanaan yang digunakan sering kali menyimpang dari kompleksitas interaksi strategis di dunia nyata. Agen nyata baik manusia, biologis, maupun artifisial umumnya beroperasi di bawah ketidakpastian, memiliki observabilitas terbatas, dan menyesuaikan perilakunya secara dinamis. Kondisi-kondisi ini mendorong meningkatnya perhatian terhadap *opponent modelling*, yang membekali agen dengan kemampuan untuk menginferensi strategi, kecenderungan perilaku, atau intensi lawan berdasarkan aksi yang teramati. Survei-survei terkini menunjukkan spektrum luas teknik *opponent modelling* dalam domain adversarial dan multi-agen, serta menegaskan perannya dalam memungkinkan pengambilan keputusan adaptif pada permainan berulang, negosiasi, sistem otonom, dan lingkungan pembelajaran kompetitif [1]. Dalam konteks IPD, kemampuan ini bersifat krusial: keberlanjutan kerja sama bergantung pada kemampuan mengenali mitra kooperatif, mendeteksi pihak yang eksploitatif, serta beradaptasi secara efektif terhadap strategi yang non-stasioner atau menipu.

Oleh karena itu, penelitian ini bertujuan untuk melakukan tinjauan sistematis terhadap literatur yang ada mengenai *opponent modelling* dalam *Iterated Prisoner's Dilemma*. Tinjauan ini dipandu oleh pertanyaan penelitian berikut:

- **RQ1:** Asumsi perilaku lawan apa saja yang digunakan dalam *opponent modelling* untuk *Repeated Games*?
- **RQ2:** Pendekatan metodologis apa saja dalam *opponent modelling* yang telah diterapkan pada *Repeated Games*?
- **RQ3:** Bagaimana efektivitas strategi *opponent modelling* dievaluasi dalam *Repeated Games*?

Untuk menjawab pertanyaan-pertanyaan tersebut, Bagian 2 menguraikan metodologi *Systematic Literature Review* (SLR) yang digunakan untuk mengidentifikasi, melakukan penyaringan, dan mensintesis studi-studi yang relevan. Bagian 3 menyajikan hasil dan pembahasan tinjauan dari tiga perspektif analitis: (i) perbandingan asumsi perilaku lawan, (ii) perbandingan pendekatan metodologis, dan (iii) perbandingan berbasis evaluasi. Bagian 4 menyimpulkan tinjauan literatur ini.

II. Methodology

Tinjauan sistematis digunakan untuk menjawab pertanyaan-pertanyaan penelitian yang diajukan. Metode ini menerapkan prosedur yang transparan dan terstruktur untuk mengumpulkan studi-studi yang relevan, menilai kualitasnya, serta mengintegrasikan temuan-temuannya guna menjawab pertanyaan penelitian yang terdefinisi dengan jelas [8].

A. Inclusion and Exclusion Criteria

Studi dimasukkan ke dalam tinjauan apabila memenuhi seluruh kriteria berikut:

- **Relevansi topik:** Studi menyajikan pembahasan substantif, metode, atau analisis empiris mengenai *opponent modelling*, pengambilan keputusan yang sadar terhadap lawan (*opponent-aware decision making*), atau penalaran eksplisit terhadap strategi agen lain dalam *konteks interaksi strategis berulang*. Konteks ini mencakup, namun tidak terbatas pada, *Iterated Prisoner's Dilemma* (IPD), permainan *general-sum* berulang, permainan *zero-sum* berulang, serta dilema sosial berulang lain yang secara struktural sebanding, di mana agen mengondisikan perilaku berdasarkan aksi lawan yang teramati.
- **Kesebandingan struktural:** Permainan atau lingkungan interaksi melibatkan pertemuan berulang, saling ketergantungan strategis, serta aksi lawan yang dapat diamati, sehingga memungkinkan adaptasi atau penalaran terhadap perilaku lawan dari waktu ke waktu. IPD diperlakukan sebagai instansiasi kanonik dalam kelas permainan strategis berulang yang lebih luas ini.
- **Kualitas akademik:** Karya dipublikasikan pada jurnal bereputasi atau prosiding konferensi internasional yang melalui proses *peer-review*.
- **Rentang publikasi:** Studi dipublikasikan antara tahun 2019 hingga 2025. Rentang waktu ini dipilih untuk menangkap pendekatan *opponent modelling* kontemporer yang berkembang pada era modern sistem multi-agen berbasis pembelajaran.
- **Aksesibilitas:** Teks lengkap tersedia dalam bahasa Inggris dan dapat diakses untuk dianalisis.

