Bálint Gyevnár

School of Informatics, University of Edinburgh
balint.gyevnar@ed.ac.uk | gbalint.me | github.com/gyevnarb

Education

University of Edinburgh — Edinburgh, UK
PhD in Natural Language Processing with Integrated Studies — *Sep. 2021 – May 2025 (est.)*
Supervisors: Stefano V. Albrecht, Shay B. Cohen, and Christopher G. Lucas

University of Edinburgh — Edinburgh, UK
Integrated Master of Informatics — *Sep. 2016 – May 2021*
Supervisor: Maria Wolters

Nanyang Technological University — Singapore
Exchange Student in Computer Science — *Aug. 2018 – May 2019*

Experience

Teaching Assistant — Sep. 2020 – present
University of Edinburgh — *Edinburgh, UK*

- Delivering and moderating online tutorial sessions of ∼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.

Assistant Supervisor for Master's Students — Sep. 2022 – present
University of Edinburgh — *Edinburgh, UK*

- Assistant supervisor for two master's student working on robust planning for autonomous vehicles.

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.

Research Output

Awards

- UKRI Trustworthy Autonomous Systems Early Career Researcher Awards, Knowledge Transfer Track, £4000, 2023
- **B. Gyevnar**. Cars that Explain: Building Trust in Autonomous Vehicles through Explanations and Conversations; "Shape the Future of ITS" Competition, $1000, *IEEE Intelligent Transportation Systems Society*, 2022

Conference

- 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

- **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. Brewit, S. Garcin, M. Tamborski, and S.V. Albrecht. Interpretable Goal-based Prediction and Planning (IGP2) Code Repository; *Github*, 2022

Workshop

- **B. Gyevnar**, N. Ferguson. Aligning Explainable AI and the Law: The European Perspective, *AAMAS 2023 Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems (EXTRAAMAS)*, 2023
- **B. Gyevnar**, C. Wang, C.G. Lucas, S.B. Cohen, S.V. Albrecht. Causal Explanations for Stochastic Sequential Multi-Agent Decision-Making, *IJCAI 2023 Workshop on Explainable Artificial Intelligence*, 2023
- **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
- C. Brewitt, S.V. Albrecht, J. Wilhelm, **B. Gyevnar**, F. Eiras, M. Dobre, S. Ramamoorthy. Autonomous Driving with Interpretable Goal Recognition and Monte Carlo Tree Search, *RSS 2020 Workshop on Interaction and Decision-Making in Autonomous-Driving*, 2020

Projects

Explainable Autonomous Vehicle Intelligence Sep. 2021 – Present

- Cross-disciplinary collaboration towards trustworthy autonomous vehicles via explanations and conversations.
- Introduced CEMA: a novel system to generate causal explanations for multi-agent decision-making.
- Integrating CEMA with dialogue systems to deliver relevant natural language explanations.
- Evaluating methods with human participants to measure the effect of explanations on trust and understanding.
- Leading the work with a team of 5+ people within the Autonomous Agents group.

Aligning Explainable AI and the Law Nov. 2021 – Present

- Surveyed the explainability and broader transparency requirements of upcoming legislative frameworks for AI.
- Reviewed the legal considerations behind modern XAI techniques and paradigms.
- Identified shared concepts and notional discrepancies between XAI and the Law.
- Paper in review at ECAI 2023.

Lead Developer and Maintainer of Interpretable Goal-Based Prediction and Planning for Autonomous Vehicles May 2021 – Present

- Lead developer and maintainer of open-source Python package for AV prediction and planning.
- Author of comprehensive documentation and users' manual.
- Python package on GitHub with 62 stars and 17 forks.

Acquisition & Communication of Colour Naming Systems Sep. 2021 – Nov. 2022

- Understanding the effects of communicative efficiency and acquisition on the patterns of human colour naming systems via computational information-theoretic measures.
- Simulating acquisition patterns of colour term learning using self-organising maps and the World Colour Survey.
- Lead author of a team of five with a published journal paper in *Entropy*.

Other Skills

Programming Languages: Fluent in **Python**. Experienced with **C#**. Some **C++**, Java, Bash, R, etc.
Natural Languages: English (fluent), German (advanced), Japanese (intermediate), Hungarian (fluent).
Software: CARLA, RoadRunner, Shapely, PyTorch, Pandas, Matplotlib, Django, etc.