

# Interpretation der Trainingsergebnisse

DQN + Exploration-Strategien in Versicherungs-/Gym-Umgebung

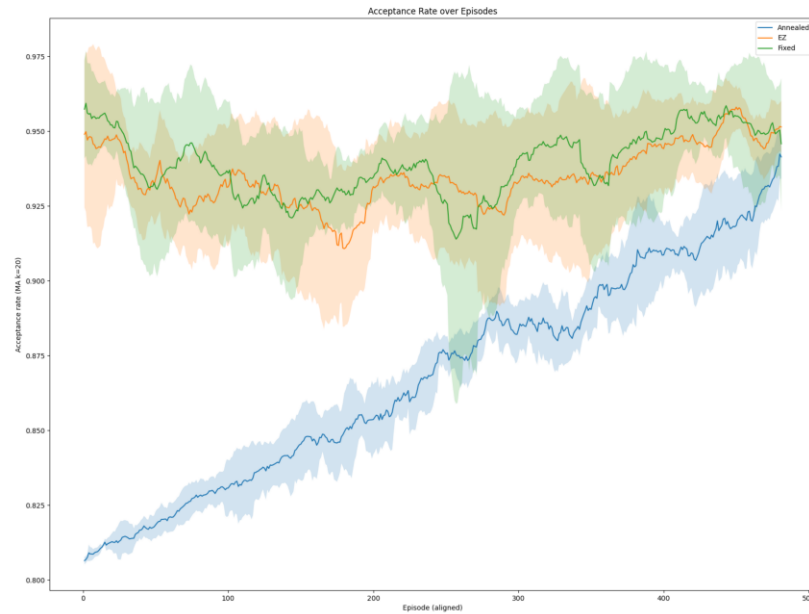
Erstellt am 11.08.2025 13:27

Verglichene Varianten:

- Annealed  $\epsilon$ -greedy (lineares Absenken der Exploration)
- EZ-greedy (konstante, aber strukturierte Exploration)
  - Fixed- $\epsilon$  (starrer  $\epsilon$ -Wert)

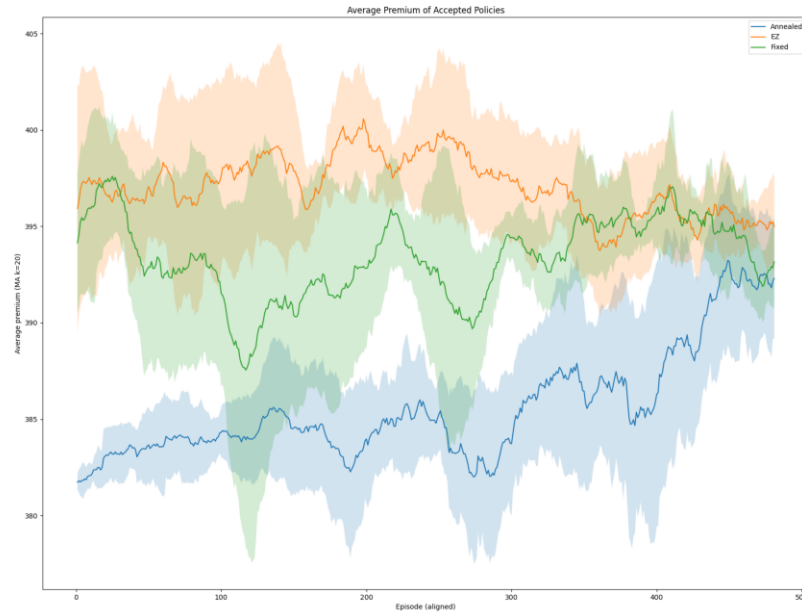
## Kurzfazit

- EZ-greedy erreicht die höchsten Erträge und die höchsten akzeptierten Prämien, zeigt jedoch größere Varianz.
- Annealed  $\epsilon$ -greedy lernt stetig und stabil: steigende Acceptance, sinkende Insolvenzrate in der Greedy-Evaluation.
- Fixed- $\epsilon$  weist einen klaren Stabilitäts-/Risikofehler auf (großer Kapital-Drawdown um Episoden 240–270).
- Für reinen Profit → EZ-greedy; für robustes, risikoärmeres Verhalten → Annealed; Fixed- $\epsilon$  derzeit nicht empfehlenswert.



