Julian Skirzyński

Curriculum Vitae — February 2026

jskirzynski@ucsd.edu
www.jskirzynski.com

Education

University of California, San Diego 2022 – Present
Ph.D. Candidate in Computer Science & Engineering
Thesis: Foundations of Human AI Interaction
Advisor: Berk Ustun

McGill University 2017 – 2020
M.S. in Computer Science
Thesis: Language-Conditional Imitation Learning
Advisor: David Meger

University of Warsaw 2012 – 2018
M.S. in Cognitive Science
B.S. in Mathematics, Cognitive Science
Advisors: Andrzej Skowron; Piotr Wasilewski

Academic Positions

Max Planck Institute for Intelligent Systems, Germany 2019 – 2023
Research Scientist
Projects: Interpretable RL Policies, Improving Human Planning, Discovering Human Planning Strategies
Advisor: Falk Lieder

Research Interests

Areas: Machine Learning, Cognitive Science, Human-Computer Interaction
Topics: Decision-Making, Interpretability, Explainability, Reinforcement Learning, Experimental Design
Applications: Social Sciences, Medicine, Consumer Finance, Criminal Justice

Awards & Honors

Pierre Arbour Foundation Scholarship 2018 – 2019
McGill University Graduate Excellence Award 2018
McGill - University of Warsaw Exchange Scholarship 2015
University of Warsaw Academic Excellence Scholarship 2014 – 2017

Preprints

$\alpha - \beta$ Alphabetical

1. Measuring What Matters: Synthetic Benchmarks for Concept Bottleneck Models
   $\alpha - \beta$ Harry Cheon, Shreyas Kadekodi, **Julian Skirzyński**, Meredith Stewart, Berk Ustun
   In Submission, 2025

2. On the Value of Interpretability in Human Decision-Making
   Julian Skirzyński, Elena Glassman, Berk Ustun
   In Preparation, 2025

Papers

*Equal contribution

Google Scholar

3. Quantifying Cognitive Bias Induction in LLM-Generated Content
   Abeer Alessa, Akshaya Lakshminarasimhan, Param Somane, **Julian Skirzyński**, Julian McAuley, Jessica Echterhoff
   International Joint Conference on Natural Language Processing & Asia-Pacific Chapter of the Association for Computational Linguistics, 2025

4. Discrimination Exposed? On the Reliability of Explanations for Discrimination Detection
   Julian Skirzyński, David Danks, Berk Ustun
   ACM Conference on Fairness, Accountability, and Transparency, 2025

5. Automatic Discovery and Description of Human Planning Strategies
   Julian Skirzyński, Yash Raj Jain, Falk Lieder
   Behavior Research Methods, 2023

6. Boosting Human Decision-making with AI-Generated Decision Aids
Frederic Becker*, **Julian Skirzyński***, Bas van Opheusden, Falk Lieder
Computational Brain & Behavior, 2022

7. Automatic Discovery of Interpretable Planning Strategies
Julian Skirzyński, Frederic Becker, Falk Lieder
Machine Learning, 2021

8. Object [Re] Cognition with Similarity
Łukasz Sosnowski, **Julian Skirzyński**
International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2018

9. A Framework for Analysis of Granular Neural Networks
Julian Skirzyński
International Joint Conference on Rough Sets, 2017

Refereed Workshop Papers

9. On Interpretability and Overreliance
Julian Skirzyński, Elena Glassman, Berk Ustun
Interpretable AI: Past, Present and Future, NeurIPS Workshop, 2024

10. Language-Conditional Imitation Learning
Julian Skirzyński, Bobak Baghi, David Meger
Visually Grounded Interaction and Language, NAACL Workshop, 2021

Teaching

UCSD Computer Science & Engineering Department 2026
CSE101 – Design & Analysis of Algorithms
Teaching Assistant
Led discussion sessions for 300+ undergraduate students on the basics of theoretical computer science (graph theory, runtime, NP-completeness, etc.). Held weekly office hours and corrected homework.

UCSD Halıcıoğlu Data Science Institute 2023
DSC291 – Interpretability & Explainability in Machine Learning
Guest Lecturer & Teaching Assistant
Co-designed curriculum and held weekly office hours for serving 20+ PhD/MS students. Delivered guest lectures on ML interpretability methods and cognitive biases in AI-assisted decision-making. Completed teaching development workshop on graduate-level instruction.

Software

GitHub
Strategy Extraction from RL Policies – Algorithm to extract interpretable decision trees from RL policies
Human Planning Strategy Analysis – Framework for identifying strategies used in human planning tasks

Selected Industry Positions

Educational Entertainment One, Warsaw, Poland 2021 – 2024
Lead Technical Architect
Designed algorithms (AI, NLP) and supported the production process for a story-driven mobile game for learning English.

Academic Service

Journal Reviewing

Machine Learning 2022

Conference Program Committee

NeurIPS – Conference on Neural Information Processing Systems 2023 – Present
ICML – International Conference on Machine Learning 2025 – Present
ICLR – International Conference on Learning Representations 2024 – Present

AAAI Conference on Artificial Intelligence	2025 – Present
FAccT – ACM Conference on Fairness, Accountability and Transparency	2022 – Present
ICML Workshop RL4RealLife – International Conference on Machine Learning	2021
IPMU – Information Processing and Management of Uncertainty in Knowledge-Based Systems	2018

Personal

Language Skills : English, Polish, German (Conversational)
Software Skills : Python, R, C++, Flask, AWS, PyTorch, CPLEX, JavaScript, Jira
Interests : Soccer, Groundhopping, Traveling, Fantasy Literature, Record Collecting
Other : Peer tutoring, Co-author of "Triozy polskie", a textbook for learning Polish by foreigners