Beyond Predictions: Alignment between Prior Knowledge and Machine Learning for Human-Centric Augmented Intelligence

Doctoral Dissertation
Xia Chen

Supervisor:
Prof. Dr.-Ing. Philipp Geyer

2024

Summary

As machine learning (ML) and artificial intelligence (AI) continue to achieve new advancements across a variety of domains, it is clear that data-driven, end-to-end, connectionist approaches can match and sometimes surpass human performance in many tasks. However, while these systems have achieved significant performance milestones, they fundamentally differ from human intelligence, especially in logic, symbolism, and reductionist mindsets. In design, engineering, and scientific research, this often results in two methodological developments running in parallel: data-driven methods and first-principles approaches, each with its unique strengths, natural limitations, and performance gaps. Inherently, both methodologies are capable of proficiently describing phenomena, or what we call modeling, differently yet without any intrinsic mutual exclusivity, mirroring the long-standing debate between connectionism and symbolism in AI. I recognize that separate modeling in either methodology is not a sustainable approach for long-term development. Instead, the key is to reconcile and integrate our prior knowledge into ML methods for engineering human-centric alignment. This integration is crucial in empirical-dominant domains, niche fields, and scientific explorations that align with logical positivism.

As we are facing an era where machine intelligence is reshaping our understanding of the world and our position within it, recognizing human and machine's unique strengths and limitations is essential in harnessing their combined potential. This dissertation proposes machine assistance for human intelligence augmentation, primarily aiding users such as engineers, designers, and researchers in better decision-making. The core effort is to establish the alignment between human insights and machine capabilities to surpass the limitations inherent in each separated methodology while utilizing information more comprehensively. In this context, this dissertation is structured around three distinct yet essential objectives: Decision-Making Processes Alignment, Methodological Paradigms Alignment, and Interaction Patterns Alignment. Each tackles real-world specific problems as case studies to demonstrate the practical application of the proposed paradigm and methodologies.

The first alignment aims to design a set of fundamental mechanisms to align ML models with complex user decision-making processes. Inspired by the estimation processes of the human nervous system, the framework compiles key mechanisms of uncertainty identification and quantification analysis with incomplete input acceptance in a dynamic, human-in-the-loop environment. It lays a groundwork for aligning information flow from different knowledge-based and data-driven methods to interact with users while enabling a recursive multi-objective optimization, showcasing its practical informational assistance utility in the decision-making process under uncertainty.

The second alignment emphasizes methodological integration by proposing a paradigm to systematically embed human prior knowledge and domain insights into data-driven models. The paradigm first identifies naturally inherited uncertainties from data acquisition conditions, data-driven model mechanisms, and prior domain knowledge. These identifications set the foundation for underscoring their complementary roles in information representation. Building upon this, I organize three hierarchical knowledge integration levels named "Ladder of Knowledge-integrated Machine Learning," corresponding to the three stages in advancing data-driven models with respect to data augmentation, modeling process enhancement, and knowledge discovery. These knowledge types are: modeling knowledge for system description (level 1), inductive logic and disentanglement for extrapolation (level 2), and abstract reasoning and deductive logic (level 3). With applications in engineering case studies analysis, I affirm the framework's efficacy in interpolation, extrapolation, and information representation tasks. With the ladder

ascendance, the methodological paradigm in ML shifts from pattern finding to model building to describe a system, revealing key factors in ML aligning to the human cognitive and learning process.

The last alignment of this dissertation extends the topic to the human-computer interaction (HCI) domain to investigate an ecology of symbiosis between humans and machine assistance, exchanging information among phenomena, data, and prior knowledge, and how such interaction patterns can be enhanced through diverse information collection and processing methods. Furthermore, I primitively investigate potential possibilities of how data-driven methods decode implicit information, such as electroencephalograms, to broaden the bandwidth of information exchange between humans and machines.

To conclude, this dissertation aims to lay the groundwork for aligning human intelligence and machine capacities to advance human-centric intelligence augmentation, while providing a more holistic understanding of the interaction between human intelligence and computational methods.

Keywords

Data-Driven Method; Knowledge Engineering; Knowledge-Integrated Machine Learning; Human-Computer Interaction; Decision-Making Support; Alignment;

. . . . . .
Author's signature

Zusammenfassung

Da maschinelles Lernen (ML) und künstliche Intelligenz (KI) weiterhin Fortschritte in einer Vielzahl von Bereichen erzielen, wird deutlich, dass datengesteuerte, durchgängige, konnektionistische Ansätze in vielen Aufgaben die menschliche Leistung erreichen und manchmal übertreffen können. Diese Systeme unterscheiden sich jedoch grundlegend von der menschlichen Intelligenz, insbesondere in Bezug auf Logik, Symbolismus und reduktionistische Denkweisen. In Design, Ingenieurwesen und wissenschaftlicher Forschung führt dies oft zu zwei parallel verlaufenden methodischen Entwicklungen: datengesteuerte Methoden und Methoden, die auf ersten Prinzipien beruhen, jede mit ihren einzigartigen Stärken, natürlichen Einschränkungen und Leistungslücken. In beiden Methodologien können Phänomene unterschiedlich, aber ohne gegenseitigen Ausschluss beschrieben werden, was die langjährige Debatte zwischen Konnektionismus und Symbolismus in der KI widerspiegelt. Ich erkenne, dass eine separate Modellierung in beiden Methodologien kein nachhaltiger Ansatz für eine langfristige Entwicklung ist. Der Schlüssel liegt vielmehr darin, unser vorhandenes Wissen in ML-Methoden zu integrieren, um eine menschzentrierte Ausrichtung zu erreichen. Diese Integration ist in empirisch dominierten Bereichen, Nischenfeldern und wissenschaftlichen Erkundungen, die mit dem logischen Positivismus übereinstimmen, von entscheidender Bedeutung.

Da wir uns in einer Ära befinden, in der maschinelle Intelligenz unser Verständnis der Welt und unsere Position darin neu gestaltet, ist es wesentlich, die einzigartigen Stärken und Grenzen von Mensch und Maschine zu erkennen, um ihr kombiniertes Potenzial zu nutzen. Diese Dissertation schlägt maschinelle Unterstützung zur menschlichen Intelligenzsteigerung vor, hauptsächlich zur Unterstützung von Nutzern wie Ingenieuren, Designern und Forschern bei besseren Entscheidungsprozessen. Der Kern dieser Bemühung besteht darin, die Ausrichtung zwischen menschlichen Erkenntnissen und maschinellen Fähigkeiten herzustellen, um die Einschränkungen der getrennten Methodologien zu überwinden und Informationen umfassender zu nutzen. In diesem Kontext ist diese Dissertation um drei unterschiedliche, aber wesentliche Ziele strukturiert: Ausrichtung der Entscheidungsprozesse, Ausrichtung der methodischen Paradigmen und Ausrichtung der Interaktionsmuster. Jedes Ziel behandelt spezifische reale Probleme als Fallstudien, um die praktische Anwendung des vorgeschlagenen Paradigmas und der Methoden zu demonstrieren.

Das erste Ziel besteht darin, grundlegende Mechanismen zu entwerfen, um ML-Modelle mit komplexen Benutzerentscheidungsprozessen abzustimmen. Inspiriert von den Schätzungsprozessen des menschlichen Nervensystems, umfasst das Rahmenwerk wesentliche Mechanismen zur Identifizierung und Quantifizierung von Unsicherheiten sowie zur Annahme unvollständiger Eingaben in einer dynamischen, menschzentrierten Umgebung. Es legt das Fundament für die Ausrichtung des Informationsflusses aus verschiedenen wissensbasierten und datengesteuerten Methoden zur Interaktion mit den Benutzern und ermöglicht eine rekursive multi-objektive Optimierung, die ihren praktischen Informationsnutzungsnutzen im Entscheidungsprozess unter Unsicherheit zeigt.

Die zweite Ausrichtung betont die methodische Integration durch das Vorschlagen eines Paradigmas, das menschliches Vorwissen und domänenspezifische Erkenntnisse systematisch in datengesteuerte Modelle einbettet. Das Paradigma identifiziert zunächst natürlich vererbte Unsicherheiten aus den Bedingungen der Datenerfassung, den Mechanismen datengesteuerter Modelle und dem domänenspezifischen Vorwissen. Diese Identifikationen bilden die Grundlage, um ihre komplementären Rollen in der Informationsdarstellung hervorzuheben. Darauf aufbauend organisiere ich drei hierarchische Wissensintegrationsebenen, die als "Leiter des Wissensin-

tegrierten Maschinellen Lernens" bezeichnet werden und den drei Stufen zur Verbesserung datengesteuerter Modelle in Bezug auf Datenanreicherung, Modellierungsprozessverbesserung und Wissensentdeckung entsprechen. Diese Wissensarten sind: Modellierungswissen zur Systembeschreibung (Stufe 1), induktive Logik und Entflechtung zur Extrapolation (Stufe 2) und abstraktes Denken und deduktive Logik (Stufe 3). Mit Anwendungen in ingenieurwissenschaftlichen Fallstudien bestätige ich die Wirksamkeit des Rahmenwerks in Aufgaben der Interpolation, Extrapolation und Informationsdarstellung. Mit dem Aufstieg auf der Leiter verschiebt sich das methodische Paradigma im ML von der Mustererkennung zur Modellbildung zur Beschreibung eines Systems und enthüllt Schlüsselfaktoren, die das ML an den menschlichen kognitiven und Lernprozess angleichen.

Die letzte Ausrichtung dieser Dissertation erweitert das Thema auf den Bereich der Mensch-Computer-Interaktion (HCI), um eine Symbiose zwischen Menschen und maschineller Unterstützung zu untersuchen, wobei Informationen zwischen Phänomenen, Daten und Vorwissen ausgetauscht werden und wie solche Interaktionsmuster durch diverse Methoden der Informationssammlung und -verarbeitung verbessert werden können. Darüber hinaus untersuche ich primitive Möglichkeiten, wie datengesteuerte Methoden implizite Informationen wie Elektroenzephalogramme entschlüsseln können, um die Bandbreite des Informationsaustauschs zwischen Mensch und Maschine zu erweitern.

Abschließend zielt diese Dissertation darauf ab, die Grundlagen für die Ausrichtung von menschlicher Intelligenz und maschinellen Fähigkeiten zu legen, um eine menschzentrierte Intelligenzsteigerung voranzutreiben und ein umfassenderes Verständnis der Interaktion zwischen menschlicher Intelligenz und computergestützten Methoden zu vermitteln.

Schlüsselwörter

Datengesteuerte Methode; Wissensingenieurwesen; Wissensintegriertes Maschinelles Lernen; Mensch-Computer-Interaktion; Entscheidungsunterstützung; Ausrichtung;

Acknowledgement

On my academic journey, I would like to first express my greatest gratitude to Prof. Dr.-Ing. Philipp Geyer for his guidance in my Ph.D.. He provided me with an incredibly free, inspiring, and supportive academic environment that allowed me to develop my research interests and receive his wholehearted academic support. This rare opportunity has been a great honor for me. He actively helped me plan my academic career, promoted my exposure to academic events and opportunities, and provided top-notch platforms for me to gain rich experience in research, teaching, cross-team collaboration, and academic presentations. Moreover, he offered invaluable advice and shared experiences from a friend and mentor's perspective, both in my academic and life. For my dissertation, he continuously helped me refine my work and provided constructive suggestions and philosophical guidance. On the eve of my departure to the USA, we had an in-depth discussion for three and a half hours. He mentioned that the perspective of post-humanism might face challenges in humanist Europe, but he unconditionally supported me from an academic freedom standpoint. He believed that my experiences in North America might provide new insights, and having research experiences in Asia, Europe, and America would enrich my academic profile. I am deeply grateful to have him as my mentor.

I am also grateful to my friends and colleagues during my Ph.D.. They have been integral to maintaining my work-life balance. Whether through engaging in stimulating academic discussions, emotional support during hard times, or sharing moments of relaxation and joy outside of work, their presence has been invaluable. As my dissertation nears completion, I also want to thank Xinwei Zhuang at the University of California, Berkeley, for offering meaningful discussion and providing professional and thoughtful feedback from an academic rigor perspective.

Furthermore, I want to wholeheartedly thank my partner during my Ph.D., my girlfriend Junyi Jiang. She fully supported my research work and was a constant source of encouragement, especially when I felt lost or stressed. Her belief in me, her willingness to listen to my thoughts, and her insightful feedback provided both inspiration and motivation. Junyi's positive attitude toward life and caring nature were especially vital during the challenging times of the pandemic. Her presence made the difficult times much easier and the achievements more meaningful, which helped me navigate the complexities of my academic journey. For all of this and more, I am deeply grateful and loving.

On my academic journey, I recall my mentors at the RWTH E.ON Energy Research Center, Univ.-Prof. Dr.-Ing. Aaron Praktiknjo and Univ.-Prof. Dr. rer. soc. oec. Reinhard Madlener. Working with them was the first time I felt the rigor, passion, and responsibility in the pursuit of knowledge. They generously shared their experiences, igniting my passion for academia.

Lastly, I owe the deepest gratitude to my parents, whose unwavering support and love have been the cornerstone of my journey. They raised me and provided the best conditions for me to explore the world independently. During my ten-year stay in Germany, they supported me financially, remotely cared for me in life, and continuously encouraged me spiritually. They have set excellent examples of the pursuit of knowledge, dreams, and a positive attitude toward life for me. Without their sacrifices and dedication, I would not be who I am today. I love them very much.

