

Hochschule Hof  
Multimodal AI  
Wintersemester 2025 / 26

# Pre-Training Evaluation

Taha Faroukh Khapra   Fadhil Fasalu Rahiman   Naren Prakash  
00010623                    00006523                    00015623

**Ultimate Tic Tac Toe**

**Modell:** unsloth/Qwen3-VL-8B-Instruct-bnb-4bit

# 1 Merkmale des Datensatzes

Der Datensatz umfasst **4004** Datenpunkte. **1001** davon wurden manuell durch das Spielen des algorithmischen Bots gegen die Teammitglieder erstellt, die übrigen **3003** durch Drehung des ursprünglichen Datensatzes um  $90^\circ$ ,  $180^\circ$  und  $270^\circ$ . Alle Koordinaten und erforderlichen Informationen wurden entsprechend aktualisiert. Es handelt sich um einen **generativen VQA-Datensatz**, der es ermöglicht, während des Trainings und der Evaluierung dynamische Fragen zu stellen.

## 2 Fragen zur Evaluation

Das Modell wird anhand von vier hierarchischen Aufgaben evaluiert:

<b>Allowed Square</b>	Zuordnung der grünen Markierung in den Bildern zu logischen globalen Koordinaten (row, col)
<b>Move</b>	Strategische Auswahl des optimalen lokalen Zuges (row, col) basierend auf dem aktuellen Spielstand.
<b>State</b>	Lokale optische Zeichenerkennung zur Darstellung eines 3x3-Teilgitters als Textmatrix [X, O, .], wobei „.“ ein leeres Feld darstellt.
<b>Legality</b>	Mehrstufige Prüfung, ob ein vorgeschlagener Zug innerhalb der „Allowed Square“ liegt und ob das Teilgitter belegt ist.

## 3 Ergebnisse der Pre-Training Evaluation

Frage	Metrics	Results (%)
Allowed Square	Accuracy <sup>1</sup>	26,5
Move	Legality <sup>2</sup>	14,5
	Accuracy <sup>3</sup>	2,0
State	Similarity Score <sup>4</sup>	40,6
	Accuracy <sup>5</sup>	2,0
Legality	Logic <sup>6</sup>	43,0

<sup>1</sup> Die korrekte Identifizierung des grün markierten 3x3 Feldes.

<sup>2</sup> Der Anteil der vorhergesagten Züge, die den offiziellen Spielregeln entsprechen.

<sup>3</sup> Die exakte Übereinstimmung des gewählten Zuges mit dem berechneten „Best Move“ des Ground Truth.

<sup>4</sup> Die strukturelle Ähnlichkeit der vorhergesagten Matrix mit dem Ground Truth via Levenshtein-Distanz.

<sup>5</sup> Die gesamte Matrix gilt als korrekt, wenn alle 9 Felder zu 100% übereinstimmen..

<sup>6</sup> Die Korrektheit der mehrstufigen Begründung (Occupancy/Constraint) für die Gültigkeit eines Zuges.

Tabelle 1: Ergebnisse der Evaluation

## 4 Beispieldgebnis der Evaluation

.	.	.	0	.	X	0
.	.	.	.	.	.	X
.	.	X	.	.	.	.
<hr/>		.	.	X	.	.
.	.	0	.	0	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	0
<hr/>		.	.	.	.	.
.	.	.	.	X	.	X
.	.	.	.	.	.	.

