

# **Hochschule Darmstadt**

– Fachbereich Informatik –

## **Konzeption und Evaluation eines Reinforcement-Learning-Systems für vollautomatische Konfliktlösung in Realzeitsimulationen**

Abschlussarbeit zur Erlangung des akademischen Grades  
Bachelor of Science (B.Sc.)

vorgelegt von

**Lars Przybylek**

Matrikelnummer: 1117822

Referent : Prof. Dr. Gunter Grieser  
Korreferent : Adriatik Gashi



## ERKLÄRUNG

---

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

*Darmstadt, 3. November 2025*

---

Lars Przybylek

# INHALTSVERZEICHNIS

---

## I Thesis

1	Einleitung	2
1.1	Motivation . . . . .	2
1.2	Ziel der Arbeit . . . . .	2
1.3	Gliederung . . . . .	3
1.4	Restliche Punkte für Vorhabensbeschreibung . . . . .	3
2	Flugsicherungsgrundlagen	5
2.1	Aufgabe der Flugsicherung . . . . .	5
2.2	Aufbau des Luftraums . . . . .	5
2.2.1	Sektorisierung . . . . .	5
2.2.2	Flugzeugübergabe . . . . .	5
2.3	Flugzeugkonflikte . . . . .	5
2.3.1	Definition . . . . .	5
2.3.2	Konfliktarten und deren Lösungsstrategien . . . . .	5
2.3.3	Kreuzende Routen . . . . .	5
2.3.4	Vereinen von Routen . . . . .	5
2.3.5	Identische Koordinationspunkte . . . . .	5
2.4	Simulator . . . . .	5
2.4.1	Nutzung . . . . .	5
2.4.2	Aufgaben der Adjazenzlotsen . . . . .	5
2.4.3	Pilot Commands . . . . .	5
3	Reinforcement Learning Grundlagen	6
3.1	Einführung in Reinforcement Learning . . . . .	6
3.2	RL-Algorithmen . . . . .	6
3.2.1	Q-Learning . . . . .	6
3.2.2	Policy Gradient Methoden . . . . .	6
3.3	Open Source RL Bibliotheken . . . . .	6
3.4	Probleme im Reinforcement Learning . . . . .	6
3.5	Trainingsstrategien . . . . .	7
4	Konzeption	8
4.1	Definition des Gyms . . . . .	8
4.1.1	Konfliktlösung mittel Headingänderung . . . . .	8
4.1.2	Konfliktlösung mittel Headingänderung und Speedänderung . . . . .	8
4.1.3	Konfliktlösung mittel Headingänderung, Speedänderung und Leveländerung . . . . .	8
4.2	Trainingsumgebung . . . . .	8
4.3	Evaluationsumgebung . . . . .	8
4.4	Simulatorintegration in das RL-Gym . . . . .	8
5	Evaluation	9
5.1	Quantitative Bewertung . . . . .	9

5.1.1	Bewertungskriterien . . . . .	9
5.1.2	Versuch 1 . . . . .	9
5.1.3	Versuch 2 . . . . .	9
5.1.4	Versuch 3 . . . . .	9
5.1.5	Vergleich . . . . .	9
5.2	Qualitative Bewertung . . . . .	9
5.2.1	Lotsenfeedback . . . . .	9
6	Zusammenfassung und Ausblick	10
6.1	Zukünftige Arbeiten . . . . .	10

## II Appendix

Literatur	12
-----------	----

## ABKÜRZUNGSVERZEICHNIS

---

DFS Deutsche Flugsicherung GmbH

RL Reinforcement Learning

Teil I

THESIS

## EINLEITUNG

---

### 1. Ausgangssituation

Diese Arbeit wurde in Rahmen des Dualen Studiums der Informatik an der Hochschule Darmstadt in Kooperation mit der Deutsche Flugsicherung GmbH (DFS) erstellt. Die Inhalte dieser Arbeit basieren zu Teil auf den Anforderungen, Erfahrungen und Erkenntnissen, die im Rahmen Praxisprojekts der dritten Praxisphase in dem Team TI/TD der DFS gesammelt wurden. Dieses Projekt baut auf einem vorherigen Bachelorarbeit[2] auf, die Reinforcementlearning als Lösungsansatz für die teilautomatisierte Flugverkehrssteuerung in der Adjazent Position untersucht hat.

#### 1.1 MOTIVATION

Das Team TI/TD der DFS entwickelt Simulatoren, die in der Aus- und Weiterbildung von Fluglotsen eingesetzt werden. Während Simulatorübungen arbeiten die Trainees mit den Adjazenlotsen zusammen. Adjazenlotsen müssen im Simulator eine Vielzahl von Aufgaben gleichzeitig bewältigen, dies führt zu einer hohen kognitiven Belastung.

Adjazenlotsen steuern nicht nur den Simulator, sondern übernehmen auch Kommunikationsaufgaben und die Kontrolle des Flugverkehrs in angrenzenden Sektoren. Diese Aufgaben erfordern ständige Aufmerksamkeit, schnelle Reaktionen und die Fähigkeit, mehrere Informationen gleichzeitig zu verarbeiten.

