

# Bálint Gyevnár

PhD Candidate, School of Informatics, University of Edinburgh

[balint.gyevnar@ed.ac.uk](mailto:balint.gyevnar@ed.ac.uk) | [gbalint.me](https://gbalint.me) | [github.com/gyevnarb](https://github.com/gyevnarb)

## EDUCATION

---

### University of Edinburgh

Sep. 2021 – May 2025 (est.)

*PhD in Natural Language Processing with Integrated Studies*

*Edinburgh, UK*

*Supervisors: Stefano V. Albrecht, Shay B. Cohen, and Christopher G. Lucas*

### University of Edinburgh

Sep. 2016 – May 2021

*Integrated Master of Informatics*

*Edinburgh, UK*

*Supervisor: Maria Wolters*

### Nanyang Technological University

Aug. 2018 – May 2019

*Exchange Student in Computer Science*

*Singapore*

## EXPERIENCE

---

### Research Assistant

Jul. 2023 – present

*University of Edinburgh*

*Edinburgh, UK*

- Researching the intersection of AI safety and AI ethics to build bridges among research problems.
- Large scale quantitative literature analysis with unsupervised natural language processing tools.
- Curation, topic coding, and qualitative analysis of large corpora of papers.
- Research assistant to Dr. Atoosa Kasirzadeh.

### Sports Club Executive Member

Sep. 2022 – Jun. 2025 (est.)

*Edinburgh University Volleyball Club*

*Edinburgh, UK*

- (2024-25; Secretary) Responsible for public outreach to and networking with alumni members and organizing a two-day event series.
- (2023-24; VP) Large-scale event organization, public speaking, timetabling, and human resource management of 8 teams, 10 coaches, and 220+ members.
- (2022-23; Treasurer) Managing a cash flow of approximately £70k, setting up an annual budget, and managing thousands of transactions.

### Teaching Assistant

Sep. 2020 – present

*University of Edinburgh*

*Edinburgh, UK*

- Assistant supervisor for two master's students.
- Online tutorial sessions to ~12 students for introductory machine learning course.
- Coursework and exam marker for courses in the School of Informatics, including Doing Research in NLP, Reinforcement Learning, Computer Systems, and Machine Learning.

### Research Assistant

May 2020 – Oct. 2020

*Five AI Ltd.*

*Edinburgh, UK*

- Developed and evaluated IGP2, a goal-based interpretable prediction and planning system for autonomous vehicles with intuitive explanations.
- Scenario-based and open-world testing and evaluation of IGP2.
- Publication at International Conference on Robotics and Automation (ICRA), 2021.

### Conference

- [under review] **B. Gyevnar**, S. Droop, T. Quillien, S.B. Cohen, N.R. Bramley, C.G. Lucas, S.V. Albrecht. “People Attribute Purpose to Autonomous Vehicles When Explaining Their Behavior”; ACM Conference on Human Factors in Computing Systems (CHI), 2025.
- **B. Gyevnar**, C. Wang, S.B. Cohen, C.G. Lucas, S.V. Albrecht. “Causal Explanations for Sequential Decision-Making in Multi-Agent Systems”, *Association for the Advancement of Artificial Intelligence (AAMAS-24)*, 2024.
- **B. Gyevnar**, N. Ferguson, B. Schafer. “Bridging the Transparency Gap: What Can Explainable AI Learn From the AI Act?”, *European Conference on Artificial Intelligence (ECAI)*, 2023.
- S.V. Albrecht, C. Brewitt, J. Wilhelm, **B. Gyevnar**, F. Eiras, M. Dobre, S. Ramamoorthy. “Interpretable Goal-based Prediction and Planning for Autonomous Driving”, *International Conference on Robotics and Automation (ICRA)*, 2021;
- C. Brewitt, **B. Gyevnar**, S. Garcin., S.V. Albrecht. “GRIT: Fast, Interpretable, and Verifiable Goal Recognition with Learned Decision Trees for Autonomous Driving”, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

### Journal

- A. Kuznietsov\*, **B. Gyevnar\***, C. Wang, S. Peters, S.V. Albrecht. “Explainable AI for Safe and Trustworthy Autonomous Driving: A Systematic Review”, *IEEE Transactions on Intelligent Transportation Systems*, 25 (12), 19342-19364, 2024. [\* Equal contribution]
- **B. Gyevnar**, G. Dagan, C. Haley, S. Guo, F. Mollica. “Communicative Efficiency or Iconic Learning: Do communicative and acquisition pressures interact to shape colour-naming systems?”, *Entropy*, 24(11), 1542, 2022.

### Other

- **B. Gyevnar**, C. Brewitt, S. Garcin, M. Tamborski, and S.V. Albrecht. Code Repository for Interpretable Goal-based Prediction and Planning (IGP2); *Github*, 2022.
- **B. Gyevnar**, M. Tamborski, C. Wang, C.G. Lucas, S.B. Cohen, S.V. Albrecht. “A Human-Centric Method for Generating Causal Explanations in Natural Language for Autonomous Vehicle Motion Planning”, Runner-up for best paper, *IJCAI 2022 Workshop on Artificial Intelligence for Autonomous Driving*, 2022.

### Awards

- AI100 Early Career Essay Competition Featured Essay, “Love, Sex, and AI”, *Standing Committee of the One Hundred Year Study on Artificial Intelligence (AI100)*, Stanford University, 2023.
- Trustworthy Autonomous Systems Early Career Researcher Awards for £4000, Knowledge Transfer Track, *UK Research & Innovation*, 2023.
- Shape the Future of ITS Competition for \$1000, “Cars that Explain: Building Trust in Autonomous Vehicles through Explanations and Conversations”, *IEEE Intelligent Transportation Systems Society*, 2022.