Jordi Vallverdú

Causality for Artificial Intelligence

From a Philosophical Perspective

Springer

Causality for Artificial Intelligence

Jordi Vallverdú

Causality for Artificial Intelligence

From a Philosophical Perspective

Springer

Jordi Vallverdú
Philosophy Department
ICREA/Universitat Autònoma de Barcelona
Barcelona, Catalonia, Spain

ISBN 978-981-97-3186-2 ISBN 978-981-97-3187-9 (eBook)
https://doi.org/10.1007/978-981-97-3187-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

灯りし家があれば小川の流れいる
Tomorishi ie ga areba ogawa no nagare iru
If there is a lighted house, a stream will flow.

Gekkōshi

To Heura and Isard, my lovely kids

Preface

Causality, the inherent linkage between events, constitutes an intrinsic cognitive imperative, not only refined for sensory comprehension but integral to the profound construction of meaning within our cognitive apparatus. The relentless human pursuit of understanding the underpinnings of occurrences is heightened by the evolving landscape of Artificial Intelligence tools in contemporary research. This technological paradigm shift has not only extended the purview of our inquiries but has also significantly elevated the refinement and enhancement of our investigatory methodologies. The ensuing challenge lies in the astute assimilation of our conceptual modalities for information processing, strategically orchestrated to harness the precision and analytical prowess intrinsic to computational tools.

This book marks the inaugural volume in a series dedicated to exploring the philosophical underpinnings and contributions to artificial intelligence, machine learning, and, more specifically, deep learning techniques. Throughout the writing of this book, the dynamic landscape in the field has been characterized by the emergence of Generative AI and Large Language Models (LLM). These transformative developments have captivated researchers, sparking a revolutionary shift in how humans interact with AI tools and approach causal analysis within these new systems. Acknowledging the rapid evolution of technology, particularly in Generative AI and LLM, this series aims to delve deeper into the philosophical implications, challenges, and ethical considerations arising from the integration of these innovative technologies into the intricate tapestry of AI and machine learning. One pivotal aspect under my scrutiny has been the profound impact of counterfactual reasoning in addressing complexities at the intersection of causality and machine learning. This book endeavors to shed light on the open debates surrounding these critical issues, contributing meaningfully to the scholarly discourse on the multifaceted relationship between causality and artificial intelligence.

The main concepts and fields encapsulating the essence of this work include several topics: epistemology, deep learning, science, statistics, causality, cognition, and heuristics. Through collaborative efforts and the support of an ICREA2019 award, this book aims to contribute significantly to advancing knowledge at the intersection of academia and cutting-edge research.

In this intellectual milieu, it is imperative to recognize that traditional philosophical frameworks, though rooted in antiquity, persist as invaluable assets. Their enduring relevance becomes manifest as they offer unique perspectives on contemporary challenges. Indeed, these age-old philosophical paradigms serve as an enduring repository of human models of reality. Imbued with detailed conceptual frameworks, they contribute to the construction of more sophisticated approaches for navigating our present reality, now augmented by the computational prowess of computers and AI systems. Thus, the fusion of time-honored philosophical insights with modern technological tools becomes paramount in our pursuit of a more nuanced understanding and adept handling of the complex tapestry of causality in our ever-evolving world.

This book is crafted with a specific audience in mind—primarily engineers and individuals engaged in the realm of Artificial Intelligence, seeking a profound comprehension of the intricate dynamics inherent in causal events and their analytical intricacies. Simultaneously, it extends its embrace to my esteemed colleagues immersed in the philosophical nuances of computing. Beyond these specialized circles, the narrative extends its reach to any individual intrigued by the intricacies of contemporary work, especially in light of the transformative influence wielded by computational systems. The exploration within these pages aspires to provide a bridge between the technical intricacies of engineering and the profound philosophical underpinnings, offering a holistic perspective for a diverse readership with varied interests.

I hope that the reading experience of this book is significantly more accessible and enjoyable compared to the challenges encountered during the writing process.

Barcelona, Catalonia, Spain Jordi Vallverdú

Acknowledgments

I am deeply grateful for the invaluable support extended by the ICREA2019 award from the Institució Catalana de Recerca i Estudis Avançats (ICREA). This generous grant, amounting to €200,000 over a dedicated five-year period (2020–2024), has played a pivotal role in enabling me to embark on a thorough exploration of the intricate causal challenges inherent in machine learning, with a specific focus on deep learning. The significance of this award extends beyond mere financial support; it has afforded me the luxury of time—akin to enjoying five consecutive sabbatical years. This temporal freedom has proven essential, allowing for a deliberate and thoughtful examination of the profound dimensions of causality within the dynamic context of machine learning. As the saying goes, sometimes, all we need is time to carefully ponder over significant ideas. The commitment and expertise that guided this research effort have been integral to advancing our comprehension of the intricate interplay between statistics and causal graph reasoning, notably through the lens of Directed Acyclic Graphs (DAGs). The ICREA2019 award, in serving as a cornerstone, has provided the necessary support and resources to delve deeply into these complex realms, contributing to the evolving landscape of artificial intelligence. In expressing my gratitude for this support, I recognize the role of the ICREA2019 award not only in facilitating financial backing but also in affording the luxury of time—an invaluable resource for contemplative exploration and the cultivation of groundbreaking ideas.

Special recognition is extended to my esteemed editor, Celine Chang, whose contributions have been pivotal in shaping the content and quality of this book. Celine's insightful, precise, and invaluable advice has been instrumental in refining the narrative and ensuring clarity in the complex interplay between computer sciences and philosophy. Her trust and confidence in the fruitful crossfield studies undertaken in this case, bridging the realms of computer sciences and philosophy, have been both motivating and reassuring. Celine's dedication to maintaining a high standard of excellence has greatly enriched the intellectual depth and coherence of the manuscript. In expressing my gratitude to Celine Chang, I acknowledge not only her editorial expertise but also her commitment to fostering a collaborative and constructive working relationship. Her role goes beyond meticulous editing; it

encapsulates a shared vision for the impactful convergence of computer sciences and philosophy in the context of AI and deep learning. This book stands as a testament to the synergy between academic disciplines, and Celine's contribution has been indispensable in bringing forth a work that aligns with the highest standards of scholarship. I am sincerely grateful for her partnership throughout this journey, and I look forward to continued collaboration on future endeavors.

I also thank the intellectual support of the following experts: Judea Pearl for his early suggestions about Bayesian analysis, even with his criticisms against my position. To Professor Clark Glymour for his suggestions and advice at an early stage of this research. To Professor Vladimir Naumovich Vapnik, for his kindness in answering my questions and giving me advice about this research, especially on computational abduction. I thank to Professor Christian List, Chair of the MCMP, for offering a silent workspace and intellectual support during my stay at the Munich Center for Mathematical Philosophy, from the Ludwig-Maximilian-Universität München, Germany, in April of 2023.

In addition to the aforementioned scholars, I would like to acknowledge the indirect influence of Demis Hassabis, with whom a direct visitation was not possible. His ideas have served as a source of inspiration throughout the course of this work. Likewise, I extend appreciation to Bernhard Schölkopf, whose impressive ongoing research on causal AI, particularly within the realm of Generative AI, has continued to inspire and surprise me month after month. Their contributions, though indirect, have left a lasting impact on the development of this manuscript.

Finally, it is essential to acknowledge that any potential errors or oversights within this exploration must be attributed to the limitations of my comprehension and not to the esteemed individuals and bodies whose insights have shaped this discourse. Their contributions remain invaluable, and any shortcomings in interpretation are solely the responsibility of the present writer.

Contents