Um die kognitive Belastung zu reduzieren und die Effizienz der Adjazenlotsen zu steigern, untersucht diese Arbeit die Konzeption und Evaluation eines Reinforcement Learning (RL)-basierten Systems, das die Adjazenlotsen bei ihren Aufgaben unterstützt.

#### 1.2 ZIEL DER ARBEIT

### 2. Zielsetzung

Ziel dieser Arbeit ist die Konzeption und Evaluation eines RL-Systems, das die konfliktfreie Steuerung des Flugverkehrs in den umliegenden Sektoren der Trainees übernimmt. Das System soll Pilot Commands generieren, um Flugkonflikte zu vermeiden und den Verkehr effizient zu steuern. Dabei werden die Realisierbarkeit der Anweisungen, physikalische Limitierungen und die Anweisungsfrequenz berücksichtigt. Die Arbeit soll folgende Fragestellungen beantworten:



- Wie kann ein **RL**-System konzipiert werden, um die konfliktfreie Steuerung des Flugverkehrs in Realzeitsimulationen zu übernehmen?
- Welche Zustands- und Aktionsräume sowie Belohnungsfunktionen sind für die Konfliktlösung geeignet?
- Wie sollen Trainings- und Evaluationsumgebungen gestaltet sein, um eine effektive Ausbildung des **RL**-Agents zu gewährleisten?
- Welchen Einfluss haben verschiedene RL-Algorithmen auf die Leistung des Systems?
- Wie effektiv kann der **RL**-Agent Flugkonflikte lösen?
- Welche Anforderungen der Problemdomäne machen die Umsetzung eines **RL**-Systems herausfordernd?

Diese Arbeit verfolgt nicht das Ziel, ein operativ einsetzbares System zur vollständigen Ablösung des Adjazenzlotsen zu entwickeln. Vielmehr wird ein Proof of Concept vorgestellt, der auf einem stark vereinfachten Szenario basiert und die grundlegende Machbarkeit sowie zentrale Herausforderungen der Anwendung von **RL** im Flugverkehrsmanagement untersucht.

### 1.3 GLIEDERUNG

Zu Beginn der Arbeit wird in Kapitel 2 die Grundlagen der Problemdomäne Flugverkehrssteuerung und die Rolle der Adjazenzlotsen in der Ausbildung von Fluglotsen am Simulator erläutert. Anschließend werden in Kapitel 3 die theoretischen Grundlagen der Lösungsdomäne (**RL**) und relevanten Algorithmen vorgestellt. Im Kapitel 4 werden die Definition von Zustands- und Aktionsräumen sowie der Belohnungsfunktion behandelt, sowie deren Zusammenhang mit verschiedenen Konfliktlösungsmaßnahmen erläutert. Kapitel 5 widmet sich der Evaluation des **RL**-Systems, in dem die durchgeführten Experimente und deren Ergebnisse präsentiert und analysiert werden. Abschließend werden in Kapitel 6 die Ergebnisse der Arbeit zusammengefasst und ein Ausblick auf mögliche zukünftige Arbeiten gegeben.

### 1.4 RESTLICHE PUNKTE FÜR VORHABENS BESCHREIBUNG

## 3. Vorgehen

Das Vorgehen zur Umsetzung des **RL**-Systems gliedert sich in mehrere Schritte:

- Optimierung der Rewardfunktion, Action und Observation spaces mit dem Ziel, die Konfliktlösungsfähigkeit mit wenigen Aktionen zu ermöglichen. (3 Wochen)
- Unterstützung weiterer Konfliktlösungsmaßnahmen wie Speed- und Leveländerungen. (1 Woche)

- Vergleich verschiedener [RL](#)-Algorithmen hinsichtlich ihrer Eignung. (1 Woche)
- Validierung der Agenten mittels Domänen spezialisten. (0,5 Woche)

## FLUGSICHERUNGSGRUNDLAGEN

---

### 2.1 AUFGABE DER FLUGSICHERUNG

### 2.2 AUFBAU DES LUFTRAUMS

#### 2.2.1 *Sektorisierung*

#### 2.2.2 *Flugzeugübergabe*

### 2.3 FLUGZEUGKONFLIKTE

#### 2.3.1 *Definition*

#### 2.3.2 *Konfliktarten und deren Lösungsstrategien*

Ziel Abstand von 6 - 9 NM horizontal

#### 2.3.3 *Kreuzende Routen*

60NM Regel: 60 NM vor Kollisionpunkt 1 grad heading Änderung resultiert in etwa 5 NM Abstand am Kollisionpunkt

#### 2.3.4 *Vereinen von Routen*

dürfen sich nicht annähern und mindestens 10 NM Abstand halten

#### 2.3.5 *Identische Koordinationspunkte*

Direct Routing (Koordinationspunkt auslassen)