Studi dikecualikan apabila memenuhi salah satu dari kondisi berikut:

- **Fokus tidak relevan:** Studi tidak membahas *opponent modelling*, adaptasi yang sadar terhadap lawan, atau penalaran terhadap agen lain, atau tidak berada dalam konteks interaksi berulang.
- **Pembelajaran multi-agen generik:** Studi menerapkan *multi-agent reinforcement learning* atau teknik *learning-in-games* tanpa mengintegrasikan adaptasi yang sadar terhadap lawan, baik secara eksplisit maupun implisit.
- **Murni teoretis tanpa penalaran lawan:** Makalah hanya menyajikan analisis teoretis yang bersifat abstrak atau normatif dan tidak mengusulkan, menganalisis, maupun mengoperasionalkan mekanisme apa pun untuk *opponent modelling*, inferensi lawan, atau pengambilan keputusan yang sadar terhadap lawan dalam konteks interaksi berulang.
- **Di luar rentang publikasi:** Studi dipublikasikan di luar rentang waktu enam tahun yang telah ditetapkan.
- **Non-Inggris atau tidak dapat diakses:** Teks lengkap tidak tersedia dalam bahasa Inggris atau tidak dapat diakses.

B. Information Sources and Search Strategy

Penelusuran literatur dilakukan menggunakan empat basis data akademik utama, yaitu IEEE Xplore, ScienceDirect, Scopus, dan SpringerLink. Strategi pencarian memprioritaskan formulasi dilema berulang yang bersifat kanonik serta karya-karya yang sadar terhadap lawan (*opponent-aware*) untuk menjaga

fokus konseptual. Literatur pembelajaran multi-agen yang lebih luas turut dipertimbangkan apabila terdapat diferensiasi lawan, namun tidak ditinjau secara ekstensif. Kueri berikut diterapkan secara konsisten di seluruh basis data:

```
("opponent modelling" OR "opponent
 modeling" OR "opponent model" OR
        "agent modeling")
              AND
("prisoner's dilemma" OR "iterated
 prisoner's dilemma" OR "IPD" OR
"repeated game" OR "social dilemma")
```

Filter spesifik basis data diterapkan untuk memastikan relevansi dan aksesibilitas publikasi yang diperoleh:

1) **IEEE Xplore** (04–11–2025): Pencarian diterapkan pada seluruh bidang.
2) **ScienceDirect** (04–11–2025): Pencarian diterapkan pada seluruh bidang dan dibatasi pada publikasi *Open Access*.
3) **Scopus** (04–11–2025): Pencarian diterapkan pada seluruh bidang.
4) **SpringerLink** (04–11–2025): Pencarian diterapkan pada seluruh bidang dan dibatasi pada publikasi *Open Access*.

C. Research Procedure

Penelusuran literatur dilakukan dengan mengikuti strategi pencarian yang telah ditetapkan, dengan penyesuaian kata kunci terhadap kapabilitas kueri spesifik dari masing-masing basis data. Seluruh rekaman publikasi yang teridentifikasi dikumpulkan dan dikelola menggunakan *reference manager* untuk menghilangkan duplikasi serta entri yang tidak memenuhi kriteria pada tahap penyaringan awal. Studi yang tersisa kemudian disaring berdasarkan judul dan abstraknya untuk menentukan relevansi. Versi teks lengkap dari studi yang lolos pada tahap ini selanjutnya diperoleh dan dinilai kelayakannya sebagai bagian dari tinjauan sistematis.

Ekstraksi data difokuskan pada pengumpulan informasi yang relevan dengan *Opponent Modelling* dalam *Iterated Prisoner's Dilemma*, termasuk pengaturan lingkungan, pendekatan pemodelan, dan metrik evaluasi. Data yang diekstraksi kemudian dianalisis dan dibandingkan untuk mengidentifikasi pola dari berbagai pendekatan yang ada. Temuan-temuan tersebut selanjutnya digunakan untuk menjawab pertanyaan penelitian yang telah dirumuskan dalam tinjauan ini.

