

Julian Skirzyński

CURRICULUM VITAE — JULY 2025


jskirzynski@ucsd.edu

www.jskirzynski.com

EDUCATION	University of California, San Diego <i>Ph.D. Candidate in Computer Science & Engineering</i> Thesis: Foundations of Human AI Interaction Advisor: Berk Ustun	2022 – PRESENT
	McGill University <i>M.S. in Computer Science</i> Thesis: Language-Conditional Imitation Learning Advisor: David Meger	2017 – 2020
	University of Warsaw <i>M.S. in Cognitive Science</i> <i>B.S. in Mathematics, Cognitive Science</i> Advisors: Andrzej Skowron; Piotr Wasilewski	2012 – 2018
ACADEMIC POSITIONS	Max Planck Institute for Intelligent Systems, Germany <i>Research Scientist</i> Projects: Interpretable RL Policies, Improving Human Planning, Discovering Human Planning Strategies Advisor: Falk Lieder	2019 – 2023
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 McGill University Graduate Excellence Award McGill - University of Warsaw Exchange Scholarship University of Warsaw Academic Excellence Scholarship	2018 – 2019 2018 2015 2014 – 2017
PREPRINTS	<ol style="list-style-type: none">1. How Much Content Do LLMs Generate That Induces Cognitive Bias in Users? Abeer Alessa, Akshaya Lakshminarasimhan, Param Somane, Julian Skirzyński, Julian McAuley, Jessica Echterhoff <i>In Submission</i>, 20252. On the Value of Interpretability in Human Decision-Making Julian Skirzyński, Elena Glassman, Berk Ustun <i>In Submission</i>, 2025	
PAPERS	<ol style="list-style-type: none">3. Discrimination Exposed? On the Reliability of Explanations for Discrimination Detection Julian Skirzyński, David Danks, Berk Ustun <i>ACM Conference on Fairness, Accountability, and Transparency</i>, 20254. Automatic Discovery and Description of Human Planning Strategies Julian Skirzyński, Yash Raj Jain, Falk Lieder <i>Behavior Research Methods</i>, 20235. Boosting Human Decision-making with AI-Generated Decision Aids Frederic Becker*, Julian Skirzyński*, Bas van Opheusden, Falk Lieder <i>Computational Brain & Behavior</i>, 2022	

*EQUAL CONTRIBUTION



	6.	Automatic Discovery of Interpretable Planning Strategies Julian Skirzyński, Frederic Becker, Falk Lieder <i>Machine Learning</i> , 2021	
	7.	Object [Re] Cognition with Similarity Łukasz Sosnowski, Julian Skirzyński <i>International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems</i> , 2018	
	8.	A Framework for Analysis of Granular Neural Networks Julian Skirzyński <i>International Joint Conference on Rough Sets</i> , 2017	
REFEREED WORKSHOP PAPERS	9.	On Interpretability and Overreliance Julian Skirzyński, Elena Glassman, Berk Ustun <i>Interpretable AI: Past, Present and Future, NeurIPS Workshop</i> , 2024	
	10.	Language-Conditional Imitation Learning Julian Skirzyński, Bobak Baghi, David Meger <i>Visually Grounded Interaction and Language, NAACL Workshop</i> , 2021	
TEACHING		UCSD Halicioğlu Data Science Institute DSC291 – Interpretability & Explainability in Machine Learning <i>Guest Lecturer & Teaching Assistant</i> 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.	2023
SOFTWARE		Strategy Extraction from RL Policies – Algorithm to extract interpretable decision trees from RL policies  GitHub Human Planning Strategy Analysis – Framework for identifying strategies used in human planning tasks	
SELECTED INDUSTRY POSITIONS		Educational Entertainment One, Warsaw, Poland <i>Lead Technical Architect</i> Designed algorithms (AI, NLP) and supported the production process for a story-driven mobile game for learning English.	2021 – 2024
ACADEMIC SERVICE		JOURNAL REVIEWING Machine Learning	2022
		CONFERENCE PROGRAM COMMITTEE NeurIPS – Conference on Neural Information Processing Systems ICML – International Conference on Machine Learning ICLR – International Conference on Learning Representations FAccT – ACM Conference on Fairness, Accountability and Transparency ICML Workshop RL4RealLife – International Conference on Machine Learning IPMU – Information Processing and Management of Uncertainty in Knowledge-Based Systems	2023 – PRESENT 2025 – PRESENT 2024 – PRESENT 2022 – PRESENT 2021 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	