

Hochschule Hof
Multimodal AI
Wintersemester 2025 / 26

Post-Training Evaluation

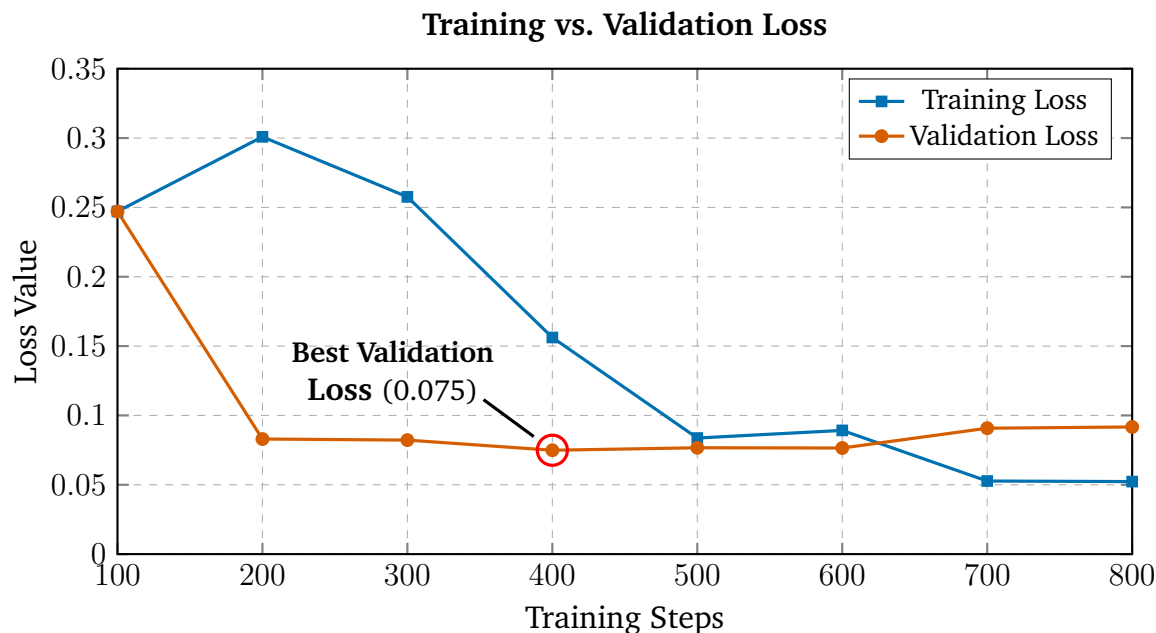
Taha Faroukh Khapra	Fadhil Fasalu Rahiman	Naren Prakash
00010623	00006523	00015623

Ultimate Tic Tac Toe

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

1 Training

Zur Optimierung wurde das Modell zunächst auf 1001 manuell erstellten Proben evaluiert, um die räumliche Integrität und Formatstabilität sicherzustellen. Die anschließende Skalierung auf 4004 Datenpunkte diente der Steigerung der Robustheit und der Verbesserung der Zustandserkennung. Parallel wurden die LoRA-Parameter von $r = 32$, $\alpha = 32$ über 64 auf 128 gesteigert, um eine ausreichende Modellkapazität für die komplexe hierarchische Spiellogik zu gewährleisten.



2 Ergebnisse der Post-Training Evaluation

Fragen	Metriken	Pre-Training (%)	Post-Training (%)
Allowed Square	Accuracy ¹	26,5	100,0
Move	Legality ²	14,5	93,8
	Accuracy ³	2,0	13,2
State	Similarity Score ⁴	40,6	92,0
	Accuracy ⁵	2,0	82,5
Legality	Logic ⁶	43,0	95,0

¹ Die korrekte Identifizierung des grün markierten 3x3 Feldes.
² Der Anteil der vorhergesagten Züge, die den offiziellen Spielregeln entsprechen.
³ Die exakte Übereinstimmung des gewählten Zuges mit dem berechneten „Best Move“ des Ground Truth.
⁴ Die strukturelle Ähnlichkeit der vorhergesagten Matrix mit dem Ground Truth via Levenshtein-Distanz.
⁵ Die gesamte Matrix gilt als korrekt, wenn alle 9 Felder zu 100% übereinstimmen..
⁶ Die Korrektheit der mehrstufigen Begründung (Occupancy/Constraint) für die Gültigkeit eines Zuges.

Tabelle 1: Ergebnisse der Evaluation

3 Beispielergebnis der Evaluation

0	0	0	X	.	.	X	.	X
0	0	0
.	X	0	X	X	X	X	.	.
X	.	.	X	.	.	X	.	0
.	.	.	.	X	0	.	.	.
X	.	X	0	X	.	X	X	.
0	0	0	0	0	.	X	0	.
0	X	X	0	0	0	.	.	.
X	0	0	0	X	.	X	X	0

Tabelle 2: Visualisierung des globalen Spielfelds

Allowed Square **Prompt:** Identify the allowed square (active global subgrid) based on the green highlight. If none is highlighted, report null.