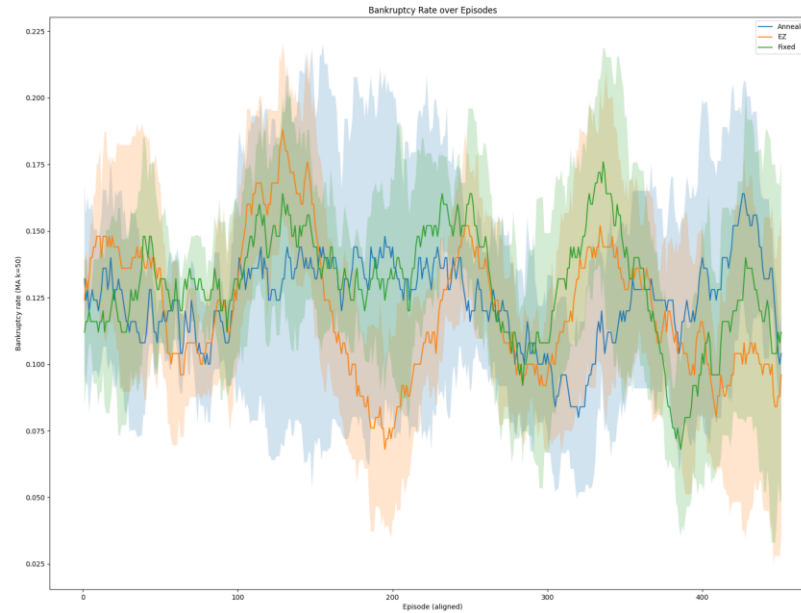
## Acceptance Rate über Episoden

- Annealed: klare Lernkurve von  $\sim 0.81 \rightarrow \sim 0.94-0.95$  (erst explorieren, dann ausnutzen).
- EZ & Fixed: starten hoch ( $\sim 0.95$ ), bleiben in  $0.92-0.96$  mit Wellen.
- Am Ende liegen alle nahe beieinander; Annealed holt sichtbar auf.



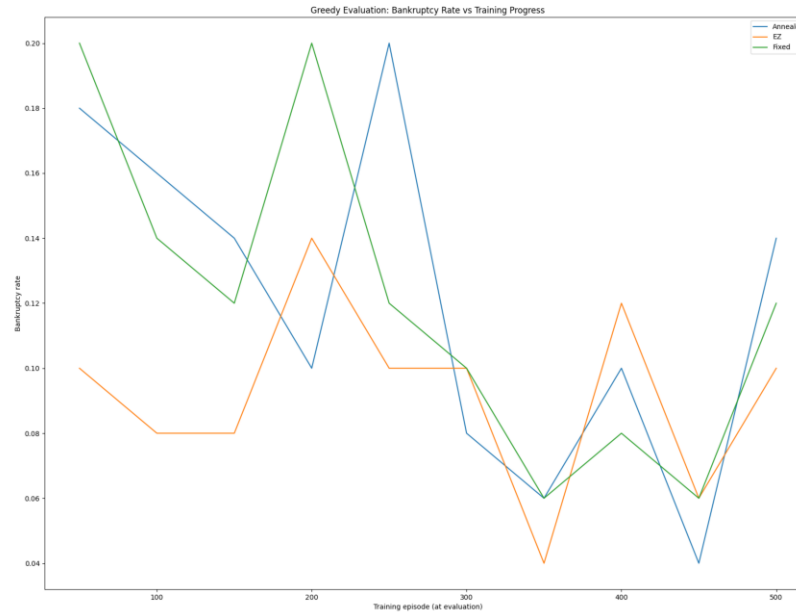
## Durchschnittsprämie akzeptierter Policen

- EZ: durchweg höchste durchschnittliche Prämien der akzeptierten Policen (~398-401 → ~395-398).
- Fixed: tiefer Dip (~Eps 150-200), später Erholung auf ~395-397.
- Annealed: startet am niedrigsten (~382) und steigt stetig Richtung ~392-395.