III. Results

Pada setiap tahap proses seleksi yang diilustrasikan pada Gambar 1, simbol n menyatakan jumlah studi yang dipertimbangkan. Pada tahap identifikasi, sebanyak 10 rekaman dihapus sebagai duplikasi berdasarkan judul dan abstraknya. Selain itu, 48 rekaman dikeluarkan karena tidak memenuhi kriteria inklusi. Pada tahap penyaringan (*screening*), 26 makalah dieliminasi karena tidak relevan dengan penelitian ini (misalnya tidak melibatkan *opponent modelling* atau tidak menggunakan IPD sebagai konteks penelitian), dan 6 makalah dikeluarkan karena merupakan studi non-primer, sehingga total terdapat 32 makalah yang dieliminasi pada tahap ini. Satu rekaman tidak dapat diperoleh karena keterbatasan akses. Dengan demikian, sebanyak 20 rekaman dipilih untuk ditinjau pada tahap teks lengkap (*full-text review*).

Dari 20 artikel teks lengkap yang diperoleh, 2 dikeluarkan karena berfokus pada simulasi populasi IPD atau tidak menyajikan bentuk *opponent modelling* apa pun. Dua artikel lainnya dikeluarkan karena merupakan studi non-primer. Tiga makalah tambahan dieliminasi karena hanya membahas aspek teoretis dari IPD.

Setelah seluruh kriteria eksklusi diterapkan, sebanyak 13 studi dimasukkan ke dalam tinjauan sistematis final. Judul dan karakteristik dari studi-studi tersebut disajikan pada Tabel I.

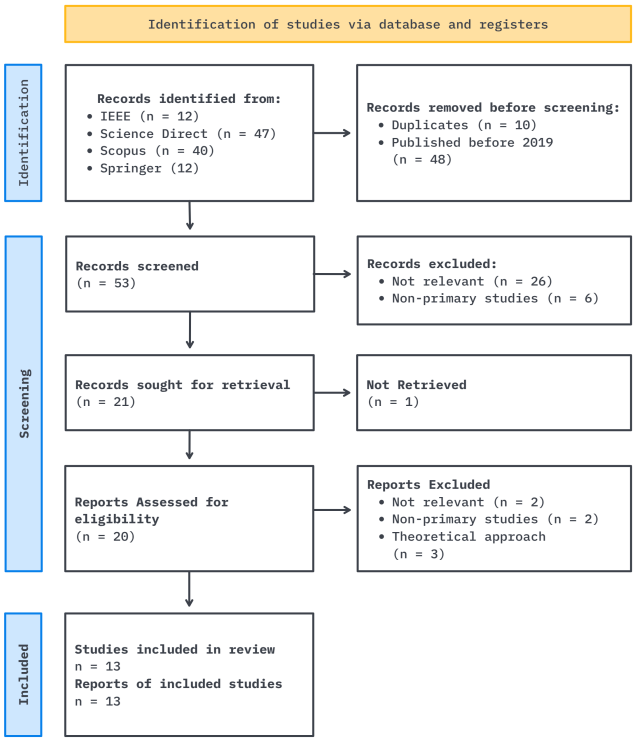

Fig. 1. Diagram alur seleksi studi.

IV. Synthesis / Discussion

Bagian ini mengintegrasikan temuan dari berbagai perspektif analitis untuk mengidentifikasi keterbatasan yang bersifat persisten serta merumuskan celah penelitian. Meskipun sejumlah atribut diekstraksi dari studi-studi yang disertakan, hanya dimensi yang bersifat diskriminatif secara metodologis yang digunakan untuk perbandingan langsung; atribut lingkungan dan evaluasi dirangkum secara terpisah.