Xia Chen
May 20, 2024
in Berkeley, USA

Contents

Introduction **2**
1.1 Contextual Background . . . 2
1.2 Research Questions and State-of-the-Art . . . 4
1.3 Objectives of the Dissertation . . . 7
1.4 Structure of the Dissertation . . . 9
1.5 Summary of the Publications . . . 10
Decision-Making Process Alignment: Machine Assistance Framework **15**
2.1 Machine Assistance Framework . . . 18
2.2 Assistance Extension: Recommendation System . . . 19
Methodological Paradigms Alignment: Pathway toward Prior Knowledge-Integrated Machine Learning **20**
3.1 Overview Pathway . . . 25
3.2 Level 1: Modeling Knowledge for Interpolation - Data Augmentation . . . 26
3.3 Level 2: Disentanglement for Extrapolation - Modeling Process Modification . . . 28
3.4 Level 2: Disentanglement for Extrapolation - Compositionality Extraction . . . 30
3.5 Level 3: Abstraction Reasoning for Representation - Causal Inference . . . 32
3.6 AI for Science: Knowledge-Integrated Machine Learning Application . . . 35
Communication Alignment: Advanced Human-Computer Interaction **36**
4.1 Symboisis between Users and Machines . . . 40
4.2 Data-driven Approaches for User Implicit Signal Decoding . . . 41
Conclusion **42**
5.1 Summary of Contributions . . . 42
5.2 Remaining Challenges and Future Directions . . . 43

List of Acronyms

- AI - Artificial Intelligence
- ML - Machine Learning
- HCI - Human-Computer Interaction
- KIML - Knowledge-Integrated Machine Learning
- LOI - Level-of-Information
- DSS - Decision Support Systems
- MCDM - Multiple Criteria Decision Making
- PINNs - Physics-Informed Neural Networks
- CNNs - Convolutional Neural Networks
- RNNs - Recurrent Neural Networks
- PEMWE - Proton Exchange Membrane Water Electrolysis
- MMM - Multimodal Models
- LLMs - Large Language Models
- AGI - Artificial General Intelligence
- EEG - Electroencephalogram
- RLHF - Reinforcement Learning from Human Feedback
- CBML - Component-Based Machine Learning

1. Introduction

1.1 Contextual Background

> *"By the late twentieth century, our time, a mythic time, we are all chimeras, theorized and fabricated hybrids of machine and organism; in short, we are cyborgs. The cyborg is our ontology; it gives us our politics. the cyborg is a condensed image of both imagination and material reality, the two joined centers structuring any possibility of historical transformation."*
>
> *Donna J. Haraway, 1985*

In recent times, our understanding of the universe and our place within it has been extensively reshaped by advances in cybernetics, artificial intelligence (AI), and informatics. These fields augment human capabilities and open new frontiers where machines excel in tasks that were once exclusively human. This evolving partnership between humans and machines invites us to rethink our role in an age where technology extends our physical and cognitive abilities. As H.G. Wells, a visionary writer, once aptly warned, we are indeed *"in a race between education and catastrophe."* [1] The challenge ahead is not just technological but also ethical, philosophical, and deeply human.

Following this profound observation, it becomes increasingly clear that the advancements in AI and machine learning (ML) are not just milestones in proving computational capabilities but are reshaping our understanding of intelligence and cognition. The rapid progress in ML, particularly in data-driven, end-to-end connectionist approaches, has been nothing short of remarkable. These advancements have led to systems capable of carrying out complex tasks with a proficiency that sometimes outperforms human ability. From mastering strategic games [2, 3] to advancing in generative works [4, 5, 6], AI's capabilities have expanded dramatically, showcasing a level of adaptability and efficiency that was once the sole domain of human intellect.

However, despite these significant milestones, AI and ML systems fundamentally diverge from human intelligence in several critical aspects of organizations [7], learning mechanisms [8], and information processing patterns [9]. Our cognitive processes are not solely data-driven; they are also steeped in a rich blend of prior knowledge, analogy, and, crucially, the capacity to generalize from a few examples [10]. These abilities are deeply rooted in a logical, symbolistic, reductionist mindset and modeling [11, 7], which allows for a profound understanding of abstract concepts, moral reasoning, the ability to interact with complex social dynamics [12],

and decision-making [13]. In contrast, current bloomed AI systems primarily excel at pattern recognition and statistical inference in a connectionist, data-driven, and end-to-end manner [14], often lacking an inherent understanding akin to human thoughts in the dimensions mentioned above. Their most notable differences lie in learning efficiency of solving problems [15], data reliance, flexibility in solving diverse problems, and reasoning ability [16]. These existing differences restrict the integration and mutual understanding between AI and human users, limiting the full access to the exploitation of prior knowledge, data, and machine computational capacities to solve complex tasks collaboratively. Such tasks are widespread in real-world engineering, creative design, and cutting-edge scientific discovery scenarios, which often encounter insufficient or biased understanding, along with limited data availability.

This dichotomy between human and machine intelligence underscores a crucial phase of our journey into the AI era. Machine intelligence is developing at an unprecedented speed and scale, intriguingly, the emergence phenomenon seems to hint at one potential direction for the genesis of intelligence [17, 18] which is that complex behaviors could arise from a system operated under simple principles on a large scale [19]. However, these purely data-driven approaches do not imply that AI will spontaneously understand and align with humans' knowledge formation, the purposes of our actions, values, and the path we comprehend the world. This divergence is not just a technical limitation; it reflects the fundamentally different approach in which machines and humans perceive, process, and interact within the same context. From this perspective, integrating human prior knowledge with data-driven methods addresses more than just the efficiency in problem-solving by AI. It could fundamentally set AI's base cognition, the prior consciousness [20] and regularization [7]. How to position AI as a collaborator not only supports but also enhances human capabilities by bridging the gap between machine and human way of thinking, rather than an adversarial role in its developmental trajectory alongside us, becomes an urgent essential topic.

In this new era, where the boundaries between human and machine intelligence blur, the human-in-the-loop landscape of design, engineering, and scientific research stands at the tip of a transformative wave. Integrating AI in these fields is not merely a technological upgrade but a paradigm shift in our approach to creation and discovery. This dissertation focuses on exploring the role of AI as a collaborative assistant in these realms. We delve into how AI can augment human efficiency in informational access and creativity with its unparalleled computational power and data processing capabilities. In this context, Recognizing the unique strengths and limitations of human and machine intelligence is essential in harnessing their combined potential. The journey ahead is not about comparing or replacing human cognition with machine intelligence but revealing a path to create a symbiotic relationship where each complements the other, leading to deeper insights and enhanced capabilities beyond individual limitations.

This inquiry is rooted in a critical exploration of AI's capability to complement complex design thinking processes [21], engineering problem-solving, and scientific exploration. In the context of a specific task, how can we access, share, and exchange all available information from the phenomenological data and prior knowledge, along with feasible modeling methodologies, though they might be in different formalizations, for the solution? The goal is to establish a symbiotic relationship where AI's analytical strength merges with human intuition and expertise, leading to groundbreaking advancements and innovations. This dissertation investigates how we can better align and utilize data-driven methods in human-centric subject development toward augmented intelligence, supported by applying methodological frameworks with case studies in practical scenarios. We are entering a future where the joint contribution of AI alongside our

mind becomes pivotal in navigating the complexities and challenges of design, engineering, and scientific endeavors.

1.2 Research Questions and State-of-the-Art

In this section, I identify three critical research questions and hypotheses in this dissertation through research challenges with existing work backgrounds. Sequentially, I highlight corresponding solutions that were developed during my Ph.D.. They are formalized into three major aspects, along with the corresponding state-of-the-art background:

Framework Gaps in Model-Based Systems and Human Decision-Making Alignment

Q1: How can we design a machine assistance framework to align with the human decision-making process?

Cybernetic Process in Decision-Making

In design, engineering, and scientific discovery tasks, our approaches are naturally embedded with a cybernetic process of observation, intervention, and feedback loop. Although current flourished ML models exceed human performance in many domains, most of them lack such process and are trained to focus primarily on solving specific tasks in an end-to-end behavior, essentially mapping inputs to outputs sequentially [14]. Among various ML approaches, deep reinforcement learning [22] follows the cybernetics mindset the most: By setting a data-driven model (agent) in dynamic virtual environments, the agent with predefined rewards is trained to interact with the environment to behave optimally unattended. Interestingly, the most recent emerging techniques of reinforcement learning from human feedback (RLHF) [23] achieve significant advantages for generative AI [24]. Even so, it only involves human feedback during the training phase. Enlightened by this leap, I observed that most fields still lack a systematic, mutual informative exchange framework for aligning these models with human priors in an interactive feedback loop mechanism. This misalignment hinders ML's effectiveness in human-centric control, design, and decision-making process assistance.

Uncertainties Incorporation

Furthermore, this gap is also reflected in the absence of uncertainty analysis in ML models. In real-world problems, various aleatoric and epistemic uncertainties exist [25] - the former natural-inherited and ineradicable (e.g., 50% uncertainty of the result in coin flip), the latter reducible through enhanced cognition, tools, and information. For instance, in building design and engineering, a discussion about a framework of Human-Building interaction reported the engagement of envisioning designers' thoughts regarding the interface and constraints with uncertainty assessment [26]. The importance of uncertainty estimation is proven as it greatly shapes our perception of the world [27]. A hierarchical predictive coding framework has been proposed to minimize the error between "perception" and "expectation" [28]. Since sensory input is typically ambiguous and noisy, the ability to make predictions with uncertainties based on prior information is a key feature of brain function [27, 29]. This capability should also be incorporated into machine assistance frameworks to provide robust results with enhanced model flexibility and applicability, allowing the frameworks to process information in a more human-like manner.

Model Truthfulness and Transparency

However, most ML models are designed for an end-to-end predictive behavior with fixed input formats, deterministic outputs, and static model structures after the training phase. Many tasks in engineering design include decision-making processes towards a certain optimal under uncertainties. Although Decision Support Systems (DSS) [30] use techniques such as Multiple Criteria Decision Making (MCDM) for complex scenarios [31], the general knowledge-based approaches with hard-coding contextual constraints find it hard to exhaust all possibilities in the background of data-booming digitalization progress and multi-disciplinary requirement. From the process perspective, this form also contradicts the dynamic human interaction of observation, intervention, and feedback, which entails continuous information updating and iteration, incorporating probabilistic model features such as Bayesian [32] and mutual information [33, 34] characteristics. In this context, although some noticeable advancements in probabilistic ML models [35, 36] stand out, their black-box nature raises the model trust issues for engineers, designers, and researchers [37], causing them to stand on the opposite side and prioritize adapting knowledge-based, first-principles models only in real-world practice [38]. The key factor is the relationship between the user's understanding and machine process transparency, which refers to the clarity of the input processing mechanism. In this context, expediency, transparency, and explainability requests have raised concerns and argued for rebalancing through a task-dependent symbiosis of fitting and interpreting [39, 40], in which two sources of information guide learning: (a) outcomes from data and trained data-driven model (b) engineering interpretation from knowledge-based models of how data are generated.

Thus, a fundamental gap remains open: the absence of a data-driven framework that dynamically handles incomplete inputs and aligns with the user interaction process and prior understanding consistency. Such a framework is essential for decision-making support under uncertainties, serving as a foundation form of intelligence augmentation.

Methodological Challenges in Integrating Domain Knowledge into Data-Driven Approaches

Q2: How do we incorporate prior knowledge into data-driven methods to minimize the performance gap, enhance applicability and user truthfulness, and fully utilize available information?

Prior Knowledge Integration

ML and AI have flourished, primarily driven by a connectionism mindset, the growth of data availability, and the development of computation optimization approaches, surpassing the limitations of rule-based, prior knowledge encoding, and symbolic methods [11]. Engineering inverse problems with hidden complex physics sometimes are effort-expensive for simulation or prediction [41] to solve "how" problems in semantic description through traditional approaches. However, the real-world implementation and performance of ML models heavily depend on the scope of data [7, 42], leading to model inefficacy in cases of extrapolation, sparse, inconsistent, and noisy input. This is because the model only extracts information from solely the given data sources. Furthermore, ML approaches usually aim to recognize patterns rather than build models to describe a phenomenon [18], which means they miss an essential conceptualization ability to comprehend data that encapsulates the information representing a system or phenomenon. For instance, comparing human learning and AI in tasks like handwritten character recognition [43] reveals fundamental differences: Humans learn from fewer examples and develop richer,

more flexible representations, indicating a need for AI systems to allow flexible knowledge application [18]. In practical engineering, design, and scientific discovery, there is a need for utilizing information from prior knowledge and learning under a small-sample dataset to achieve sufficient generalization, extrapolation, and explainability based on induction – attributes inherent in symbolic, domain-specific knowledge. This need involves a series of techniques that enable transferring domain knowledge into acknowledged formats for data-driven methods, which allows for the full utilization of available information from different perspectives. Despite many research efforts to address this issue in practice, most of them remain on a specific domain or case scope. A systematic framework for pinpointing the beneficial characteristics of integrating our prior knowledge into data-driven methods remains underdeveloped. The questions of what prior knowledge to integrate, how to incorporate it into data-driven methods, and which advantages and problem-solving capabilities it could bring to specific domains have yet to be systematically discussed.

Reasoning and Knowledge Discovery

Initiated by medical statistic development [44], causal inference is a critical research topic in many domains given observational, multivariate data to reveal an effect and the cause that's influencing it, making interventions ("What-if..") and counterfactual reasoning ("What if I had…") possible [45, 46]. Most engineering optimization tasks use a set of design parameters to represent a system; however, the dependencies among the different parameters are not carefully considered, taking into account their influence on simulation results. The well-known mantra in statistics, "correlation does not imply causation" [45, 47], remains under-appreciated and rarely discussed in the context of ML applications. Causality is commonly confused with correlation, but the former presents a different interpretation from observational data: it analyzes the asymmetric change and response between cause and effect. Such reasoning ability is essential for knowledge discovery directly from the data and could overcome potentially biased prior knowledge of the individual, especially with the trend of multi-disciplinary requirements in many domains. This causal mindset or knowledge used for knowledge discovery reveals a broader topic: what fundamental factors contribute to human-like thinking for such unsupervised knowledge discovery and representation? More in-depth views emphasize the importance of early inductive biases and learning algorithms that extract knowledge from limited training data [48, 49]. Combined with knowledge integration, such a path toward closing the loop between knowledge discovery and application remains underdeveloped.

Exploration in Interaction Forms and Symbiosis of Human-Machine Modes

Q3: How do we incorporate and improve the user-computer-interaction patterns?