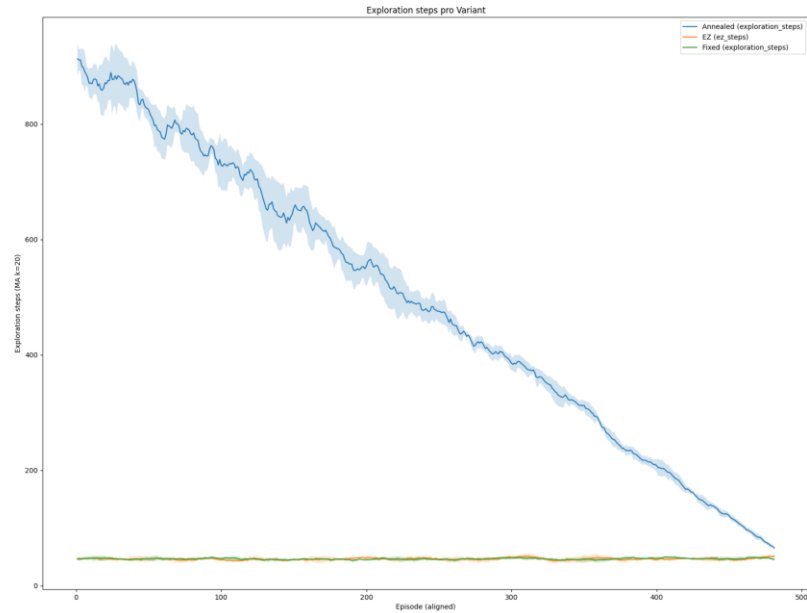
## Insolvenzrate - Training

- Alle Strategien oszillieren zwischen ca. 6–20% (Regime-/Schadens-Cluster sichtbar).
- Annealed zeigt über das Training die konsistenteste Abnahme der Insolvenzrate.



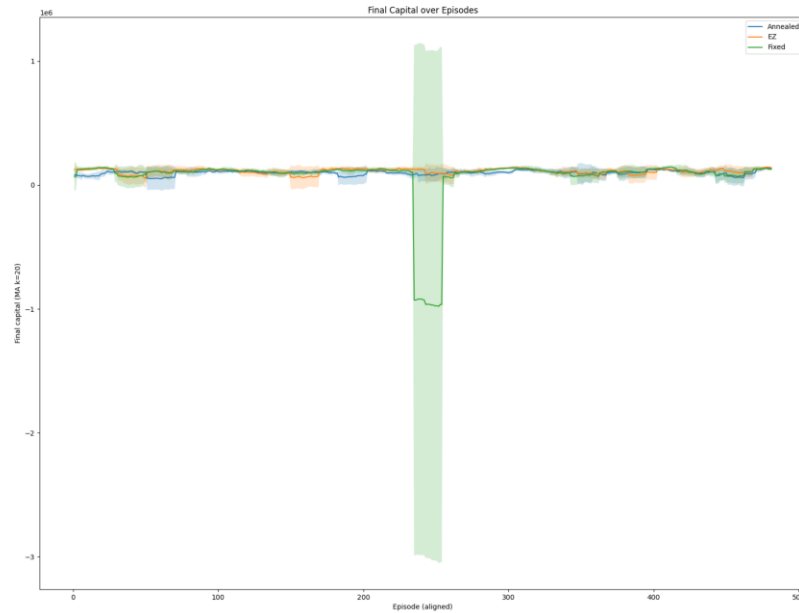
## Insolvenzrate - Greedy-Evaluation

- Greedy-Evaluation: ähnliche Größenordnung; Annealed erreicht Tiefstwerte (~4-6% um Ep. 400-450).
- Kein eindeutiger Gesamtsieger, aber Annealed zeigt die klarste Verbesserungstendenz.