Tabelle 2: Visualisierung des globalen Spielfelds

<b>Allowed Square</b>	<b>Prompt:</b> Identify the allowed square (active global subgrid) based on the green highlight. If none is highlighted, report null.
	<b>Raw Response:</b> "The allowed square is [3, 2]. "
	→ FAIL Ground Truth: [2, 2]
<b>Move</b>	<b>Prompt:</b> Visually analyze the board. It is Player O's turn. Identify the allowed square (green highlight) and then select the best move.
	<b>Raw Response:</b> "The allowed square is [2, 2] and the local subgrid [2, 2] is empty. The optimal move is [2, 2, 2, 2]. "
<b>State</b>	<b>Prompt:</b> Examine Global Subgrid (2, 2). Represent the 3x3 local grid state as a matrix of 'X', 'O', or '.' (Empty).
	<b>Raw Response:</b> ["X", "O", ". ."], [". .", ". .", "X"], ["X", "X", ". ."]
	→ FAIL Ground Truth: [". .", ". .", ". ."], ["X", ". .", ". ."], [". .", ". .", ". ."]
<b>Legality</b>	<b>Prompt:</b> Is it legal for Player O to play at Global(2,2), Local(1,1)? Inspect the square state, then check the allowed square constraint to give the verdict.
	<b>Raw Response:</b> "The target square is Local (1,1) within Global (2,2). This local grid contains a red X at (0,0) and a blue O at (0,1). The center (1,1) is empty. In Ultimate Tic-Tac-Toe, a player can only move in a local grid if their opponent has not yet won that local game. In Local(1,1), the current state is: [X, O; ?, ?]. There is no winner yet (no three-in-a-row/column/diagonal). Therefore, the move is legal. "
	→ FAIL Ground Truth: False / Illegal

## 5 Qualitative Fehleranalyse

Die Ergebnisse der Evaluation verdeutlichen die spezifischen Herausforderungen bei der Verarbeitung komplexer, geschachtelter Spielbretter durch VLMs.

### 1. Räumliche Lokalisierung:

Die geringe Genauigkeit bei der Identifizierung des Allowed Square deutet darauf hin, dass das Modell Schwierigkeiten hat, visuelle Marker (die grüne Hervorhebung) direkt in den Koordinatenraum zu übersetzen. Das Auftreten von „Out-of-Bounds“-Koordinaten mit Werten größer als 2 belegt zudem eine mangelnde interne Repräsentation der Spielfeldbegrenzung.

### 2. Regelverständnis vs. Perzeption:

Die niedrigen Legality-Werte resultieren hauptsächlich aus einer fehlerhaften Objekterkennung. Zwar versteht das Modell das Konzept der Legalität, es scheitert jedoch daran, X und O korrekt auf dem Brett zu lokalisieren. Dadurch werden bereits belegte Felder fälschlicherweise als leer („Empty“) eingestuft.

### 3. Strategische Tiefe:

Es ist zu erwarten, dass die Move Accuracy hinter einem klassischen Minimax-Algorithmus zurückbleibt. Während Minimax-Algorithmen für die Tiefensuche in Spielbäumen optimiert sind, fehlt einem 8B-Parameter-Modell ohne dedizierte SSearch-and-ReasoningArchitektur die Fähigkeit zur langfristigen Vorausplanung.

### 4. Zustands-Halluzinationen:

Die Diskrepanz zwischen Similarity Score und Exact Accuracy zeigt, dass das Modell zwar das allgemeine Layout erkennt, aber bei der Detailerfassung der lokalen Subgrids zu Halluzinationen neigt und fiktive Spielzustände generiert.

## 5.1 Optimierungsansätze zur Fehlerreduktion

### 1. Multi-Task Curriculum Learning:

Um die räumliche Wahrnehmung zu schärfen, sollte das Training das Modell gezielt mit Teilaufgaben konfrontieren. Durch die Gewichtung von Aufgaben wie der Identifizierung „Allowed-Square“ oder der Erkennung lokaler Zustände wird das Modell gezwungen, zunächst die visuelle Ebene zu stabilisieren, bevor es strategische Entscheidungen trifft.

### 2. Chain-of-Thought Reasoning:

Um Fehler zu vermeiden, sollte das Modell darauf trainiert werden, seine Antwort schrittweise aufzubauen. Anstatt direkt eine Koordinate auszugeben, muss das Modell zunächst einen visuellen Scan des relevanten Bereichs durchführen und alle belegten Felder explizit auflisten.

### 3. JSON Enforcement:

Die Einführung strenger Ausgaberegeln während des Feinabstimmungsprozesses sollte sicherstellen, dass das Modell lernt, Fehlalarme außerhalb des gültigen Bereichs zu unterdrücken und Antworten in einem maschinenlesbaren, gültigen Format zu liefern.

### 4. Dynamic Region-of-Interest Zooming:

Um die OCR-Präzision zu erhöhen, sollte ein hochauflösender Ausschnitt des aktiven Teilgitters als zusätzliche Eingabe dienen. Dies ermöglicht eine detailliertere Erfassung der X- und O-Markierungen, die in der Gesamtansicht unscharf sein könnten.