### 2.4 SIMULATOR

#### 2.4.1 *Nutzung*

#### 2.4.2 *Aufgaben der Adjazenlotsen*

#### 2.4.3 *Pilot Commands*

## REINFORCEMENT LEARNING GRUNDLAGEN

---

### 3.1 EINFÜHRUNG IN REINFORCEMENT LEARNING

- Einordnung im Maschinelles Lernen
- Ziel des Reinforcement Learnings
- Markov-Entscheidungsprozesse
- Agenten und Umgebung
- Belohnungssignal

### 3.2 RL-ALGORITHMEN

#### 3.2.1 *Q-Learning*

#### 3.2.2 *Policy Gradient Methoden*

[7]

### 3.3 OPEN SOURCE RL BIBLIOTHEKEN

- Open AI Gym
- Bluesky Gym
- Stable Baselines3
- RL Baselines3 Zoo
- Optuna

[1] [4] [6]

### 3.4 PROBLEME IM REINFORCEMENT LEARNING

[5]

- Exploration vs. Exploitation
- Sparse Rewards
- Sample Efficiency
- Überanpassung (Overfitting)
- Belohnungsgestaltung (Reward Shaping)

### 3.5 TRAININGSSTRATEGIEN

- Curriculum Learning
- Domain Randomization
- Self-Play

## KONZEPTION

---

### 4.1 DEFINITION DES GYMS

#### 4.1.1 *Konfliktlösung mittel Headingänderung*

##### 4.1.1.1 *Zustandsraum*

##### 4.1.1.2 *Aktionsraum*

##### 4.1.1.3 *Belohnungsfunktion*

#### 4.1.2 *Konfliktlösung mittel Headingänderung und Speedänderung*

##### 4.1.2.1 *Zustandsraum*

##### 4.1.2.2 *Aktionsraum*

##### 4.1.2.3 *Belohnungsfunktion*

#### 4.1.3 *Konfliktlösung mittel Headingänderung, Speedänderung und Leveländerung*

##### 4.1.3.1 *Zustandsraum*

##### 4.1.3.2 *Aktionsraum*

##### 4.1.3.3 *Belohnungsfunktion*

### 4.2 TRAININGSUMGEBUNG

### 4.3 EVALUATIONSUMGEBUNG

### 4.4 SIMULATORINTEGRATION IN DAS RL-GYM

## EVALUATION

---

### 5.1 QUANTITATIVE BEWERTUNG

#### 5.1.1 *Bewertungskriterien*

#### 5.1.2 *Versuch 1*

#### 5.1.3 *Versuch 2*

#### 5.1.4 *Versuch 3*

#### 5.1.5 *Vergleich*

### 5.2 QUALITATIVE BEWERTUNG

#### 5.2.1 *Lotsenfeedback*

## ZUSAMMENFASSUNG UND AUSBLICK

---

### 6.1 ZUKÜNFTIGE ARBEITEN

- Erweiterung des RL-Systems um zusätzliche Konfliktlösungsstrategien.
- Imitationslernen basierend auf menschlichen Lotsenverhalten. [3]
- Integration trainierter Policies mittels ONNX in der Java Umgebung der Simulatoren.



Teil II

APPENDIX

## LITERATUR

---

- [1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta und Masanori Koyama. "Optuna: A Next-Generation Hyperparameter Optimization Framework". In: *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, S. 2623–2631.
- [2] Dominik Diez. "Reinforcement Learning als Lösungsansatz für die Automatisierung von Aufgaben der Adjacent Position in Flugsicherungs-Realzeitsimulationen". 2025.
- [3] Adam Gleave, Mohammad Taufeque, Juan Rocamonde, Erik Jenner, Steven H. Wang, Sam Toyer, Maximilian Ernestus, Nora Belrose, Scott Emmons und Stuart Russell. *imitation: Clean Imitation Learning Implementations*. arXiv:2211.11972v1 [cs.LG]. 2022. arXiv: [2211.11972](https://arxiv.org/abs/2211.11972) [cs.LG]. URL: <https://arxiv.org/abs/2211.11972>.
- [4] DJ Groot, G Leto, A Vlaskin, A Moec und J Ellerbroek. *BlueSky-Gym: Reinforcement Learning Environments for Air Traffic Applications*. 2024.
- [5] Alex Irpan. *Deep Reinforcement Learning Doesn't Work Yet*. <https://www.alexirpan.com/2018/02/14/rl-hard.html>. 2018.
- [6] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus und Noah Dormann. "Stable-Baselines3: Reliable Reinforcement Learning Implementations". In: *Journal of Machine Learning Research* 22.268 (2021), S. 1–8. URL: <http://jmlr.org/papers/v22/20-1364.html>.
- [7] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford und Oleg Klimov. "Proximal Policy Optimization Algorithms". In: *CoRR* abs/1707.06347 (2017). arXiv: [1707.06347](https://arxiv.org/abs/1707.06347). URL: [http://arxiv.org/abs/1707.06347](https://arxiv.org/abs/1707.06347).