A. Asumsi perilaku lawan apa saja yang digunakan dalam opponent modelling untuk Repeated Games?

Untuk mengoperasionalkan *opponent behavior assumptions* pada formulasi yang heterogen, tinjauan ini merangkum properti perilaku yang dapat diamati, alih-alih mekanisme pembela-

TABLE I
KARAKTERISTIK STUDI YANG DISERTAKAN

Ref.	Title	Characteristics
[9]	O2M: Online Opponent Modeling in Online General-Sum Matrix Games	*Online opponent modelling* untuk permainan *general-sum* dinamis.
[10]	Evolutionary Dynamics of Cooperation and Extortion on Networks With Fitness-Dependent Rules	Menganalisis strategi *zero-determinant*.
[11]	Inducing Coordination in Multi-Agent Repeated Game through Hierarchical Gifting Policies	Mengusulkan mekanisme *hierarchical gifting*.
[12]	Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach	Memodelkan kerja tim sukarela.
[13]	The coupling effect between the environment and strategies drives the emergence of group cooperation	Menganalisis keterkaitan memori dan lingkungan.
[14]	Learning to cooperate against ensembles of diverse opponents	Pendekatan RL yang skalabel untuk mendorong kerja sama.
[15]	Exploiting a No-Regret Opponent in Repeated Zero-Sum Games	Kerangka untuk mengeksploitasi lawan *no-regret*.
[16]	Modeling Theory of Mind in Dyadic Games Using Adaptive Feedback Control	Mengusulkan agen *Theory of Mind* berbasis kontrol.
[17]	Modeling opponent learning in multi-agent repeated games	Memperkenalkan metode berbasis Stackelberg.
[18]	The effects of population size and information update rates on the emergent patterns of cooperative clusters in a large-scale social particle swarm model	Pembaruan informasi cepat dalam jaringan SNS.
[19]	Achieving cooperation through deep multiagent reinforcement learning in sequential prisoner's dilemmas	Memperkenalkan PD sekuensial dan metode *deep RL*.
[20]	Achieving collective welfare in multi-agent reinforcement learning via suggestion sharing	Metode MARL dengan berbagi saran aksi.
[21]	Higher-order theory of mind is especially useful in unpredictable negotiations	Menganalisis *higher-order Theory of Mind* dalam lingkungan yang tidak terprediksi.

jaran internal. Tabel di bawah menyajikan kategorisasi deskriptif berdasarkan apakah aksi lawan (i) bergantung pada kondisi lingkungan, (ii) dimediasi oleh interaksi pada tingkat populasi, (iii) merespons secara langsung aksi agen fokus, dan (iv) menunjukkan divergensi aksi pada riwayat interaksi terkini yang ekuivalen. Kriteria-kriteria ini secara sengaja berorientasi pada perilaku dan dievaluasi pada tingkat jejak interaksi (*interaction traces*), tanpa mengasumsikan adanya akses terhadap model internal lawan, aturan pembaruan, maupun tujuan optimisasi yang digunakan.

Kategorisasi ini tidak dimaksudkan sebagai taksonomi formal atau komprehensif dari metode *opponent modelling*. Sebaliknya, kategorisasi ini berfungsi sebagai alat analitis untuk mendukung perbandingan lintas studi yang mengadopsi pengaturan permainan, paradigma pembelajaran, dan abstraksi pemodelan yang berbeda. Dalam kasus di mana sebuah makalah tidak mendefinisikan asumsi lawan secara eksplisit, klasifikasi diturunkan secara konservatif dari pengaturan eksperimen dan dinamika interaksi yang dilaporkan.

TABLE II
ASUMSI PERILAKU LAWAN BERDASARKAN DEPENDENSI PERILAKU YANG DAPAT DIAMATI.

Env.	Pop.	Agent	Div.	Kategori Perilaku Lawan	Ref.
–	–	✓	–	Reactive	[20]
–	✓	–	–	Population-Conformist	[12]
–	✓	✓	–	Contextual Reactive	[18]
–	–	✓	✓	Learning Opponent	[9], [11], [15]–[17], [19], [21]
–	✓	✓	✓	Population-Contextual Strategic	[14]
✓	✓	–	✓	Heterogeneous Collective Behavior	[10]
✓	✓	✓	✓	Environment-Conditioned Strategic	[13]

Catatan:
Env. — perilaku bervariasi lintas lingkungan dalam permainan yang sama;
Pop. — perilaku dimediasi oleh interaksi tingkat populasi;
Agent — perilaku merespons secara langsung aksi agen fokus;
Div. — divergensi aksi terjadi pada riwayat interaksi terkini yang ekuivalen.
Tanda centang menunjukkan keberadaan suatu dependensi.