Informatic Symbiotic Construct

Finally, this dissertation refocuses on utilizing the advantages of ML approaches by linking them to reinforce our cognition and interaction patterns. As an end-to-end learning behavior, ML models are trained on comprehensive datasets, enabling the capture of complex patterns and relationships directly from the data, which bypasses the need for manual feature understanding in data as a prerequisite before they can be used. This holistic learning method aligns closely with the nature of human cognitive processing, which also operates from perception to decision-making in a continuous flow. The advent of representation learning [50] in theoretical exploration, i.e., learning representations of the data that make it easier to extract useful information when building classifiers or other predictors, and further multimodal

learning [51, 52] expands the capabilities of AI by utilizing diverse data types – visual, textual, auditory, and more – much like human sensory processing. However, a critical gap remains in integrating this multimodal approach with domain-specific insights. While AI can process and analyze data from various modalities, incorporating human reasoning and decision-making by few-shot learning via induction, as mentioned in the last subsection, is still challenging. These rich, theory-like knowledge structures, characteristic of human thought, are difficult for current AI models to replicate. In this context, constructing a human-in-the-loop and human-centric symbiosis construct that shares information, computation, and decision actions with enhanced efficacy is vital. Despite early exploration of Mixed-Initiative Human-Machine Systems [53], nowadays, only limited discussions in specific domains to prove the advantages in decision augmentation [54, 55]. With the emerging advanced methods mentioned above, discussions on interaction forms and enhancing the information bandwidth between humans and machines for domain-specific practice are still in their early stages.

1.3 Objectives of the Dissertation

As we confront structuring a historical transformation moment with any possibility as the rapid development of ML and AI, the integration of these computational models with human intelligence becomes increasingly significant. This dissertation aims to address critical gaps and explore new paradigms that ensure the advancement of AI technologies not only aligns with but also benefits human-centric decision-making processes as intelligence augmentation. By offering practical contributions across various fields, I validated their efficacy in addressing domain tasks in real-world scenarios. The following key objectives drive this research:

1. **Decision-Making Processes Alignment**: This objective focuses on developing a machine assistance framework for human decision-making support. It seeks to tailor ML mechanisms that adapt and align effectively with human-involved collaboration processes, especially in complex tasks in engineering and design. It emphasizes identifying and quantifying uncertainty, handling incomplete inputs, and ensuring consistency to facilitate informed decision-making under dynamic conditions. The underlying research question investigates how these models can be adapted to complement and reinforce human decision-making, ensuring the integration of AI capabilities with human intuition and expertise.

2. **Methodological Paradigms Alignment**: The second objective aims to construct a paradigm that systematically incorporates domain-specific insights and first principles into data-driven approaches. This integration is designed to embed ML models with an understanding similar to human prior knowledge, thereby improving their applicability and effectiveness in real-world tasks. This integration involves first distinguishing different knowledge regarding their inherent characteristics and novelties for decomposition guidance. Based on the categorized levels, proposing a hierarchical framework – Ladder of Knowledge-Integrated Machine Learning (KIML): By acknowledging the limitations and characteristics between different modeling approaches, KIML sorts comprehensive methodologies of how to combine domain-specific insights and prior knowledge with data-driven models. The combined model presents distinguished improvements in various engineering contexts in interpolation, extrapolation, and information representation. With the ascending hierarchy in knowledge integration, the model representation paradigm shifts from pattern recognition to building models. The research question here delves into the methods and impacts of integrating different types of domain knowledge into data-driven models, enhancing their predictive accuracy, contextual relevance, and advanced reasoning and knowledge discovery abilities.

3. **Interaction Pattern Alignment**: This objective explores interaction pipelines between humans and AI systems to improve communication and collaboration efficiency. It focuses on investigating possibilities of multimodal data interpretation and refining interaction patterns to reduce information exchange loss. This includes studying advanced HCI representations on how we use ML to better recognize implicit, useful user feedback. The aim is to broaden the information exchange bandwidth that facilitates seamless collaboration with human cognitive abilities utilizing AI's computational power in engineering and scientific contexts in a complementary manner. The corresponding research question seeks to understand how these interaction patterns can be optimized for mutual understanding and collaborative problem-solving, breaking through the information exchange limitation between human intuition and machine intelligence.

Ultimately, this research aspires to develop a human-centric decision-support framework capable of human capabilities across multiple domain information sources. This tool is designed to align the data-driven approach with our prior domain knowledge and cognitive characteristics, along with the ML method's computation capacity and scalability to address real-world complex tasks, as presented in Figure 1.1. The core idea is to maximize the utilization of all known, differently represented information to achieve intelligence augmentation.

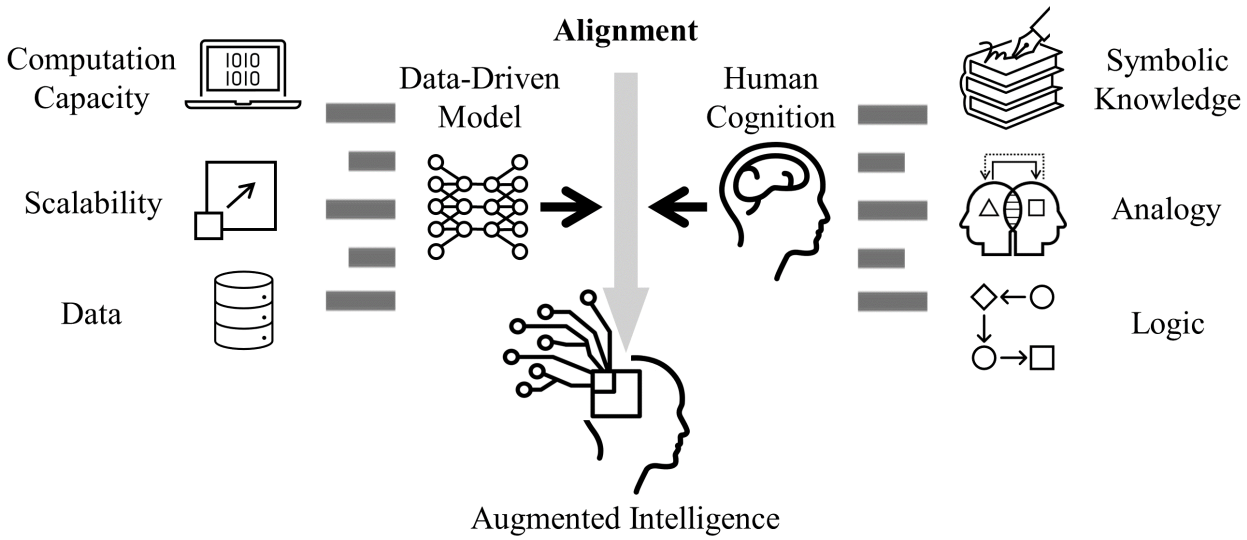

Figure 1.1: Dissertation Objectives: Alignment of Human Cognitive Characteristics with Data-Driven Approaches as A Human-Centric Decision-Support Tool

The main contributions in this dissertation are summarized as follows:

- Design a machine assistance framework that aligns AI with the human decision-making process and aids in engineering and design scenarios.
- Propose a hierarchical "Ladder of Knowledge-integrated Machine Learning" paradigm that systematically integrates prior knowledge into data-driven methods, improving their real-world applicability.
- Explore advanced interaction patterns between human and machine assistance, focusing on efficient information exchange and collaboration in solving problems in complex environments.

1.4 Structure of the Dissertation

In this dissertation, the introduction section sets the time background, the current development of the topic, and the research objectives. It highlights the need and drive for a more integrated approach to decision-making processes in the context of AI and ML.

As the core of this dissertation, Figure 1.2 presents the structure and bullet point of the organization between chapters 2 and 4.

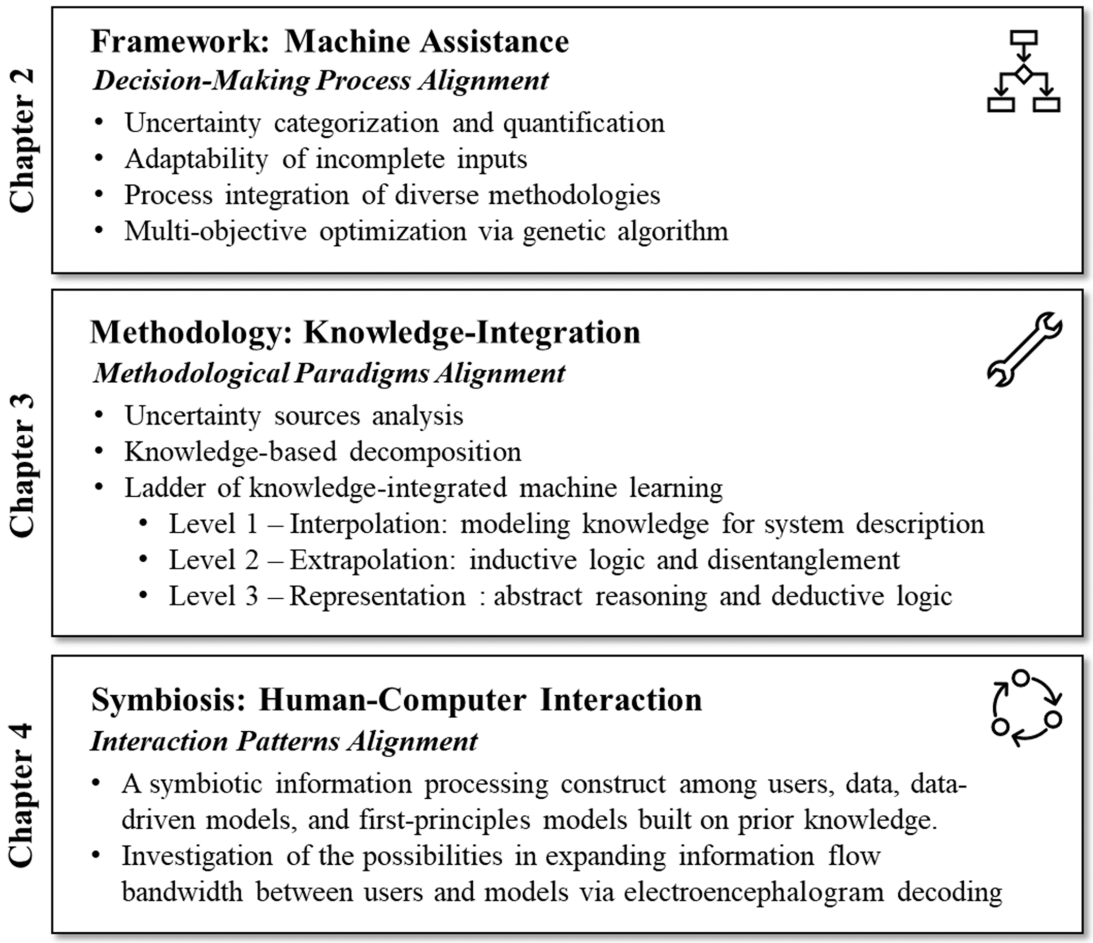

Figure 1.2: Dissertation Structure and Bullet Points of main Chapters

Finally, the conclusion section summarizes the impact of this dissertation's contributions and exemplifies remaining challenges and future directions.

1.5 Summary of the Publications

The work presented in this dissertation is based on twelve peer-reviewed publications in international, scientific journals and conference proceedings:

1.5.1 Publications in international journals

1. Chen, X., Teng, X., Chen, H., Pan, Y., & Geyer, P. (2024). Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. Biomedical Signal Processing and Control, 87, 105475.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration.*
    - *Xiangbin Teng: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Resources.*
    - *Han Chen: Data Curation, Software, Validation.*
    - *Yafeng Pan: Supervision, Methodology, Validation, Writing - Review & Editing, Resources.*
    - *Philipp Geyer: Supervision, Conceptualization, Writing - Review & Editing, Project administration, Funding acquisition.*

2. Chen, X., Sun, R., Saluz, U., Schiavon, S., & Geyer, P. (2023). Using causal inference to avoid fallouts in data-driven parametric analysis: A case study in the architecture, engineering, and construction industry. Developments in the Built Environment, 100296.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration.*
    - *Ruiji Sun: Investigation, Writing - Review & Editing.*
    - *Ueli Saluz: Data Curation, Software, Validation.*
    - *Stefano Schiavon: Supervision, Validation, Writing - Review & Editing, Resources.*
    - *Philipp Geyer: Supervision, Conceptualization, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

3. Chen, X., & Geyer, P. (2022). Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty. Applied Energy, 307, 118240.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Philipp Geyer: Supervision, Conceptualization, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

4. Chen, X., Guo, T., Kriegel, M., & Geyer, P. (2022). A hybrid-model forecasting framework for reducing the building energy performance gap. Advanced Engineering Informatics, 52, 101627.

