

Magic: The Gathering als Benchmark für strategische Intelligenz in künstlichen Systemen: der Rahmen des Projekts cEDH

Richard Lam
Unabhängiger Forscher
Januar 2026

Abstract—Spiele haben historisch als Meilensteine zur Messung des Fortschritts der künstlichen Intelligenz gedient. Schach demonstrierte die Effizienz von Suchverfahren, Go hob tiefes bestärkendes Lernen hervor, und Poker förderte das Schließen unter verborgener Information. Diese Umgebungen bleiben jedoch durch feste Regeln und begrenzte Mehragentendynamiken eingeschränkt.

Dieser Beitrag schlägt kompetitives Magic: The Gathering, insbesondere das Mehrspieler-Commander-Format bekannt als competitive Elder Dragon Highlander (cEDH), als potenziellen Benchmark zur Bewertung umfassenderer Formen strategischer Intelligenz vor. Die Domäne vereint unvollständige Information, stochastische Dynamiken, symbolische Regelinteraktionen, sich entwickelnde strategische Metas sowie Mehragentenwettbewerb innerhalb einer einzigen Umgebung.

Wir stellen Projekt cEDH vor, ein deterministisches Simulationsframework, das Experimente mit KI-Agenten in dieser Domäne unterstützt. Ein struktureller Vergleich mit bestehenden Benchmarks sowie eine mathematische Einordnung der Entscheidungskomplexität legen nahe, dass konsistente Leistung in cEDH die Integration probabilistischen Schließens, symbolischer Manipulation, Gegnermodellierung und Planung über lange Horizonte erfordert. Diese Eigenschaften deuten darauf hin, dass die Domäne als aussagekräftige Testumgebung zur Bewertung allgemeiner Entscheidungssysteme dienen kann.

Index Terms—Künstliche Intelligenz, Spielkomplexität, Mehragentensysteme, Bestärkendes Lernen

I. EINLEITUNG

Spiele bieten kontrollierte Umgebungen zur Bewertung von Systemen künstlicher Intelligenz. Erfolge in Schach und Go zeigten Fortschritte in Suche und Lernen [1], [2], während Poker das Schließen unter Unsicherheit einführte [3], [4]. Die meisten als Benchmarks verwendeten Spiele basieren auf festen Regeln, Zwei-Spieler-Wettbewerb und relativ stabilen strategischen Räumen.

Reale Entscheidungsprobleme beinhalten häufig mehrere Agenten, Unsicherheit, symbolische Regelinteraktionen und sich verändernde Umgebungen. Kompetitives Magic: The Gathering, insbesondere das Mehrspieler-Commander-Format, weist diese Eigenschaften gleichzeitig auf. Im Gegensatz zu traditionellen Benchmarks verlangt diese Domäne von Agenten, sowohl über strukturierte Regeln als auch über emergentes strategisches Verhalten zu schließen. Wir argumentieren, dass diese Kombination sie zu einem geeigneten Kandidaten für die Bewertung allgemeinerer Formen strategischen Schließens und adaptiver Entscheidungsfindung macht.

II. STRUKTURELLE EIGENSCHAFTEN DER DOMÄNE

Magic vereint mehrere Dimensionen von Komplexität, die in bestehenden KI-Benchmarks selten gemeinsam auftreten.

Unvollständige Information: verdeckte Hände, unbekanntes Ziehreihenfolge und unvollständiges Wissen über Gegner.

Stochastische Entwicklung: zufällige Ziehungen erfordern Planung über Ergebnisverteilungen.

Symbolische Regelinteraktionen: Tausende von Karteneffekten interagieren über zeitliche Schichten und bedingte Logik.

Mehragentenwettbewerb: Spiele umfassen typischerweise drei bis vier Spieler und führen Koalitionsdynamiken sowie Nicht-Nullsummen-Anreize ein.

Sich entwickelnder strategischer Raum: neue Karten und Archetypen verändern kontinuierlich das optimale Spiel.

Diese Faktoren erzeugen eine Entscheidungsumgebung, die komplexen stochastischen Mehragentensystemen näherkommt als deterministischen Brettspielen. Insbesondere die Wechselwirkung zwischen symbolischen Regeln und probabilistischen Zustandsübergängen schafft Planungsprobleme, die weder rein kombinatorisch noch rein statistisch sind, sondern die Integration beider Schlussweisen erfordern.

III. MATHEMATISCHE EINORDNUNG DER KOMPLEXITÄT

Seien S der Zustandsraum und A die Aktionsmenge. Mehrspieler-Magic lässt sich natürlich als partiell beobachtbares stochastisches Spiel modellieren.

$$|S| \approx \sum_{d \in D} \prod_{i=1}^n |Z_{i,d}| \quad (1)$$

wobei D die Deckkonfigurationen darstellt und $Z_{i,d}$ die Zonenpermutationen für Spieler i bezeichnet. Da beide Terme groß sind, wächst der effektive Zustandsraum kombinatorisch.

$$b = f(|H|, |B|, |M|, |R|) \quad (2)$$

wobei H die Handgröße, B den Zustand des Spielfelds, M das verfügbare Mana und R reaktive Optionen bezeichnet. Dies erzeugt unregelmäßige Suchbäume, die sich von klassischen deterministischen Spielen unterscheiden.