Sejumlah pola yang konsisten muncul dari kategorisasi perilaku pada Tabel II. Sebagian besar studi terkini memodelkan lawan yang aksinya secara eksplisit responsif terhadap agen dan menunjukkan divergensi aksi [9]–[11], [13]–[17], [19], [21], yang mengindikasikan asumsi adanya pembelajaran atau adaptasi strategis. Hal ini mencerminkan meningkatnya penekanan riset pada ketahanan terhadap lawan yang non-stasioner dan adaptif, dibandingkan optimisasi terhadap strategi yang tetap.

Selain itu, perilaku yang dimediasi oleh populasi terutama muncul dalam studi-studi pada lingkungan evolusioner atau berbasis jaringan, di mana aksi lawan dibentuk secara tidak langsung melalui dinamika agregat alih-alih adaptasi bilateral secara langsung [10], [12]–[14], [18]. Pengaturan semacam ini sering kali memisahkan responsivitas agen individual dari perubahan pada tingkat populasi, sehingga menghasilkan pola perilaku yang berbeda secara kualitatif dibandingkan dengan skenario pembelajaran berpasangan. Secara khusus, hanya sebagian kecil karya yang mengombinasikan perilaku yang bergantung pada lingkungan, dimediasi populasi, dan responsif terhadap agen secara simultan [13], yang menunjukkan bahwa lawan strategis yang sepenuhnya terkondisi oleh konteks masih relatif kurang dieksplorasi.

B. Pendekatan metodologis apa saja dalam opponent modelling yang telah diterapkan pada Repeated Games?

Distribusi asumsi perilaku lawan pada Tabel II merefleksikan pergeseran metodologis yang lebih luas dalam riset *opponent modelling*, dari asumsi yang terkontrol dan stasioner menuju lawan yang kaya secara perilaku, adaptif, dan heterogen. Meskipun karakterisasi tersebut menyoroti sifat lawan yang dipertimbangkan, pendekatan ini belum menangkap bagaimana agen dirancang untuk menghadapi kompleksitas tersebut. Oleh karena itu, pada Tabel III penelitian-penelitian terdahulu direorganisasi dari perspektif pemodelan di sisi agen, dengan mengelompokkan pendekatan berdasarkan paradigma *opponent modelling* yang dominan serta mekanisme pembelajaran yang menyertainya.

TABLE III
PERBANDINGAN PENDEKATAN *opponent modelling* BERDASARKAN PARADIGMA PEMODELAN DOMINAN DAN MEKANISME PEMBELAJARAN.

Paradigma Pemodelan	**Mekanisme Pembelajaran / Pembaruan**	**Makalah**
Reactive reinforcement learning	RL berbasis nilai (DQN)	[11]
Gradient-based opponent shaping	*Gradient descent* / pembaruan sadar lawan	[9], [17], [19]
Recursive belief reasoning	Pembaruan keyakinan Bayesian / *cognitive hierarchy*	[16], [21]
Population-based training	*Policy-gradient* MARL (sampling populasi)	[14]
Evolutionary population dynamics	Imitasi strategi berbasis *fitness*	[10], [13], [18]
Bandit-based learning-in-games	*Multi-armed bandit* / *smooth best response*	[12]
System identification	Model autoregresif dengan input eksogen (NARX)	[15]
Communication-driven coordination	*Policy-gradient* dengan objektif komunikasi	[20]

Tabel III mengelompokkan karya-karya terdahulu berdasarkan paradigma pemodelan dan mekanisme pembelajaran. Pendekatan-pendekatan tersebut juga berbeda dalam hal apakah adaptasi terhadap lawan ditangani secara eksplisit atau implisit. Pendekatan *opponent modelling* yang eksplisit, seperti *gradient-based opponent shaping* [9], [17], [19], *recursive belief reasoning* [16], [21], dan *system identification* [15], membangun representasi internal terhadap perilaku lawan. Sebaliknya, pendekatan implisit termasuk *reactive reinforcement learning* [11], *population-based training* [14], dinamika evolusioner [10], [13], [18], serta *bandit-based learning-in-games* [12], beradaptasi tanpa mempertahankan model lawan yang eksplisit. Perbedaan ini menyoroti filosofi perancangan alternatif dalam menghadapi ketidakstasioneran perilaku lawan.