- *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
- *Tong Guo: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
- *Martin Kriegel: Supervision, Validation.*
- *Philipp Geyer: Supervision, Validation, Writing - Review & Editing, Project administration, Funding acquisition*

5. Chen, X., Singh, M. M., & Geyer, P. (2024). Utilizing domain knowledge: robust machine learning for building energy performance prediction with small, inconsistent datasets. Knowledge-Based Systems, 294, 111774.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Manav Mahan Singh: Methodology, Software, Writing - Review & Editing, Data Curation, Resources.*
    - *Philipp Geyer: Supervision, Conceptualization, Validation, Investigation, Writing - Review & Editing, Project administration, Funding acquisition.*

6. Chen, X., Abualdenien, J., Singh, M. M., Borrmann, A., & Geyer, P. (2022). Introducing causal inference in the energy-efficient building design process. Energy and Buildings, 277, 112583.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Jimmy Abualdenien: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing.*
    - *Manav Mahan Singh: Methodology, Software, Writing - Review & Editing, Data Curation, Resources.*
    - *André Borrmann: Supervision, Validation, Writing - Review & Editing.*
    - *Philipp Geyer: Supervision, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

7. Chen, X., Rex, A., Woelke, J., Eckert, C., Bensmann, B., Hanke-Rauschenbach, R., & Geyer, P. (2024) Machine Learning in Proton Exchange Membrane Water Electrolysis–a Knowledge-Integrated Framework. Applied Energy, under review. Available at SSRN 4743024.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Alexander Rex: Conceptualization, Methodology, Formal analysis, Validation, Writing - Original Draft, Writing - Review & Editing, Data Curation, Resources.*
    - *Janis Woelke: Conceptualization, Methodology, Validation, Writing - Original Draft, Writing - Review & Editing, Resources.*

- *Christoph Eckert: Validation, Writing - Review & Editing.*
- *Boris Bensmann: Validation, Writing - Review & Editing.*
- *Richard Hanke-Rauschenbach: Supervision, Resources, Project administration, Validation, Funding acquisition.*
- *Philipp Geyer: Supervision, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

1.5.2 Publications in international conference proceedings

1. Chen, X., & Geyer, P. (2023). Sustainability recommendation system for building design alternatives under multi-objective scenarios, accepted by 30th International Workshop on Intelligent Computing in Engineering, EG-ICE 2023, London, UK.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Philipp Geyer: Supervision, Conceptualization, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

2. Chen, X., Guo, T., & Geyer, P. (2021). A hybrid-model forecasting framework for reducing the building energy performance gap. In 28th International Workshop on Intelligent Computing in Engineering, EG-ICE 2021. Berlin, Germany, 2021, special issue on Advanced Engineering Informatics.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Tong Guo: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Philipp Geyer: Supervision, Validation, Writing - Review & Editing, Project administration, Funding acquisition*

3. Chen, X., Singh, M.M. & Geyer, P. (2021). Component-based machine learning for predicting representative time-series of energy performance in building design. In 28th International Workshop on Intelligent Computing in Engineering, EG-ICE 2021. Berlin, Germany.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Manav Mahan Singh: Methodology, Software, Writing - Review & Editing, Data Curation, Resources.*
    - *Philipp Geyer: Supervision, Conceptualization, Validation, Investigation, Writing - Review & Editing, Project administration, Funding acquisition.*

4. Chen, X., Cai, X., Kümpel, A., Müller D., & Geyer, P., (2022). A Dynamic Feedforward Control Strategy for Energy-efficient Building System Operation. In Passive and Low Energy Architecture, PLEA 2022, Santiago de Chile, Chile.

- *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.*
- *Xiaoye Cai: Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Visualization.*
- *Alexander Kümpel: Supervision, Resources.*
- *Dirk Müller: Supervision, Validation.*
- *Philipp Geyer: Supervision, Validation, Writing - Review & Editing, Project administration, Funding acquisition.*

5. Chen, X., & Geyer, P. (2023). Pathway toward prior knowledge-integrated machine learning in engineering. In 18th International IBPSA conference and Exhibition, Building Simulation 2023, Shanghai, China. arXiv preprint arXiv:2307.06950.

    - *Xia Chen*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.*
    - *Philipp Geyer: Supervision, Validation, Investigation, Writing - Review & Editing, Project administration, Funding acquisition.*

The following lists provide the remaining publications I (co-)authored during my Ph.D..

Publications in international journals

1. Zong, C., Chen, X., Deghim, F., Staudt, J., Geyer, P., & Lang, W. (2024). A holistic two-stage decision-making methodology for passive and active building design strategies under uncertainty. Building and Environment, 111211.

    - *Xia Chen: Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft.*

2. Guo, T., Chen, X., Geyer, P., & Kriegel, M. Coupling Component-Based Machine Learning Framework with Temporal Uncertainties for Efficient District Heating System Optimization. Available at SSRN 4819804.

    - *Xia Chen: Conceptualization, Methodology, Validation, Writing - Review & Editing.*

3. Geyer, P., Singh, M.M. & Chen, X., (2021). Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning. arXiv preprint arXiv:2108.13836.

    - *Xia Chen: Methodology, Software, Validation, Writing - Original Draft.*

4. Chen X., Zhang Y., & Cai X. (2022). Frontiers of carbon neutrality in EU-German building sector, Heating Ventilating & Air Conditioning, TU-023; X322.

    - *Xia Chen*: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.*

Publications in international conference proceedings

1. Chen X., Saluz U., Staudt J., Margesin M., Lang W., & Geyer P. (2022). Integrated data-driven and knowledge-based performance evaluation for machine assistance in building design decision support, In 29th International Workshop on Intelligent Computing in Engineering, EG-ICE 2022. Aarhus, Denmark.

    - *Xia Chen*: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.*

2. Guo, T., Chen, X., Geyer, P., & Kregel, M. (2023). Performance investigation of different topology organizations in district heating systems with component-based machine learning. In 18th International IBPSA conference and Exhibition, Building Simulation 2023, Shanghai, China.

    - *Xia Chen*: Conceptualization, Methodology, Investigation, Writing - Original Draft.*

3. Wang, S., Chen, X., & Geyer, P. (2023). Feasibility Analysis of POD and Deep-autoencoder for Indoor Environment CFD Prediction. In 18th International IBPSA conference and Exhibition, Building Simulation 2023, Shanghai, China.

    - *Xia Chen: Conceptualization, Methodology, Software, Validation, Writing - Original Draft, Visualization.*

2. Decision-Making Process Alignment: Machine Assistance Framework

"Is the car part of our body when we drive?"

When driving a car, the boundary between the human body and the vehicle becomes blurred, not in a physical sense, but in a cybernetic context. This fusion of driver and vehicle is conceptualized as a singular system of perception, feedback, and action; a notion that aligns with the pioneering principles of cybernetics proposed by Norbert Wiener [56]. This integrated system enables us to transcend our individual capabilities and reach great distances faster and safer, which exemplifies the essence of cybernetics' first wave: challenges the traditional perception of human boundaries as fixed and points out their fluid, constructed nature [57]. Analogically, the subject of knowledge is perceived as a construct, which means the reconfiguration of our bodies into information systems implies that our cognitive processes can be augmented as well. This paradigm shift is mirrored in our increasing reliance on external processed information input as assistance. In contemporary practice, assisting the process of decision-making based on data-driven methods with numerical analysis and modeling has become one of the important means in engineering, design, and scientific discovery domains.

As the cybernetics principle emphasizes, it investigates the science of control and communication in critical information feedback loops between systems and their environments, instead of the system itself. For instance, physical design, such as automotive, human factor engineering, or ergonomics [58] plays a pivotal role, requiring in-depth research into human capabilities and behaviors to design vehicle interfaces effectively. These principles are not just confined to physical systems but are equally applicable to external information systems, necessitating the same rigorous approach in engineering interactions with humans. An illustration of this analogy is demonstrated in Figure 2.1, highlighting the objective of this chapter: investigate the key human factors to construct machine assistance.

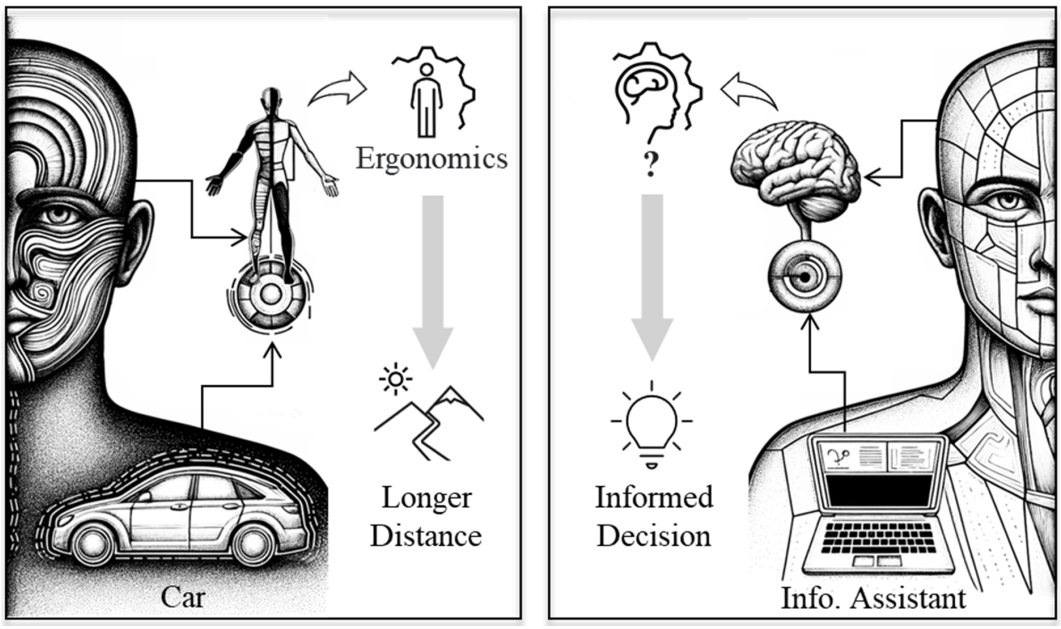

Figure 2.1: Integrating Human Factors in System Design - Analogy from Physical and Informational.

From a teleological perspective [59], choosing the decision-making process as a research scenario provides the motivation and purpose of this study; that is, building such a framework can be regarded as assisting in pursuing optimal results or decision advancements in specific engineering, design, or scientific discovery tasks. I focus on improving the decision-making quality of human users through computational mechanisms and method designs to ensure enhanced efficiency and synergy between decision-makers with optimized results, improved innovation, and process transparency. Ensure that this framework, while achieving its objectives, is also adaptive, interpretive, and ethical.

Moreover, designing this framework serves another fundamental purpose: to make consistent predictions in a dynamically changing world with continuous input variations. I identify key factors and necessary processes in frameworks to enable users to understand AI recommendations and make informed decisions. This approach not only makes the decision process more human-centric and controllable but also avoids the singular modeling approach limitation and traditional ML's end-to-end 'black box' predictive behavior. Therefore, the objective of this chapter is to design basic mechanisms for the framework to accept information, handle uncertainties, and coordinate information flows between different models in decision-making processes. This aligns with a significant school of thought in cognitive science: Parallel Distributed Processing (PDP) [60, 61], which emphasizes parallel computation and accomplishes complex computations by combining simple units [18].

Thus, this chapter comprises two research papers. The first paper (Section 2.1) presents a key general framework drawing inspiration from the human nervous system's estimation process to assist users in complex decision-making processes. It comprises basic, compositional, and scalable mechanisms to enable the framework to conduct uncertainty quantification, information flow among different data-driven and first-principles methods unification, and dynamic interaction with users. In a practical case, I applied it to the early phase of energy-efficient building design scenarios; it sets the foundation for integrating domain knowledge with ML for advancing intelligence augmentation to enhance our ability to interact, interpret, and innovate

with machine assistance in the modern world.

The second paper (Section 2.2) introduces a recommendation system, an extension based on the previous machine assistance paper within the same application scenario. The recommendation system is equipped with the drive to reach solutions in processing vast potential design spaces, analyzing scenarios, and aligning with ongoing processes; based on the given information input, this system bridges complex evaluations and sustainable outcomes via a combination of a genetic algorithm for optimal outcomes analysis, with unsupervised clustering algorithm for new design pattern discovery. The case study, grounded in real-world energy performance data, demonstrates its action and implications for holistic sustainability objectives.

In summary, this chapter presents a fundamental machine assistance framework with the following contributions:

- **Uncertainty Identification and Estimation:** With different mechanisms, the framework can identify and quantify different uncertainties, generating outputs that provide approximations and probabilistic representations of states relative to foundational statistic facts.
- **Adaptability of Incomplete Inputs:** The framework is designed to accept and process incomplete inputs, making it flexible and adaptive in broader real-world scenarios where complete data may not always be available.
- **Information Flow Alignment with User Decision-Making:** The framework ensures a coherent information flow that aligns with the user's decision-making process in a cybernetic manner, which facilitates informed and human-centric decision-making, integrating machine intelligence with human intuition.
- **Integration of Diverse Modeling Approaches:** The framework permits different modeling methodologies to integrate within the workflow, allowing comprehensive analysis drawing on the strengths of various modeling approaches.
- **Provision of Optimal Solution Sets Exploration:** The framework is integrated with evolution mechanisms to determine the set of optimal solutions for constrained or unconstrained problems, aiding users in navigating decision-making landscapes by offering well-defined choices and effective outcomes.

Outline

2.1 Machine Assistance Framework 18
Bibliography 18
2.2 Assistance Extension: Recommendation System 19
Bibliography 19

2.1 Machine Assistance Framework

Outline

Bibliography 18

Bibliographic Information

- Chen, X., & Geyer, P. (2022). Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty. Applied Energy, 307, 118240.

Bibliography

2.2 Assistance Extension: Recommendation System

Outline

Bibliography 19

Bibliographic Information

- Chen, X., & Geyer, P. (2023). Sustainability recommendation system for building design alternatives under multi-objective scenarios, accepted by 30th International Workshop on Intelligent Computing in Engineering, EG-ICE 2023, London, UK.

Bibliography

3. Methodological Paradigms Alignment: Pathway toward Prior Knowledge-Integrated Machine Learning

"How would we explain the concept of an elephant to an alien?"

In this hypothetical scenario, modeling methodologies to describe the elephant (system) is the key. One promising method is through the lens of deconstruction: we describe it by its parts, the long tusks, hard skin, massive ears, etc. It is a mindset to break a big, complex object into smaller and understandable pieces. Such a mindset is closely related to most of our knowledge-based, first-principles methodologies, using symbolic formulas and logic to construct a system and describe the problem. These methodologies follow the mindset of interpretsing a complex system as the sum of its parts; we refer to these perspectives as reductionism.

On the other hand, instead of a detailed analysis of each part, we'd look at the elephant as a complete entity. We observe its behavior, how it interacts with the environment, and how it moves. This method is similar to the black-box, ML methods we often use, collecting data from a system and describing it via an end-to-end, data-driven process. The emphasis is on overall experience rather than explainable details. Such a perspective of revealing properties of a whole system beyond those of its parts is categorized as a holistic mindset. Figure 3.1 illustratively presents both mindsets simultaneously.