## Explorationsschritte pro Variante

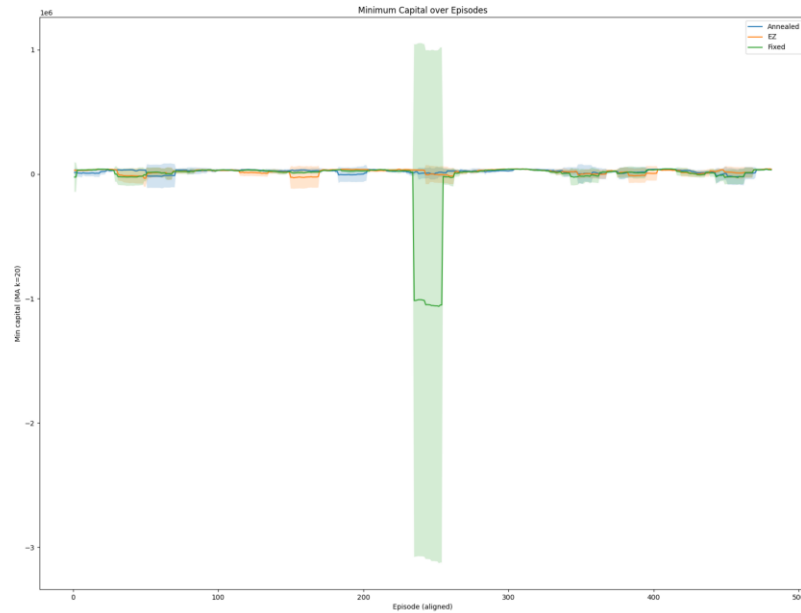
- Annealed: hohe Anfangsexploration, linear abnehmend bis  $\sim 0$  → erwartetes Verhalten.
- EZ & Fixed: fast konstante, geringe Exploration – EZ erzielt dennoch hohe Performance.



## Finales Kapital über Episoden

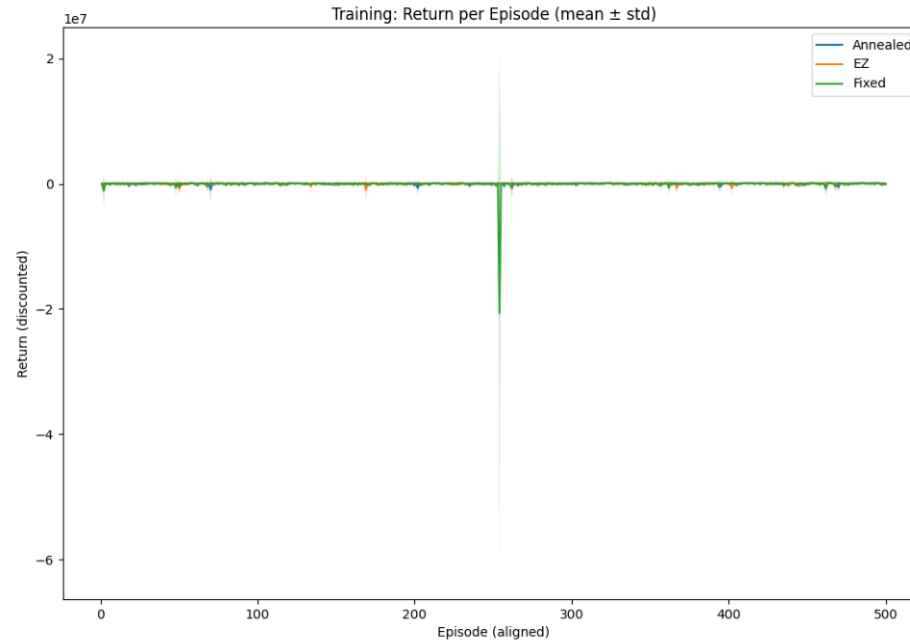
- Fixed- $\epsilon$ : massiver negativer Ausreißer um Episoden 240–270 (Kapital deutlich im Minus).
- Annealed & EZ: nahe Null mit üblichen Schwankungen – keine Katastrophen.
- Hinweis auf erhöhtes Risiko/Instabilität unter Fixed- $\epsilon$ .