C. Bagaimana efektivitas strategi opponent modelling dievaluasi dalam Iterated Prisoner's Dilemma?

Pilihan evaluasi secara implisit mendefinisikan apa yang dianggap sebagai keberhasilan dalam interaksi multi-agent yang adaptif, seperti hasil kerja sama, ketahanan terhadap eksploitasi, stabilitas perilaku, atau akurasi prediksi. Tabel IV menyajikan ikhtisar deskriptif mengenai lingkungan evaluasi dan metrik yang digunakan dalam penelitian-penelitian sebelumnya, dengan tujuan memungkinkan perbandingan lintas studi serta menyoroti pola evaluasi yang berulang maupun aspek-aspek yang masih terabaikan, alih-alih untuk melakukan standarisasi atau pemeringkatan kriteria.

TABLE IV
LINGKUNGAN EVALUASI DAN METRIK DALAM PENELITIAN

Ref.	**Lingkungan Evaluasi**	**Metrik**
[9]	Self-play simetris	MSE selama pelatihan offline; akurasi memori laten
[10]	Jaringan scale-free dengan strategi zero-determinant	Frekuensi kerja sama (C) dan eksploitasi (E)
[11]	Opponent adaptif (dirata-ratakan pada beberapa opponent)	Nilai reward
[12]	Simulator teamwork-game khusus (aggregative public good games); eksperimen sintetis	Produktivitas tim agregat; uji kecocokan χ^2 terhadap equilibrium; konvergensi ke Nash equilibrium; kontribusi individu
[13]	Simulator evolutionary game pada jaringan terstruktur	Tingkat kerja sama; fraksi kooperator; ambang fase transisi
[14]	Repeated matrix games dengan populasi opponent sintetis	Payoff rata-rata; tingkat kerja sama; robustness terhadap himpunan opponent; generalisasi
[15]	Repeated zero-sum games melawan Hedge, OMD, dan Regret Matching	Galat prediksi; payoff kumulatif; robustness terhadap non-stationarity
[16]	Repeated matrix games; simulasi robotik embodied waktu-kontinu	Efektivitas; stabilitas; akurasi prediksi
[17]	Simulasi repeated matrix game	Payoff rata-rata; kecepatan konvergensi; pemilihan equilibrium
[18]	Simulasi continuous-space skala besar (FLAME GPU)	Tingkat kerja sama; ukuran dan jumlah klaster kooperatif; kecepatan agen; stabilitas klaster
[19]	Lingkungan SPD 2D khusus (Fruit Gathering; Apple–Pear games)	Reward individu rata-rata; total kesejahteraan sosial; akurasi deteksi derajat kerja sama
[20]	Benchmark MARL (Cleanup, Harvest, Sequential PD, Tragedy of the Commons)	Return ternormalisasi; tingkat kerja sama; kecepatan konvergensi; perbedaan kebijakan (MSE)
[21]	Simulasi Colored Trails dengan peningkatan ketidakpastian lingkungan	Skor allocator; skor responder; total kesejahteraan sosial

Tabel IV menunjukkan adanya konsentrasi yang kuat pada praktik evaluasi berbasis keluaran kinerja agregat, khususnya tingkat kerja sama [13], [18]–[20], payoff rata-rata [11], [14], [15], [19], [21], serta konvergensi menuju equilibrium [12], [17], [20]. Metrik-metrik ini selaras dengan skenario interaksi berulang jangka panjang, di mana kerugian eksplorasi pada tahap awal dapat teramortisasi seiring waktu dan pola perilaku yang stabil pada akhirnya muncul.

Namun demikian, perhatian terhadap kendala interaksi relatif masih terbatas, seperti horizon permainan yang pendek, anggaran pembelajaran yang terbatas, atau biaya peluang yang timbul selama proses probing terhadap opponent. Bahkan ketika opponent non-stationary atau adaptif dipertimbangkan,

evaluasi umumnya dilakukan dalam kondisi yang mengasumsikan interaksi berkepanjangan [9], dan kinerja dinilai setelah proses konvergensi, bukan selama fase pembelajaran berlangsung.