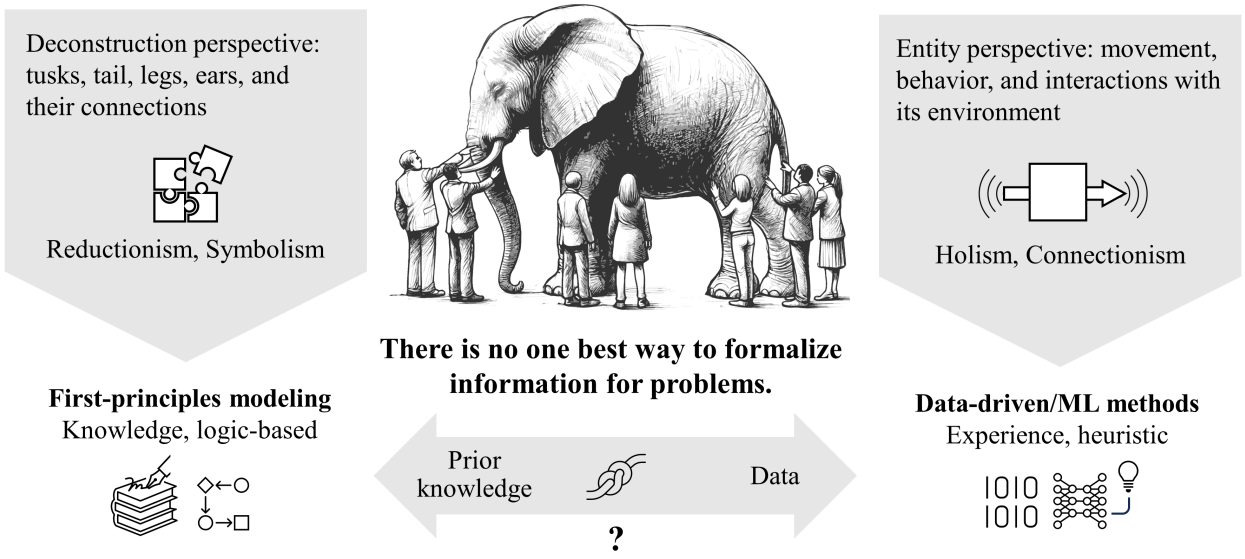

Figure 3.1: Integrating Deconstruction and Holistic Approaches in Engineering: The Elephant Analogy

However, is one necessarily better than others to describe an elephant? There is no single best approach to formalizing problem information. The decomposition mindset offers a clear, interpretable process for understanding a system, yet "The whole is greater than the sum of the parts" (Aristotle, 384 - 382 BC) is well-known to systems scientists and engineers. Correspondingly, ML-represented methodologies present a holistic pattern recognition of a system yet remain in a black-box nature for model interpretability. In algorithmic terms, this phenomenon is known as the "No-Free-Lunch Theorem" [62]. We must always utilize all available information between knowledge and data, often employing both methods to model a complex system issue. Borrowing from physicist (and ahead-of-his-time model-building proponent) Phillip Anderson, from his 1977 Nobel Prize acceptance speech [63, 19]:

"The art of model-building is the exclusion of real but irrelevant parts of the problem, and entails hazards for the builder and the reader. The builder may leave out something genuinely relevant; the reader, armed with too sophisticated an experimental probe or too accurate a computation, may take literally a schematized model whose main aim is to be a demonstration of possibility."

This leads to the key question in this chapter: How can we integrate these two complementary approaches within a specific domain to maximize the use of information from both knowledge and data? In this chapter, The focus is particularly on how to integrate symbolic, logic-based prior knowledge into data-driven, holistic ML methods. This involves addressing two main keywords in this chapter: **Machine Learning** and **Prior Knowledge**, implicitly raising two questions:

1. Why Machine Learning?

Notably, I prioritize ML as the main category of data-driven methods in this dissertation. As an important type of narrow AI, ML methods allow direct data training in an end-to-end behavior, learning data patterns through mathematical optimization (e.g., gradient descent [64]) without explicit programming. This type of method has rapidly developed and proven effective in many research and engineering applications because their designed mechanisms are better suited than the human brain to quickly search more state spaces and fit data with sufficient input in specific formalized problems (e.g., Go [2]). Here, I briefly clarify the basic mechanism of four main types of ML methods:

- **Supervised Learning** aims to improve the algorithm's pattern recognition capabilities from the labeled data.
- **Unsupervised Learning** aims to gradually match the statistical patterns within the model to the unlabelled input data's statistics.
- **Reinforcement Learning** is designed to learn optimal behaviors through set reward and error, by interacting with an environment.
- **Semi-Supervised Learning** frames the problem as a supervised learning task to generate labeled data from unlabeled data. A similar concept is *Self-supervised Learning*, which understands the underlying properties of a given dataset with some supervisory signal, often by formulating a pretext task where the data provides its own supervision. These learning techniques are designed to leverage the high cost of data labeling in real-world tasks.

2. Why Integrate Prior Knowledge?

In a nutshell, ML applications still carry various limitations in real-world practice:

1. The contradiction between pure ML methods' data reliance and the scarcity of data within niche domains in reality (Section 3.2);
2. Less flexibility and transferability of ML models trained on the given data information against the need for generalization in practice (Section 3.3);
3. The model opacity characteristic with pattern-fitting behavior conflicts with the demand for explainability and abstraction understanding for learning from a few examples in real-world task-solving (Section 3.4 & 3.5).

Essentially, these progressive limitations are due to the lack of shared information between the data-driven methods and the human prior knowledge. This information gap restraints ML and first-principles methods in real-world applications. For this reason, the core of this chapter is to bridge the gap between these two methodologies through the proposal of the Knowledge-Integrated Machine Learning (KIML) (Section 3.1) paradigm, aiming to maximize the effort of utilizing information from knowledge and data. Furthermore, KIML carries a hierarchical structure as a ladder; in the process of integrating prior knowledge, we also gradually shift the ML paradigm from "pattern recognition" to "model building" [18]. This paradigm shift allows machines to align with human ways of thinking to achieve advanced capabilities, such as reasoning and answering critical "what-if" questions in science, engineering, and design domains. Along with the ladder, critical conclusions are conducted to identify three different types of knowledge based on their characteristics throughout this chapter. Interestingly, the ascending knowledge, along with the KIML, corresponds seamlessly to a fundamental and well-known theory in child cognitive development: Piaget's stages [65], which provides solid evidence for the proposed paradigm from a cognitive science perspective. The following summarizes the corresponding knowledge types, applications, and cognitive stages:

1. Level One Knowledge: Modeling Knowledge for System Description
   (a) **Essence:** Explicit knowledge to directly describe and model the system, reinforcing data through modeling knowledge. This level focuses on using domain-specific insights to explain or predict phenomena, enhancing data's interpretability and filling in gaps for improved interpolation.
   (b) **Application:** The direct observations and domain knowledge laying the foundational understanding of a system guide the initial modeling efforts, serving as a groundwork step in making informed decisions and predictions within known prior knowledge.
   (c) **Corresponding Cognitive Stage (Piaget):** Piaget Preoperational Stage (2 to 7 years) - At this stage, Individuals begin thinking symbolically, using language and pictures, but their thinking is concrete.
2. Level Two Knowledge: Inductive Logic and Disentanglement for Extrapolation
   (a) **Essence:** At this level, knowledge leverages inductive logic and reasoning to disentangle abstract factors from diverse systems, focusing on inductive, generalizable aspects in system compositionality. The outcome of this disentanglement is used

directly for modeling process modification, enabling to make predictions beyond the observed data.

(b) **Application:** This approach mirrors the cognitive process of organizing and structuring concrete events logically, allowing for the application of specific information to broader principles. It enables the system to navigate and predict situations by understanding the underlying components and their interactions.

(c) **Corresponding Cognitive Stage (Piaget):** Piaget Concrete Operational Stage (7 to 11 years) - Individuals begin to think logically about concrete events and understand the concept of conservation. They start using inductive logic, reasoning from specific information to a general principle.

3. Level Three Knowledge: Abstract Reasoning and Deductive Logic

(a) **Essence:** Level Three Knowledge involves abstract thinking and reasoning about hypothetical problems using deductive logic, using general principles to form a specific conclusion. This level abstractly models the system, engaging with theoretical constructs and principles to guide decision-making and predictions.

(b) **Application:** Comparable to advanced human cognitive abilities, this level allows for sophisticated problem-solving and understanding of concepts that require abstract reasoning. It represents a critical step in AI's ability to mimic human thought processes, enabling machines to tackle complex, unseen problems through generalization and deduction.

(c) **Corresponding Cognitive Stage (Piaget):** Formal Operational Stage (12 years and up) - Individuals begin to think abstractly and hypothetically, using deductive logic to reason from a general principle to specific information.

Therefore, this chapter is organized as follows: Section 3.1 firstly justifies the characteristics and inherent uncertainties of prior knowledge, data, and data-driven methods; followed by the knowledge decomposition analysis, the Ladder of Knowledge-Integrated Machine Learning is proposed as an overall framework in an engineering context, along with three levels. For each level of the ladder, I discuss its application in specific engineering and design scenarios to verify its effectiveness: Section 3.2 investigates one of the representative methods in Level One, using prior modeling knowledge for feature engineering and data augmentation via simulations, utilizing data as a medium to embed the information from prior knowledge into the ML training process. Section 3.3 investigates the Level Two approach by disentanglement of ML model's objective functions and learning rules modification via domain knowledge, thereby enhancing its efficiency and adaptability in specific contexts. Section 3.4 details another Level Two approach, integrating prior inductive knowledge into the model organizational structure. Here, I focus on distinguishing the system's composable information from individual component characteristics through prior knowledge. By investigating an existing advanced method, component-based machine learning (CBML), an analogy to a "Lego-block" organizational method, I expand its applicability in general scenarios to prove its effectiveness in extrapolation and robustness against data sparseness. Section 3.5 advances to Level Three integration, where I re-examine the distinction between domain-specific prior knowledge and deductive knowledge, a type of methodology for discovering general explicit patterns as knowledge in an unsupervised behavior. I introduce causal inference for the first time in the domain of energy-efficient building design to demonstrate how identifying domain causal relationships guides designers in reasoning in frequent "what-if" questions encountered in engineering and design fields. Finally, Section 3.6 applies the KIML framework, along with all levels of methodological mindsets, to a scientific

exploration scenario by adapting in a promising sustainable energy system domain: Proton Exchange Membrane Water Electrolysis (PEMWE) development, demonstrating the across-domain adaptability of modeling methodologies described in this chapter. An organizational illustration of this chapter is presented in Figure 3.2.

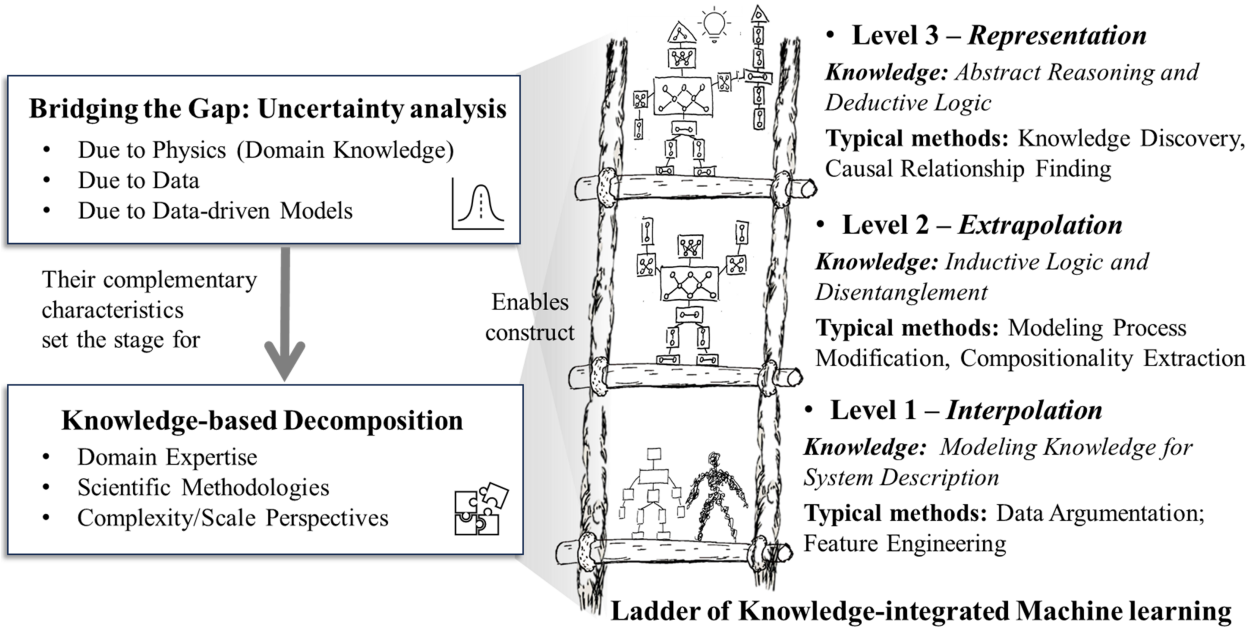

Figure 3.2: The Organization of Chapter Three

Outline

3.1 Overview Pathway 25
Bibliography 25
3.2 Level 1: Modeling Knowledge for Interpolation - Data Augmentation 26
Bibliography 27
3.3 Level 2: Disentanglement for Extrapolation - Modeling Process Modification . . 28
Bibliography 29
3.4 Level 2: Disentanglement for Extrapolation - Compositionality Extraction . . . 30
Bibliography 31
3.5 Level 3: Abstraction Reasoning for Representation - Causal Inference 32
Bibliography 34
3.6 AI for Science: Knowledge-Integrated Machine Learning Application 35
Bibliography 35

3.1 Overview Pathway

Outline

Bibliography 25

Bibliographic Information

- Chen, X., & Geyer, P. (2023). Pathway toward prior knowledge-integrated machine learning in engineering. In 18th International IBPSA conference and Exhibition, Building Simulation 2023, Shanghai, China. arXiv preprint arXiv:2307.06950.