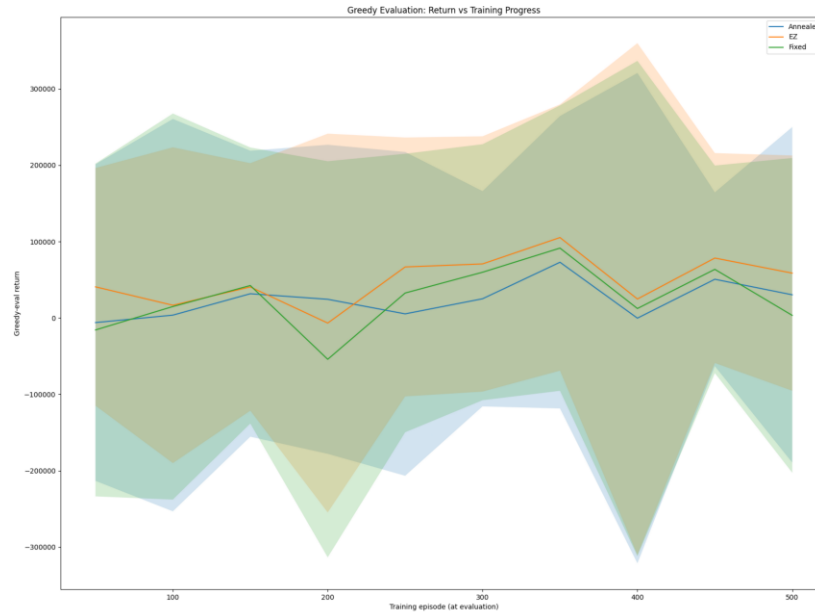
# Minimales Kapital über Episoden

- Bestätigt den Drawdown unter Fixed- $\epsilon$  im selben Zeitraum.
- Annealed & EZ halten Minimum-Kapital stabiler über das Training.



## Return pro Episode - Training

- Alle Varianten nahe 0 mit Spikes; negativer Extrem-Spike gehört zu Fixed (gleicher Zeitraum).
- Kein klarer Aufwärtstrend im Training-Return (hohe Volatilität der Episodenreturns).



## Return pro Episode - Greedy-Evaluation

- Greedy-Eval-Return steigt tendenziell; EZ erreicht wiederholt höchste Werte (bis ~300–350k), aber mit hoher Varianz.
- Annealed liegt knapp dahinter, jedoch stabiler; Fixed unterperformt und zeigt größere Risiken.

## Empfehlungen & Nächste Schritte

- Fixed- $\epsilon$  debuggen: Seeds/Logs der Episoden  $\sim 240$ – $270$  prüfen (Regime-Shift? Claim-Cluster?), Kapital-Update & Reward-Clipping verifizieren.
- Mehr Seeds ( $\geq 20$ ) für verlässlichere Konfidenzbänder; Konfidenzintervalle zusätzlich reporten.
- Fairer Vergleich: gleichen  $\epsilon$ -Floor setzen; bei EZ k-Parameter &  $\epsilon$  feinjustieren.
- Robustheit steigern: Double-DQN, Target-Soft-Updates, Gradient-Clipping, Prioritized Replay evaluieren.
- Risiko-Metriken ergänzen: CVaR/Expected Shortfall, Insolvenzhäufigkeit pro Regime, Profit per accepted policy.