Akibatnya, protokol evaluasi yang ada cenderung kurang merepresentasikan skenario di mana eksplorasi bersifat mahal, miskordinasi tidak dapat dipulihkan, atau keputusan awal sangat mendominasi total return. Kesenjangan ini menjadi sangat relevan dalam konteks online opponent modelling, di mana identifikasi perilaku opponent secara inheren memerlukan intervensi melalui aksi, tetapi biaya dari intervensi tersebut jarang diisolasi sebagai dimensi evaluasi tersendiri. Oleh karena itu, metrik yang digunakan saat ini masih memberikan wawasan yang terbatas mengenai seberapa efisien agen menyeimbangkan identifikasi opponent dengan kinerja langsung dalam pengaturan interaksi yang terkendala.

V. KESIMPULAN

Tinjauan ini mensintesis literatur opponent modelling dalam interaksi strategis berulang dengan mengkaji bagaimana perilaku opponent diasumsikan, bagaimana perilaku tersebut dimodelkan, serta bagaimana efektivitas pemodelannya dievaluasi, dengan fokus utama pada pengaturan dilema sosial kanonik yang umumnya diwujudkan melalui Iterated Prisoner's Dilemma. Analisis menunjukkan adanya keragaman yang signifikan dalam asumsi perilaku opponent, mulai dari opponent yang stasioner dan reaktif hingga dinamika yang adaptif dan dimediasi oleh populasi. Meskipun banyak penelitian secara implisit mempertimbangkan non-stationarity yang muncul akibat dinamika pembelajaran atau keterkaitan dengan lingkungan, asumsi-asumsi tersebut sering kali tertanam dalam desain eksperimen dan tidak dinyatakan secara eksplisit. Akibatnya, pendekatan pemodelan yang secara metodologis tampak serupa dapat beroperasi di bawah asumsi perilaku opponent yang secara fundamental berbeda, sehingga menyulitkan perbandingan langsung antar penelitian.

Dalam literatur yang ditinjau, beragam metodologi pemodelan telah digunakan, yang mencerminkan perbedaan pandangan mengenai tingkat abstraksi yang tepat untuk merepresentasikan opponent yang adaptif. Praktik evaluasi didominasi oleh metrik berbasis hasil, seperti tingkat kerja sama, payoff rata-rata, dan konvergensi menuju equilibrium. Metrik-metrik ini efektif untuk menilai kinerja dan stabilitas dalam horizon interaksi jangka panjang, tetapi umumnya diterapkan pada pengaturan di mana horizon interaksi tidak dibatasi atau cukup panjang untuk mengamortisasi biaya eksplorasi. Oleh karena itu, meskipun identifikasi dan adaptasi terhadap opponent merupakan motivasi utama dari opponent modelling, perhatian terhadap efisiensi dan ketepatan waktu proses identifikasi tersebut dalam horizon interaksi yang terbatas masih relatif minim. Temuan ini menyoroti peluang bagi penelitian selanjutnya untuk mengkaji efektivitas opponent modelling dalam pengaturan online, di mana kesempatan interaksi terbatas, eksplorasi bersifat mahal, dan adaptasi pada tahap awal menjadi faktor yang krusial.

REFERENCES

[1] S. B. Nashed and S. Zilberstein, "A Survey on Opponent Modeling in Adversarial Domains" unpublished.

[2] R. Axelrod, "The Evolution of Cooperation" unpublished.

[3] R. Axelrod and W. D. Hamilton, "The Evolution of Cooperation" *Science*, 1981.

[4] R. L. Trivers, "The Evolution of Reciprocal Altruism" *The Quarterly Review of Biology*, vol. 46, no. 1, pp. 35–57, 1971. Available: https://www.journals.uchicago.edu/doi/10.1086/406755

[5] Y. Chen, "Cooperation and punishment in public goods experiments (by Ernst Fehr and Simon Gächter)," in *The Art of Experimental Economics*, G. Charness and M. Pingle, Eds., London: Routledge, 2021, pp. 134–143. Available: taylorfrancis.com/books/9781003019121/chapters/10.4324/9781003019121-12

[6] A. Leung and A.-Y. Wang, "Analysis of Models for Commercial Fishing: Mathematical and Economical Aspects," *Econometrica*, vol. 44, no. 2, p. 295, 1976. Available: https://www.jstor.org/stable/1912725

[7] J. M. Smith, *Evolution and the Theory of Games*. (Publication year not provided.)