Bibliography

3.2 Level 1: Modeling Knowledge for Interpolation - Data Augmentation

Different than the human learning process, ML training process carries a significant characteristic: despite its remarkable capability as a highly adaptable approximator, it often encounters data scarcity, a situation intensified by the bias/variance dilemma [66, 67]: A foundational issue requires ML models to delicately balance between simplicity (high bias) and complexity (high variance), especially in the face of limited or noisy data in real-world applications. This phenomenon prompts an essential inquiry for this sub-chapter: How can we turn this data reliance into a useful characteristic instead of viewing it as an obstacle to ML approaches for knowledge integration?

In the context of various design, engineering, and scientific discovery domains, we have accumulated extensive experience grounded in symbolism and logical expressions, allowing for deduction, summarization, and interpretation. Unfortunately, most of such domain-specific knowledge cannot be directly integrated into ML algorithms. Yet, we can leverage the data-dependent nature of ML by utilizing our direct modeling knowledge, the developed principles of first-principle simulations, and numerical simulation methods for feature engineering strategies and generating data tailored to specific contexts as data augmentation. By using data as the medium, this approach bridges the gap between symbolic, logic-based human knowledge into forms from which ML can process and learn; meanwhile, it provides a robust method for enhancing ML's pattern recognition capabilities with the depth of human interpretation in specific contexts.

Given its intuitive nature, I categorize this type of approach as Level One in our KIML framework and position it as a foundational strategy for knowledge integration. In real-world practice, these approaches could also mean employing first-principles methods as a means of data encoding mechanisms, transforming static characteristics and predictable patterns into simulated outcomes for ML training inputs. Furthermore, grounded on this level, I introduced the concept of Level-of-Information (LOI), proposing a flexible trade-off analysis between the detail of simulation modeling and the improvement in accuracy, significantly boosting ML's learning efficiency. For instance, in a building energy modeling scenario, leveraging LOI to integrate thermal dynamics simulations with historical building performance data significantly improves prediction accuracy without overwhelming computational resources and modeling efforts.

Outline

Bibliography 27

Bibliographic Information

- Chen, X., Guo, T., Kriegel, M., & Geyer, P. (2022). A hybrid-model forecasting framework for reducing the building energy performance gap. Advanced Engineering Informatics, 52, 101627.
- Chen, X., Guo, T., & Geyer, P. (2021). A hybrid-model forecasting framework for reducing the building energy performance gap. In 28th International Workshop on Intelligent Computing in Engineering, EG-ICE 2021. Berlin, Germany, 2021, special issue on Advanced Engineering Informatics.

Bibliography

3.3 Level 2: Disentanglement for Extrapolation - Modeling Process Modification

As we ascend to Level Two of the KIML framework, we move beyond using knowledge merely for data generation or processing. In the quest to mitigate ML's pattern learning dependency on data-hungry, augmenting datasets is one approach perspective. Yet, another equally potent strategy involves redesigning the learning algorithms with domain-specific context.

Most ML algorithms' composition contains two core elements: learning rules and objective functions. The former adjusts the model's parameters based on data inputs and outputs, while the latter quantifies the discrepancies between the model's predictions and actual targets, driving the model toward minimizing these differences. Both elements could be tailored and benefited by embedding domain-specific knowledge to fit specific tasks and align with the underlying principles or system behaviors known prior, hence transcending traditional data-driven paradigms. This integration becomes particularly potent when partial system descriptions (prior knowledge) are known analytically and combined with empirical data. The integration through ML modeling allows for a comprehensive model that encapsulates the full spectrum of system information. One promising research branch is Physics-Informed Neural Networks (PINNs) [68].

Drawing upon such inspiration and adapting to the landscape of design and engineering domains, I embedded first-principles simulations and symbolic equations in ML models not just as a means for data augmentation but as an internal part of ML objective functions. In practical terms, I designed a dynamic feedforward strategy tailored for building energy system control. This strategy allows both real-time feedback, historical behavior patterns, and the simulated future signal as inputs into ML algorithms, resulting in a 'gray-box' model. This model harmonizes data with explicit system knowledge derived from simulations, tailored to learn dynamic relationships between system states and external factors over rolling timeframes. This approach is set as an engineer-friendly alternative to reinforcement learning [22]. Furthermore, it offers general adaptability to match distinct systems dynamics learned from various buildings by modifying simulations, creating a dynamic virtual environment for training ML models to approach optimal control strategies. By doing so, the development represents a significant leap towards achieving models that can generalize across different scenarios with minimal data requirements. This 'gray-box' model is a testament to the potential of integrating physical reliability and enhanced learning efficiency of ML models that are interpretable and adaptable, capable of operating within dynamically changing system conditions.

Moreover, this level of integration promises new possibilities in ML, where models are no longer confined to data availability constraints but are empowered by the depth of human understanding. This holistic approach paves the way for creating models that are both robust in their predictive capabilities and grounded in the reality of their application domains, ultimately leading to sustainable and efficient solutions across various fields of engineering and science. Hence, I mark it as Level Two in KIML.

Outline

Bibliography 29

Bibliographic Information

- Chen X., Cai X., Kümpel A., Müller D., & Geyer P., (2022). A Dynamic Feedforward Control Strategy for Energy-efficient Building System Operation. In Passive and Low Energy Architecture, PLEA 2022, Santiago de Chile, Chile.

Bibliography

3.4 Level 2: Disentanglement for Extrapolation - Compositionality Extraction

Another dilemma ML encounters in practical applications, especially those based on deep learning, is the drawback of their "black-box" nature because of the model's purely mathematical, domain-independent optimization process, inflexibility in adapting to new contexts without retraining (catastrophic forgetting [69]), and poor performance in extrapolation [7]. Unlike interpolation, which makes predictions within the known data range, extrapolation requires extending the model's predictive capability to unseen and novel scenarios beyond the boundaries of the training data.

In contrast, the human learning process exhibits remarkably flexible attributes that ML models often struggle to replicate. At this stage, I exemplified the concept of compositionality [70], a more first-principles mindset mostly discussed in language research that transcends domain-specific knowledge [71, 72]. As a mindset of an infinite number of representations constructed from a finite set of primitives, such flexibility characteristics are valuable for model organization. This general idea, derived from deduction [73, 21], allows for a broader application of models to adapt and perform across various design and engineering scenarios: In the context of building engineering, a house's components and their relations can be shared and re-used from existing concepts, such as windows, beams, roofs, etc., and recomposed the understanding to form a coherent solution [74]. This compositionality represents a more universal modeling approach, allowing for reasonable predictions across scenarios.

In many domains, fundamental laws, reductionist mindset, and system construction rules remain constant. The prior knowledge here aims to disentangle the fixed combination of information carried by basic components and those invariant rules or grammars from data. By training basic components and embedding rules/grammar into the model organizational process, we enable a shift from models that can only interpolate within fixed scenarios to those capable of reliable modeling and prediction across diverse contexts. For an intuitive understanding, it is analogous to Lego blocks: instead of modeling each assembled Lego model separately, we model the basic Lego blocks, separating the blocks and their assembly "syntax". I believe that this mindset is the key to achieving model generalization across scenarios within specific problems.

Furthermore, this modeling form, which decomposes components and organizational methods, can reduce the model's dependency on data constraint from a singular source, thereby enhancing the model's flexibility and robustness to effectively process, extract, and represent system information. I presented a case study on building energy performance evaluation by inheriting the component-based machine learning [74] approach. I proved that this approach yields several benefits that don't exist in monolithic models: enhanced interpretability, performance robustness, and resistance against incomplete, small, and sparse data inputs.

Outline

Bibliography 31

Bibliographic Information

- Chen, X., Singh, M.M., & Geyer, P. (2024). Utilizing domain knowledge: robust machine learning for building energy performance prediction with small, inconsistent datasets. Knowledge-Based Systems, 294, 111774.
- Chen, X., Singh, M.M., & Geyer, P. (2021). Component-based machine learning for predicting representative time-series of energy performance in building design. In 28th International Workshop on Intelligent Computing in Engineering, EG-ICE 2021. Berlin, Germany.

Bibliography

3.5 Level 3: Abstraction Reasoning for Representation - Causal Inference

Before ascending to Level Three knowledge, here is a short recall of the core efforts conducted in the previous two levels of KIML:

- Level One focuses on enriching or processing data inputs via domain knowledge, thereby boosting the model's interpolative capabilities.
- Level Two advances this by integrating domain knowledge directly into the modeling process, facilitating a more interpretable, flexible modeling endowed with procedural logic, composability, and extrapolative capacities.

In both contexts, "domain knowledge" refers to first-order knowledge, or "learned patterns," which directly points to the domain-specific insights addressed and derived from human induction. However, in reality, such insights could potentially carry the risk of cognitive biases [75] from individual limitations and unconscious errors in the inherent way of thinking, leading to faulty misjudgments or sub-optimal solutions. From another perspective, it also reflects that current ML approaches fundamentally lack a process akin to human reasoning or inductive biases directly learned from the data. Such capability to reason, including the ability to conceive 'what-if' scenarios, is essential for decision-making in design, engineering, and scientific discovery processes. This reasoning capability, which involves pattern recognition, concept modeling, prediction, and explanation, lies at the core of human intelligence and is fundamentally viewed as a constructive activity [18]. When humans or machines make inferences that go far beyond the data, the loop between knowledge discovery and application must close to make up the difference. So the key is: how do we not solely rely on the previously inducted knowledge by humans but also equip the machine with the induction ability to discover the knowledge from data?

So, at this level, I investigate the human capabilities of inductions used for knowledge discovery and modeling. Here, inductive biases [49, 76] based on mathematical logic play a pivotal role: they serve as a logical formula that logically encompasses the hypotheses predisposing models towards specific solutions when combined with data. In this context, the conceptual representations are abstracted and generalized from experience to enhance efficiency in problem-solving. For example, the feature extraction process of convolutional neural networks (CNNs) is inspired by the concept of human retinal process [77, 78], as well as the processing of sequential data by recurrent neural networks (RNNs) [79]. This perception, abstraction, inspiration, and model construction process reveals the significance of inductive biases that facilitate analogy and reasoning capabilities in humans. At its core, it represents the shift from "pattern learning" from data to "model building" [18]. Integrating such fundamental processes and principles for knowledge abstraction into the learning algorithm could pave the path for MLs to discover knowledge directly from data; one example of such principles would be Occam's Razor [80], the representations gained through inductive biases often provide analogously reusable and interactive content, but the principle itself is general and not attach to specific domain knowledge. In this sub-chapter, I exemplify another vital concept based on inductive biases: causal inference.

Incorporating causal inference [81, 45] into ML represents a fundamental leap, enabling models to address "what-if" questions critical for assisting in decision-making scenarios across domains. Causality grounded in the statistical identification of cause-and-effect relationships - an asymmetrical correlation between factors - offers a framework for discovering causal skeletons and

quantifying the effects of interventions within datasets. It allows for a data-driven construction of models essential for planning, optimizing, and understanding complex systems. In practice, distinguishing between correlation and causation is not just a philosophical viewpoint but a crucial methodological consideration (e.g., Simpson paradox [82]). This perspective goes beyond mere correlation, aiming to understand the mechanisms driving the observed data, considering how changes in one variable lead to changes in another, which is essential for tasks involving prediction under intervention, counterfactual reasoning, and understanding complex systems. This shift in ML approaches aligns them more closely with the human cognitive process; as we transition from combining intuition-driven decisions with data-driven methodologies, it's imperative to harness the raw potential of data itself rather than rely solely on potentially biased or incorrect prior knowledge.

Outline

Bibliography 34

Bibliographic Information

- Chen, X., Abualdenien, J., Singh, M. M., Borrmann, A., & Geyer, P. (2022). Introducing causal inference in the energy-efficient building design process. Energy and Buildings, 277, 112583.

Bibliography

3.6 AI for Science: Knowledge-Integrated Machine Learning Application

In previous sub-chapters, our implementation cases are mostly focused on the design and building engineering domain. To prove its universal applicability, I implemented the complete KIML framework with methodologies across all levels in a new scientific discovery domain - Proton Exchange Membrane Water Electrolysis (PEMWE) development.

Outline

Bibliography 35

Bibliographic Information

- Chen, X., Rex, A., Woelke, J., Eckert, C., Bensmann, B., Hanke-Rauschenbach, R., & Geyer, P. (2024) Machine Learning in Proton Exchange Membrane Water Electrolysis–a Knowledge-Integrated Framework. Applied Energy, under review. Available at SSRN 4743024.

Bibliography

4. Communication Alignment: Advanced Human-Computer Interaction

"Fake it, until you make it."

Along the human and AI cognition research journey, this adage takes on profound significance, laying the groundwork mindset for approaching understanding intelligence, mechanisms of decision-making, and free will. One of the perspectives to view this topic is via the computational irreducibility [83], a concept suggesting that the realization illuminates why individuals often perceive themselves as agents of free will: the mechanics of their decisions, even if deterministic, cannot be preemptively known without engaging in a computational endeavor as complex as the decision-making process itself [84]. This perspective provides an essential foundation for this chapter because: 1. It demonstrates the irreplaceability of the decision-making role; and 2. Formalizes the informed decision-making task into an optimization challenge. In other words, it means that the decision-making process cannot be simplified beyond its fundamental complexity but exists an informational exchange loss minimum for each step to conclude. In this context, decision-making support systems emerge as pivotal roles that aim not to circumvent the essential complexity of these processes but to improve human and artificial decision-making capabilities. Essentially, it reveals the purpose definition of our machine assistance for decision-making support: acceleration of the computation process to enhance the available information utility and approach the minimum simplified computation.

Therefore, in this chapter, the focus is shifted from enhancing the performance of ML modeling processes in specific tasks to refocusing on the decision-making process in improving human-computer interaction patterns. Building upon the foundation of the information flow established in chapter two, I primitively explore two potential directions for data-driven models/machine assistance in interaction: 1. **enhancing the information utility via symbiosis**, investigating its impact on our decision-making process, and 2. **expanding the information bandwidth of interaction flow** to minimize the information loss. An illustrative presentation is given in Figure 4.1.