Raw Response:

JSON_START

{"allowed_square ": null}

JSON_END

→ **SUCCESS** Ground Truth: None

Move **Prompt:** Visually analyze the board. It is Player O's turn. Identify the allowed square (green highlight) and then select the best move.

Raw Response:

JSON_START

{"allowed_square": null, "thinking": "No active constraint. Free move. Playing at global board (1,0) local position (0,2) secures a critical corner within that specific grid, potentially setting up future winning lines or blocking O's attempts to control that sector. This move sends the opponent to global board (0,2), where they have no immediate tactical advantage and must fight for basic territory rather than contesting more valuable central boards. By capturing this corner of the left-middle global cell, X strengthens their hold on the left column of the macro grid while forcing O into a less impactful quadrant.", "best_move": {"global_row": 1, "global_col": 0, "local_row": 0, "local_col": 2}}

JSON_END

State **Prompt:** Examine Global Subgrid (1, 1). Represent the 3x3 local grid state as a matrix of 'X', 'O', or '.' (Empty).

Raw Response:

JSON_START

{"target_global": [1, 1], "grid_matrix": [{"X", ".", "."], [".", "X", "O"], ["O", "X", "."]}]}

JSON_END

→ **SUCCESS** Ground Truth: [{"X", ".", "."], [".", "X", "O"], ["O", "X", "."]}]

Legality **Prompt:** Is it legal for Player X to play at Global(1,1), Local(1,1)? Step 1: Inspect the square state. Step 2: Check allowed square constraint. Step 3: Verdict.

Raw Response:
JSON_START
{"step_1_square_state": "X", "step_2_allowed_square": null, "is_legal": false, "reason": "SSquare is occupied by X."}
JSON_END

→ **SUCCESS** Ground Truth: False / Illegal

4 Vergleichende Analyse und Fortschrittsbewertung

Der Vergleich zeigt eine Leistungssteigerung in allen Metriken. Die Optimierungs-strategien haben die Schwächen bei räumlicher Wahrnehmung und logischer Konsistenz weitgehend behoben.

Räumliche Präzision

Während das Modell vor dem Training große Schwierigkeiten hatte, die grüne Hervorhebung zu interpretieren (26,5 % Accuracy), erreicht es nach dem Training eine perfekte Lokalisierung (100,0 %). Dies belegt, dass die Kombination aus Multi-Task-Learning und striktem Koordinatentraining die interne Repräsentation der Spielfeldbegrenzung stabilisiert hat.

Perzeptions-Halluzinationen

Ein kritischer Ausgangsfehler war die Halluzination fiktiver Spielfiguren. Durch gezieltes Training der Zustandserkennung konnte die Genauigkeit der Matrix von 2,0 % auf 82,5 % gesteigert werden. Das Modell war zudem in der Lage, das Bild erfolgreich zu scannen, was sich in dem hohen Ähnlichkeitswert von 92,0 % widerspiegelt.. Die Fehlerrate ist wahrscheinlich auf hochkomplexe Endspielszenarien mit hoher Figurendichte zurückzuführen und nicht mehr auf grundlegende Wahrnehmungsfehler.

Strategische Tiefe

Trotz der massiven Steigerung der Legalität bleibt die *Move Accuracy* mit 13,2 % vergleichsweise niedrig. Dies bestätigt die Hypothese aus der qualitativen Analyse: Ein 8B-Parameter-Modell kann zwar die Spielregeln perfekt erlernen und anwenden, besitzt jedoch ohne explizite Algorithmen nicht die strategische Tiefe, um gegen einen optimierten Minimax-Algorithmus zu bestehen. Das Modell spielt nun „regelkonform“, aber noch nicht „großmeisterlich“.

JSON Enforcement

Die Einführung des JSON-Protocolls und des Chain-of-Thought Reasoning hat die Logik-Metrik von 43,0 % auf 95,0 % gehoben. Die Modellantworten folgen nun konsistent dem maschinenlesbaren Format, was Parsing-Fehler eliminierte. Der Move Legality des Zuges stieg von 14,5 % auf 93,8 %, was zeigt, dass das Modell das Feld „Denken“ erfolgreich zur Regelprüfung nutzt.

5 Fazit

Die Evaluation zeigt, dass VLMs durch gezieltes Finetuning und strukturierte Antwortformate komplexe visuelle Logikaufgaben wie Ultimate Tic-Tac-Toe bewältigen können. Die Transformation von rein deskriptiver Bildanalyse zu einer regelbasierten Entscheidungsfindung verbesserte die von uns untersuchten Metriken und wertete das Training somit als Erfolg.