

# Julian Skirzyński

CURRICULUM VITAE — DECEMBER 2025

[jskirzynski@ucsd.edu](mailto:jskirzynski@ucsd.edu)

[www.jskirzynski.com](http://www.jskirzynski.com)

EDUCATION	<b>University of California, San Diego</b> <i>Ph.D. Candidate in Computer Science &amp; Engineering</i> Thesis: Foundations of Human AI Interaction Advisor: Berk Ustun	2022 – PRESENT
	<b>McGill University</b> <i>M.S. in Computer Science</i> Thesis: <a href="#">Language-Conditional Imitation Learning</a> Advisor: David Meger	2017 – 2020
	<b>University of Warsaw</b> <i>M.S. in Cognitive Science</i> <i>B.S. in Mathematics, Cognitive Science</i> Advisors: Andrzej Skowron; Piotr Wasilewski	2012 – 2018
ACADEMIC POSITIONS	<b>Max Planck Institute for Intelligent Systems, Germany</b> <i>Research Scientist</i> Projects: Interpretable RL Policies, Improving Human Planning, Discovering Human Planning Strategies Advisor: Falk Lieder	2019 – 2023
RESEARCH INTERESTS	<b>Areas:</b> Machine Learning, Cognitive Science, Human-Computer Interaction <b>Topics:</b> Decision-Making, Interpretability, Explainability, Reinforcement Learning, Experimental Design <b>Applications:</b> 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	<b>1. <a href="#">Measuring What Matters: Synthetic Benchmarks for Concept Bottleneck Models</a></b> $\alpha - \beta$ Harry Cheon, Shreyas Kadekodi, <b>Julian Skirzyński</b> , Meredith Stewart, Berk Ustun <i>In Submission</i> , 2025	
	<b>2. <a href="#">On the Value of Interpretability in Human Decision-Making</a></b> <b>Julian Skirzyński</b> , Elena Glassman, Berk Ustun <i>In Preparation</i> , 2025	
PAPERS	<b>3. <a href="#">Quantifying Cognitive Bias Induction in LLM-Generated Content</a></b> Abeer Alessa, Akshaya Lakshminarasimhan, Param Somane, <b>Julian Skirzyński</b> , Julian McAuley, Jessica Echterhoff <i>International Joint Conference on Natural Language Processing &amp; Asia-Pacific Chapter of the Association for Computational Linguistics</i> , 2025	
	<b>4. <a href="#">Discrimination Exposed? On the Reliability of Explanations for Discrimination Detection</a></b> <b>Julian Skirzyński</b> , David Danks, Berk Ustun <i>ACM Conference on Fairness, Accountability, and Transparency</i> , 2025	
	<b>5. <a href="#">Automatic Discovery and Description of Human Planning Strategies</a></b> <b>Julian Skirzyński</b> , Yash Raj Jain, Falk Lieder <i>Behavior Research Methods</i> , 2023	

$\alpha - \beta$   
ALPHABETICAL

\*EQUAL  
CONTRIBUTION



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 Halicioğlu 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