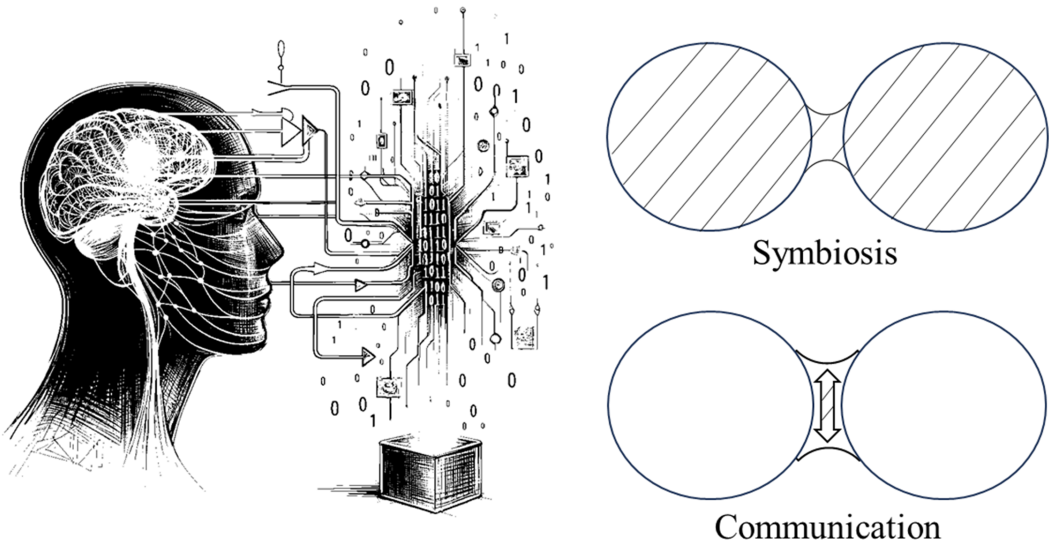

Figure 4.1: Symbiosis for Enhanced Information Utility and Expanding the Interaction Bandwidth via Data-Driven Methods

For the first direction, I present an engineering example of overcoming prior knowledge biases through machine-assisted data-driven modeling, specifically through a causal model mentioned in the last chapter, Level Three of KIML. This step is crucial for machine-assisted decision-making since previous chapters have not elaborated on the limitations of individuals' biased prior knowledge and cognitions [85], which could lead to real-world loss and risks. By raising a decision-making case in the engineering domain, I propose a symbiotic construct among users, data, causal models, data-driven ML models, and first-principles models built on prior knowledge. This construct systematically shows how these elements interact to exchange information they process and how they coexist in a symbiosis that pinpoints themselves within this framework and overcomes bias to make informed decisions beyond personal limitations.

The second direction focuses on exploring advanced information representation, transformation, and its utilization via data-driven approaches. During the decision-making process, information flows bidirectionally between humans and the external environment; they are closely interconnected and communicate in parallel. This could link to the analogy to cybernetics, where current machine assistance could be part of our constructed entity and extend our individual cognition and interaction efficiency. Essentially, engineering, design, or human communication is a game against noise between two entities exchanging information to convey ideas from one form to another [56]. Traditional methods typically require effort in communication: from the human and user perspective, learning languages, programming, and software operation for design interaction. Information transmission between different individuals often involves information transformation and loss. From the machine's perspective, converting human design and language into learnable data also demands significant engineering effort and model decoding, which makes it less efficient and direct. Thus, the second sub-chapter explores how to use data-driven methods to improve the efficiency of human-computer interaction. I examined the efficacy of various neural network (NN) models in interpreting mental constructs via electroencephalogram (EEG) signals. In this case, ML methods not only benefit from capturing information from the external system but also contribute to expanding communication bandwidth with humans by recognizing information from extra signals. In the design scenario, such new patterns would efficiently assist designers in communicating and interacting with the design work: their personal preferences, less expressible feelings, and subconscious perceptions

can be captured, represented, and transferred more seamlessly via the brain's electrical signals without explicitly formalizing in languages or actions.

The work presented in this chapter, though preliminary, demonstrates the potential for advanced human-computer interaction patterns. Together with the machine assistance framework and knowledge-integrated machine learning concepts from previous chapters, it illustrates the overarching conceptual idea of augmented intelligence for decision-making. The objective in investigating communication patterns is to provide possibilities, different from the previous two chapters, to seamlessly comprehend human intuitive, implicit yet important factors, such as intentions, preferences, and ethical standards. I acknowledge that there is still much ground to explore further, but this communication alignment chapter is crucial to provide a comprehensive view of this dissertation research, especially from the collaborative human-centric perspective.

Eventually, both directions combine together as the exemplar of advanced human-computer interaction patterns discussed in this chapter. Along with the machine assistance framework and knowledge-integrated machine learning paradigms from previous chapters, it presents the complete image of the framework for augmented intelligence in decision-making in this dissertation.

Outline

4.1 Symboisis between Users and Machines 40
Bibliography 40
4.2 Data-driven Approaches for User Implicit Signal Decoding 41
Bibliography 41

4.1 Symboisis between Users and Machines

Outline

Bibliography 40

Bibliographic Information

- Chen, X., Sun, R., Saluz, U., Schiavon, S., & Geyer, P. (2023). Using causal inference to avoid fallouts in data-driven parametric analysis: A case study in the architecture, engineering, and construction industry. Developments in the Built Environment, 100296.

Bibliography

4.2 Data-driven Approaches for User Implicit Signal Decoding

Outline

Bibliography 41

Bibliographic Information

- Chen, X., Teng, X., Chen, H., Pan, Y., & Geyer, P. (2024). Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. Biomedical Signal Processing and Control, 87, 105475.

Bibliography

5. Conclusion

Outline

5.1 Summary of Contributions 42
5.2 Remaining Challenges and Future Directions 43

5.1 Summary of Contributions

If one keyword is used to describe the main effort of this dissertation, it would be "Alignment."

This alignment effort lies throughout the dissertation, where the proposed machine assistance for extending human capabilities with knowledge integration paradigms reframes the relationship between humans and data-driven methods, suggesting a symbiotic integration in which AI systems and human thought processes augment each other. This concept is inherent in extended mind theory in cognitive science and philosophy of mind [86], which posits that objects in our environment can become extensions of our minds, contributing to cognitive processes such as memory, reasoning, and decision-making.

The focus of this dissertation is not only to harness domain-specific knowledge but also to explore how humans perceive and interpret the world for machines to align with. It can be well described by the term "**Gedankengang**" in German, literally meaning "thought path," which emphasizes the process of thinking, reasoning, or the progression of thoughts. Such need is especially urgent in an era with the rapid development of advanced AI in multimodal models (MMM) and large language models (LLMs), which has raised discussions about the potential risks of artificial general intelligence (AGI) for humanity [87]. From this perspective, the dissertation herein seeks to align AI technologies with human-centric "Gedankendgang", advocating for a collaborative effort to harness collective intelligence resources for social advancement rather than an oppositional relationship with technology. Such alignment, as an imitation game, is the critical point to ensure that AI integration into our lives and decision-making processes enriches, rather than undermines, human capabilities and societal progress autonomously. Three essential elements of this framework, emphasizing the importance of alignment research, are detailed as follows:

1. **Constructing a Cognitive Entity**: By incorporating machine assistance into part of our decision-making process, I acknowledge the role of external information processing methods in expanding our knowledge boundaries in a cybernetic concept. This constructed entity becomes a repository of extended insights, reinforced analytical prowess, and predictive outcomes, all of which are accessible with the seamless integration of AI systems and a deep understanding of human cognition that can mimic, complement, and enhance our cognitive capabilities.

2. **Prior Knowledge Integration**: Exemplified in the engineering, design, and scientific discovery scenario, I proposed a methodological paradigm to systematically incorporate human-centric knowledge into ML models. This paradigm reconciles the computational power, data processing capabilities, and pattern recognition strengths of AI, along with direct knowledge of model description, deductive logic with abstract principles, and inductive reasoning abilities in a hierarchical and aligned structure. The ultimate goal is to develop ML methods that can synthesize vast amounts of data, draw on extensive knowledge bases, and apply complex analytical models; all of this contributes to more accurate, effective, trustworthy methodological solutions, which is viewed as an extension of cognition to achieve well-informed and sophisticated decision-making.

3. **Advancement in Human-Computer Interaction**: Another necessary perspective worth investigating for completing the alignment topic in this framework is to elevate the naturalness of human-machine collaboration by improving HCI efficacy. Based on the machines' knowledge discovery potential elaborated in the methodological alignment chapter, I emphasized a symbiotic relationship between humans and machines, where each contributes to the growth and evolution of the other. Sequentially, as humans extend their cognitive capabilities through machine assistance, AI systems, in turn, become more adaptive to human thought processes, preferences, and decision-making styles that guarantee trustfulness. This entails leveraging multimodal data beyond our traditional expression methods, such as language or actions, and refining communication pathways within a cognitive construct to advancements in both entities simultaneously.

5.2 Remaining Challenges and Future Directions

This dissertation establishes a machine assistance framework to try to align with the human cognition process toward augmented intelligence for its utility in assisting engineering, design, and scientific discovery scenarios. It paves the way for data-driven methods to align with the way we interact, think, and hence, as the assistance. Yet, this dissertation merely sets the initial step for such alignment research. The path forward is marked by several key challenges and potential directions for further exploration:

- **Exploration of Fundamental Elements with Combinations**: This dissertation took a step toward aligning ML frameworks with the complexity of human intelligence. At the end scope of the knowledge-integration paradigm, one key question that arises is: *How can we incorporate the elements that make human learning flexible and efficient into machines to achieve advanced intelligence for collaboration?* The investigation of the interaction among these elements intends to lead to the creation of new ideas by combining familiar concepts, a process that reflects the essence of human intelligence with its world models and prior experiences. These characteristics allow for quick understanding, knowledge transfer, and learning from a few examples. Looking beyond this dissertation, it's clear that exploring additional foundational elements and their interactions is crucial. Potential areas for further research include scaling [88], recursion [63], emergence [19] mechanisms, etc. For instance, in understanding how to blend compositional and recursive learning to spark new concepts, causality provides the glue that gives them coherence and purpose and unlocks creative design [89, 90] - prerequisites for unexpected combinations of familiar concepts or ideas. In simpler terms, a promising potential direction lies in exploring how to make machines learn and think more like humans by studying the fundamental building blocks of human intelligence and creativity for alignment.

- **Alignment Beyond Knowledge**: While the framework proposes methodological alignment with human prior knowledge, aligning machines with societal ethics, values, and morals presents an open field requiring urgent exploration as well. This broader alignment challenges us to develop AI technologies that not only replicate human thought processes but also embody the ethical standards and moral principles that define human society [91]. Domain knowledge, while essential information that directly addresses enhancement in engineering, design, and scientific discovery problem-solving, represents just one perspective of the vast spectrum of human thoughts. The reason lies explicitly: it is crucial to ensure that AI systems are developed with a comprehensive understanding of human considerations and human-centric, while domain knowledge only covers a narrow perspective. For example, in a design scenario of architecture and urban planning, the core mission is not only to create functional, aesthetically appealing, and sustainable spaces but also to reconcile the societal impact and proposition that stands for our value - a blend of art and inspiration, science and methodology. In this context, it becomes tricky that when creative ideas are developed by machine assistance, as mentioned in the previous point, would that be a reinforcement or sabotage of user autonomy? Such factors integrated into human-centric machine assistance remain a gap in the current exploration [92]. However, the advanced interaction pattern elaborated in chapter four of this dissertation could, I believe, be one potential path to intervene with and align the machine with our reactions, which the reactions reflect our implicit standards and values and act as a new way of RLHF. Nevertheless, we must tread this path with a critical eye, reflecting not only on what our technologies can do but also on what they should do, ensuring that this integration of human and machine serves to enhance, rather than diminish, our humanity.

- **Integration with Neuroscience and Cognitive Science**: If we focus on the mechanism developed for ML and AI, the principles underlying these technologies have raised intense debate due to their lack of proof in biological systems (e.g., whether backpropagation, one fundamental learning rule for training artificial neural networks, exists in biological systems [93]). This disparity leads to the foundational difference in how machines and humans recognize the world. In this dissertation, ML works inspired by biological neural activation, such as spiking neural networks [94, 95], haven't been thoroughly investigated. Another example is the catastrophic forgetting phenomenon widely existing in connectionist ML models [69]. It contrasts sharply with natural cognitive systems where new learning rarely completely disrupts or erases previously acquired information; current ML models do not have such flexibility due to their fixed weighted structure once the training phase is done. In this context, constructing machine assistance that ensures consistent learning progress necessitates inputs and alignment from cognitive science and neuroscience inductive knowledge, such as sleeping mechanism [96]. These inputs reveal potential paths with crucial insights for developing large AI systems that could exhibit resilience against model degradation and exhausted retaining effort [97], thus maintaining the integrity of accumulated knowledge over time. This approach underscores the necessity for a multidisciplinary strategy observed/inducted in natural cognitive systems for further development.

- **Top-Down Encoding with Large Models**: The emergence of LLMs brings a new challenge for knowledge integration, as it is less feasible to fine-tune or manually craft these large-scale models. I observed some upcoming investigations proposing a top-down approach to encode knowledge and examine machine behavior, such as honesty [98], which is essential for machine assistance to gain trust and interact in human-like ways. Combined with the work conducted in this dissertation, mostly bottom-up, I refer to them

as an analogy of Platonic backhand and forehand [99]. This highlights the challenge and opportunity for embedding complex, prior knowledge into large-scale ML models. It gives a new perspective of not overemphasizing one side of methods and advocates for a balanced approach combining cybernetic thinking with a high-level overview. In other words, this critique underscores the necessity of top-down approaches to encode extensive prior knowledge, enabling AI systems to learn and solve new tasks swiftly. It calls for synthesizing innate abilities and experiential learning, leading to a comprehensive 'world mind' model [100, 101]. This balanced, integrated approach is crucial for advancing AI's understanding and interaction capabilities in a human-like manner, reflecting holistic, context-aware architectural and AI practices.