[8] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021.

[9] X. Qiao, C. Han, and T. Guo, "O2M: Online Opponent Modeling in Online General-Sum Matrix Games," in *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)*, Dec. 2024, pp. 358–361. [Online]. Available: https://ieeexplore.ieee.org/document/10899996/. doi: 10.1109/ICAIRC64177.2024.10899996.

[10] L. Zhu, Y. Zhu, and C. Xia, "Evolutionary Dynamics of Cooperation and Extortion on Networks With Fitness-Dependent Rules," in *2025 Joint International Conference on Automation-Intelligence-Safety (ICAIS) & International Symposium on Autonomous Systems (ISAS)*, May 2025, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/11051564/. doi: 10.1109/ICAISISAS64483.2025.11051564.

[11] M. Lv, J. Liu, B. Guo, Y. Ding, Y. Zhang, and Z. Yu, "Inducing Coordination in Multi-Agent Repeated Game through Hierarchical Gifting Policies," in *2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, Sep. 2023, pp. 279–287. [Online]. Available: https://ieeexplore.ieee.org/document/10298392/. doi: 10.1109/MASS58611.2023.00041.

[12] A. L. de A. Gómez, C. Sierra, and J. Sabater-Mir, "Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach," *Artificial Intelligence*, vol. 341, p. 104307, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370225000268. doi: 10.1016/j.artint.2025.104307.

[13] C. Di, Q. Zhou, J. Shen, J. Wang, R. Zhou, and T. Wang, "The coupling effect between the environment and strategies drives the emergence of group cooperation," *Chaos, Solitons & Fractals*, vol. 176, 2023, pp. 114138. doi: 10.1016/j.chaos.2023.114138.

[14] I. Perera, F. de Nijs, and J. Garcia, "Learning to cooperate against ensembles of diverse opponents," *Neural Computing and Applications*, vol. 37, no. 23, 2025, pp. 18835–18849. doi: 10.1007/s00521-024-10511-9.

[15] K. Li, W. Huang, C. Li, and X. Deng, "Exploiting a No-Regret Opponent in Repeated Zero-Sum Games" *Journal of Shanghai Jiaotong University (Science)*, vol. 30, no. 2, 2025, pp. 385–398. doi: 10.1007/s12204-023-2610-2.

[16] I. T. Freire, X. D. Arsiwalla, J. Y. Puigbò, and P. Verschure, "Modeling Theory of Mind in Dyadic Games Using Adaptive Feedback Control," *Information*, vol. 14, no. 8, 2023. doi: 10.3390/info14080441.

[17] Y. Hu, C. Han, H. Li, and T. Guo, "Modeling opponent learning in multiagent repeated games," *Applied Intelligence*, vol. 53, no. 13, 2023, pp. 17194–17210. doi: 10.1007/s10489-022-04249-x.

[18] Z. Elhamer, R. Suzuki, and T. Arita, "The effects of population size and information update rates on the emergent patterns of cooperative clusters in a large-scale social particle swarm model," *Artificial Life and Robotics*, vol. 25, no. 1, 2020, pp. 149–158. doi: 10.1007/s10015-019-00558-6.

[19] W. Wang, Y. Wang, J. Hao, and M. E. Taylor, "Achieving cooperation through deep multiagent reinforcement learning in sequential prisoner's dilemmas," in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3356464.3357712.

[20] Y. Jin, S. Wei, and G. Montana, "Achieving collective welfare in multi-agent reinforcement learning via suggestion sharing," *Machine Learning*, vol. 114, no. 8, 2025, p. 190. doi: 10.1007/s10994-025-06823-z.

[21] H. De Weerd, R. Verbrugge, and B. Verheij, "Higher-order theory of mind is especially useful in unpredictable negotiations," *Autonomous Agents and Multi-Agent Systems*, vol. 36, no. 2, 2022, p. 30. doi: 10.1007/s10458-022-09558-6.