Der erwartete Planungshorizont variiert zudem dynamisch mit der Dichte der Interaktionen auf dem Spielfeld. In hochinteraktiven Zuständen können sich die Konsequenzen einer

Entscheidung über mehrere ausgelöste Fähigkeiten und Reaktionen hinweg fortpflanzen, wodurch sich die für optimales Spiel erforderliche Suchtiefe effektiv erhöht.

IV. BENCHMARK-KONTEXT

TABLE I
REPRÄSENTATIVE EIGENSCHAFTEN VON KI-BENCHMARKS

| Spiel | Information | Agenten | Regelraum |
|--------|---------------|---------|-----------|
| Schach | Perfekt | 2 | Fest |
| Go | Perfekt | 2 | Fest |
| Poker | Unvollständig | 2–6 | Fest |
| cEDH | Unvollständig | 3–4 | Wachsend |

Traditionelle Benchmarks isolieren häufig einzelne Dimensionen der Komplexität. Magic integriert mehrere Dimensionen gleichzeitig, was darauf hindeutet, dass die Leistung in dieser Umgebung breitere strategische Fähigkeiten besser widerspiegeln kann.

V. FRAMEWORK DES PROJEKTS cEDH

Projekt cEDH implementiert eine deterministische Simulationsengine, die reproduzierbare Experimente mit KI-Agenten ermöglicht. Karteneffekte werden in strukturierter Form dargestellt, was eine konsistente Regelausführung und detaillierte Protokollierung erlaubt. Das Framework unterstützt großskaliges Selbstspiel, agentenübergreifende Evaluation und systematische Analyse des Entscheidungsverhaltens unter kontrollierten Bedingungen.

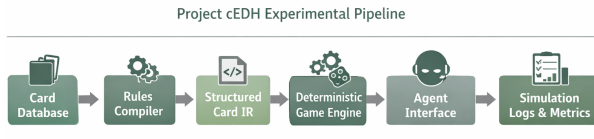


Fig. 1. Experimentelle Pipeline des Projekts cEDH. Kartendefinitionen werden in strukturierte Repräsentationen kompiliert, die deterministische Engine erzeugt reproduzierbare Spielzustände, und KI-Agenten interagieren über eine standardisierte Entscheidungsschnittstelle. Protokollierte Ergebnisse ermöglichen großskalige Evaluation und Training.

VI. EXPERIMENTELLE METHODIK

Die Evaluation erfolgt in drei Phasen. Heuristische Basisagenten etablieren Referenzleistungen und verifizieren die Regeltreue. Lernende Agenten werden durch großskaliges Selbstspiel über vielfältige Deckverteilungen trainiert, wodurch emergente Strategien und Anpassungsmuster beobachtet werden können. Die abschließende Evaluation vergleicht Agentenpopulationen mit synthetischen Gegnern

und erfahrenen menschlichen Spielern anhand von Metriken wie Gewinnrate, Entscheidungsstabilität, strategischer Vielfalt und Robustheit über unterschiedliche Spielzustände hinweg.

Dieser gestufte Ansatz ermöglicht die Messung sowohl algorithmischer Leistung als auch verhaltensbezogener Eigenschaften und erlaubt den Vergleich verschiedener Agentenarchitekturen sowie die Analyse der Entwicklung strategischer Kompetenz während des Trainings.

VII. SCHLUSSFOLGERUNG

Ein Agent, der in dieser Domäne konsistent erfolgreich ist, muss probabilistisches Schließen, symbolische Regelinterpretation, Gegnermodellierung und adaptive Planung integrieren. Diese Anforderungen gehen über die klassischer Benchmarks hinaus und ähneln den Fähigkeiten, die von allgemeineren Entscheidungssystemen erwartet werden.

Kompetitives Magic bietet daher eine vielversprechende Umgebung zur Bewertung umfassenderer Formen strategischer Intelligenz. Durch die Bereitstellung einer deterministischen experimentellen Plattform bei gleichzeitiger Wahrung hoher struktureller Komplexität ermöglicht Projekt cEDH die systematische Untersuchung von Entscheidungsfindung in Umgebungen, die realen strategischen Interaktionen näherkommen.

DANKSAGUNG

Der Autor dankt ChatGPT für die Unterstützung bei der Formatierung und der anfänglichen Ausarbeitung des Inhalts.

REFERENCES

- [1] M. Campbell, A. Hoane und F. Hsu, “Deep Blue,” *Artificial Intelligence*, 2002.
- [2] D. Silver et al., “Mastering the game of Go with deep neural networks and tree search,” *Nature*, 2016.
- [3] N. Brown und T. Sandholm, “Superhuman AI for multiplayer poker,” *Science*, 2019.
- [4] M. Moravčík et al., “DeepStack: Expert-level artificial intelligence in heads-up no-limit poker,” *Science*, 2017.
- [5] MIT Technology Review, “Magic: The Gathering and computational complexity,” 2020.