Addressing these challenges requires a multidisciplinary approach that spans machine learning, cognitive science, neuroscience, ethics, and beyond. I foresee new potential for creating machines that not only approach human thought but also embody our propositions, paving the way for a future where AI and humans coexist and collaborate symbiotically. To conclude, this dissertation contributes to the evolving potential of an intelligence augmentation system that enhances decision-making processes, embeds human knowledge, and redefines AI's role in extending and enriching our cognition. I want to propose the final question in this dissertation and end it with my answer:

"What would be the purpose of designing such a machine assistance?"
"To know you."

Bibliography

[1] H.G. Wells. *The Outline of History*. 1921.

[2] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[3] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[4] OpenAI. Openai - chatgpt [large language model], 2022.

[5] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, et al. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, page 100017, 2023.

[6] The Economist. Huge "foundation models" are turbo-charging ai progress, Jun 2022.

[7] Uri Hasson, Samuel A Nastase, and Ariel Goldstein. Direct fit to nature: an evolutionary perspective on biological and artificial neural networks. *Neuron*, 105(3):416–434, 2020.

[8] Timothy P Lillicrap, Daniel Cownden, Douglas B Tweed, and Colin J Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1):13276, 2016.

[9] Daniel Kahneman. *Thinking, fast and slow*. macmillan, 2011.

[10] Dedre Gentner and Keith J Holyoak. Reasoning and learning by analogy: Introduction. *American psychologist*, 52(1):32, 1997.

[11] Marvin L Minsky. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI magazine*, 12(2):34–34, 1991.

[12] Esther N Goody. *Social intelligence and interaction: Expressions and implications of the social bias in human intelligence*. Cambridge University Press, 1995.

[13] Berndt Brehmer. Dynamic decision making: Human control of complex systems. *Acta psychologica*, 81(3):211–241, 1992.

[14] Yoshua Bengio, Yann Lecun, and Geoffrey Hinton. Deep learning for ai. *Communications of the ACM*, 64(7):58–65, 2021.

[15] Neil C Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F Manso. The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*, 2020.

[16] Mike Oaksford and Nick Chater. The probabilistic approach to human reasoning. *Trends in cognitive sciences*, 5(8):349–357, 2001.

[17] Yi Ma, Doris Tsao, and Heung-Yeung Shum. On the principles of parsimony and self-consistency for the emergence of intelligence. *Frontiers of Information Technology & Electronic Engineering*, 23(9):1298–1323, 2022.

[18] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253, 2017.

[19] Philip W Anderson. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.

[20] Yoshua Bengio. The consciousness prior. *arXiv preprint arXiv:1709.08568*, 2017.

[21] Kees Dorst. The core of 'design thinking' and its application. *Design studies*, 32(6):521–532, 2011.

[22] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

[23] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

[24] David Foster. *Generative deep learning.* " O'Reilly Media, Inc.", 2022.

[25] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112, 2009.

[26] Hamed S Alavi, Elizabeth F Churchill, Mikael Wiberg, Denis Lalanne, Peter Dalsgaard, Ava Fatah gen Schieck, and Yvonne Rogers. Introduction to human-building interaction (hbi) interfacing hci with architecture and urban design, 2019.

[27] Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204, 2013.

[28] Beren Millidge, Anil Seth, and Christopher L Buckley. Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979*, 2021.

[29] Christoph Teufel and Paul C Fletcher. Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, 21(4):231–242, 2020.

[30] Robert H Bonczek, Clyde W Holsapple, and Andrew B Whinston. *Foundations of decision support systems.* Academic Press, 2014.

[31] Martin Aruldoss, T Miranda Lakshmi, and V Prasanna Venkatesan. A survey on multi criteria decision making methods and its applications. *American Journal of Information Systems*, 1(1):31–43, 2013.

[32] José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

[33] Joy A Thomas. Elements of information theory, 1991.

[34] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[35] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.

[36] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 2015.

[37] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8):832, 2019.

[38] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[39] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

[40] Judea Pearl. Radical empiricism and machine learning research, 2021.

[41] Alexander G Ramm. *Inverse problems: mathematical and analytical techniques with applications to engineering*. Springer Science & Business Media, 2005.

[42] Randall Balestriero, Jerome Pesenti, and Yann LeCun. Learning in high dimension always amounts to extrapolation. *arXiv preprint arXiv:2110.09485*, 2021.

[43] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6):141–142, 2012.

[44] Ramesh S Patil, Peter Szolovits, and William B Schwartz. Causal understanding of patient illness in medical diagnosis. In *Computer-Assisted Medical Decision Making*, pages 272–292. Springer, 1981.

[45] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

[46] Liuyi Yao, Zhixuan Chu, Sheng Li, Yaliang Li, Jing Gao, and Aidong Zhang. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5):1–46, 2021.

[47] John Aldrich. Correlations genuine and spurious in pearson and yule. *Statistical science*, pages 364–376, 1995.

[48] Nikolaus Kriegeskorte and Pamela K Douglas. Cognitive computational neuroscience. *Nature neuroscience*, 21(9):1148–1160, 2018.

[49] Thomas L Griffiths, Nick Chater, Charles Kemp, Amy Perfors, and Joshua B Tenenbaum. Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, 14(8):357–364, 2010.

[50] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[51] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

[52] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018.

[53] Victor Riley. A general model of mixed-initiative human-machine systems. In *Proceedings of the Human Factors Society Annual Meeting*, volume 33, pages 124–128. Sage Publications Sage CA: Los Angeles, CA, 1989.

[54] Christos Emmanouilidis, Petros Pistofidis, Luka Bertoncelj, Vassilis Katsouros, Apostolos Fournaris, Christos Koulamas, and Cristobal Ruiz-Carcel. Enabling the human in the loop: Linked data and knowledge in industrial cyber-physical systems. *Annual reviews in control*, 47:249–265, 2019.

[55] Harley Oliff, Ying Liu, Maneesh Kumar, Michael Williams, and Michael Ryan. Reinforcement learning for facilitating human-robot-interaction in manufacturing. *Journal of Manufacturing Systems*, 56:326–340, 2020.

[56] Norbert Wiener. *The human use of human beings: Cybernetics and society*. Number 320. Da capo press, 1988.

[57] N Katherine Hayles. How we became posthuman: Virtual bodies in cybernetics, literature, and informatics, 2000.

[58] Christopher D Wickens, Sallie E Gordon, Yili Liu, and J Lee. *An introduction to human factors engineering*, volume 2. Pearson Prentice Hall Upper Saddle River, NJ, 2004.

[59] Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow. Behavior, purpose and teleology. *Philosophy of science*, 10(1):18–24, 1943.

[60] David LaBerge and S Jay Samuels. Toward a theory of automatic information processing in reading. *Cognitive psychology*, 6(2):293–323, 1974.

[61] David E Rumelhart, Geoffrey E Hinton, James L McClelland, et al. A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(45-76):26, 1986.

[62] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[63] Melanie Mitchell. *Complexity: A guided tour*. Oxford university press, 2009.

[64] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[65] JEAN PlACET. Piaget's theory. *Car · michael's Manual of Child Psychology (3rd This article is intended solely for the personal use of the individual user and is not to be disseminated broadly*, 1970.

[66] Thomas G Dietterich and Eun Bae Kong. Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. 1995.

[67] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.

[68] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[69] Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999.

[70] Barbara Tversky and Kathleen Hemenway. Objects, parts, and categories. *Journal of experimental psychology: General*, 113(2):169, 1984.

[71] Theo MV Janssen and Barbara H Partee. Compositionality. In *Handbook of logic and language*, pages 417–473. Elsevier, 1997.

[72] Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. Compositionality decomposed: How do neural networks generalise? *Journal of Artificial Intelligence Research*, 67:757–795, 2020.

[73] Irving Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

[74] Philipp Geyer and Sundaravelpandian Singaravel. Component-based machine learning for performance prediction in building design. *Applied energy*, 228:1439–1453, 2018.

[75] TK Das and Bing-Sheng Teng. Cognitive biases and strategic decision processes: An integrative perspective. *Journal of management studies*, 36(6):757–778, 1999.

[76] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.

[77] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.

[78] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[79] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. Ieee, 2013.

[80] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Occam's razor. *Information processing letters*, 24(6):377–380, 1987.

[81] Judea Pearl. Causal inference in statistics: An overview. 2009.

[82] Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2):3, 2000.

[83] Stephen Wolfram et al. *A new kind of science*, volume 5. Wolfram media Champaign, IL, 2002.

[84] Marius Krumm and Markus P Müller. Free agency and determinism: Is there a sensible definition of computational sourcehood? *Entropy*, 25(6):903, 2023.

[85] Martin Hilbert. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological bulletin*, 138(2):211, 2012.

[86] Andy Clark and David J Chalmers. The extended mind. 2010.

[87] Scott McLean, Gemma JM Read, Jason Thompson, Chris Baber, Neville A Stanton, and Paul M Salmon. The risks associated with artificial general intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5):649–663, 2023.

[88] Geoffrey West. *Scale: The universal laws of life, growth, and death in organisms, cities, and companies*. Penguin, 2018.

[89] Margaret A Boden. Creativity and artificial intelligence. *Artificial intelligence*, 103(1-2):347–356, 1998.

[90] John Andrew Rehling. *Letter spirit (part two): Modeling creativity in a visual domain*. Indiana University, 2001.

[91] Peter H Ditto, David A Pizarro, and David Tannenbaum. Motivated moral reasoning. *Psychology of learning and motivation*, 50:307–338, 2009.

[92] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Ingredients of intelligence: From classic debates to an engineering roadmap. *Behavioral and Brain Sciences*, 40, 2017.

[93] Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

[94] Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, 2019.

[95] Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997.

[96] Timothy Tadros, Giri P Krishnan, Ramyaa Ramyaa, and Maxim Bazhenov. Sleep-like unsupervised replay reduces catastrophic forgetting in artificial neural networks. *Nature Communications*, 13(1):7742, 2022.

[97] Changmao Li and Jeffrey Flanigan. Task contamination: Language models may not be few-shot anymore. *arXiv preprint arXiv:2312.16337*, 2023.

[98] Andy Zou, Long Phan, Sarah Chen, James Campbell, Phillip Guo, Richard Ren, Alexander Pan, Xuwang Yin, Mantas Mazeika, Ann-Kathrin Dombrowski, et al. Representation engineering: A top-down approach to ai transparency. *arXiv preprint arXiv:2310.01405*, 2023.

[99] Camilo Andrés Cifuentes Quin. The platonic forehand and backhand of cybernetic architecture. *Leonardo*, 52(5):429–434, 2019.

[100] David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

[101] James F Allen and Johannes A Koomen. Planning using a temporal world model. In *Proceedings of the Eighth international joint conference on Artificial intelligence-Volume 2*, pages 741–747, 1983.

6. Appendix

In this dissertation, illustrations in Figures 2.1, 3.1, and 4.1 are generated with the assistance of DALL · E 2 (2024.02 Version)[1]. The following is a list of the original image prompts and their corresponding generated images.

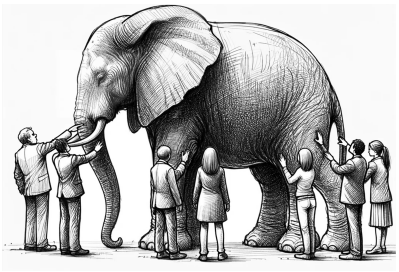

Figure 6.1: Generated illustration used in Figure 2.1

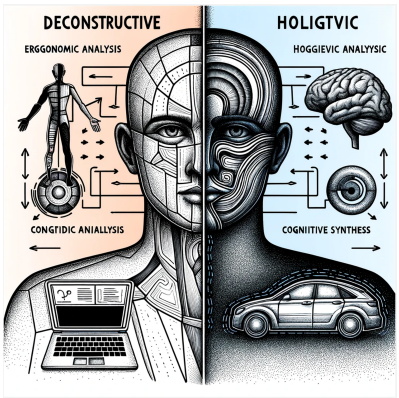

Figure 6.2: Generated illustration used in Figure 3.1

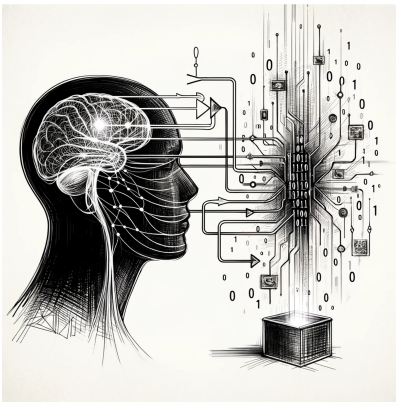

Figure 6.3: Generated illustration used in Figure 4.1

[1]https://help.openai.com/en/articles/6640875-how-should-i-credit-dall-e-in-my-work

Prompts

Figure 6.1 prompt: *Sketch a group of people standing around an elephant, each person is touching a different part of the elephant's body. The elephant is shown in side view. The people are positioned around the elephant so that one person is touching the trunk, another the ear, one person is at the side touching the elephant's belly, another at the tail, and one by the tusk. The individuals are in various poses that indicate curiosity and exploration. The scene captures the idea of exploring different perspectives and interpretations, reminiscent of the parable of the blind men and the elephant.*

Figure 6.2 prompt: *Create an academic sketch illustrating the concept of integrating deconstructive and holistic approaches in engineering. On the left side, depict a human figure divided into segments with arrows pointing to an outline of a car, signifying the ergonomic analysis of the human interaction with a vehicle. Label this segment 'Ergonomic Analysis'. On the right side, illustrate a human head with a brain, connected to a laptop representing an information assistant, symbolizing cognitive engagement and decision-making. Label this segment 'Cognitive Synthesis'. Between both sides, show a gradient or spectrum to symbolize the integration of both approaches. The style should be clean and suitable for an academic publication.*

Figure 6.3 prompt: *A clean, black and white sketch that illustrates the concept of Information Flow and Human-Computer Interaction. The image should show a human figure and a computer system with a bidirectional flow of information between them. The flow of information should be depicted as streams of binary or abstract symbols, symbolizing the exchange and enhancement of data. This represents data processing and decision-making support systems in action. The human figure should be actively engaged with the computer system, possibly with a neural interface, to depict the interconnectedness and dynamic interaction in the decision-making process.*