Synthesis Lectures on Visualization

SYNTHESIS
COLLECTION OF TECHNOLOGY

Shixia Liu · Weikai Yang · Junpeng Wang · Jun Yuan

# Visualization for Artificial Intelligence

Springer

# Synthesis Lectures on Visualization

This series publishes on topics pertaining to scientific visualization, information visualization, and visual analytics. Potential topics include, but are not limited to scientific, information, and medical visualization; visual analytics, applications of visualization and analysis; mathematical foundations of visualization and analytics; interaction, cognition, and perception related to visualization and analytics; data integration, analysis, and visualization; new applications of visualization and analysis; knowledge discovery management and representation; systems, and evaluation; distributed and collaborative visualization and analysis.

Shixia Liu · Weikai Yang · Junpeng Wang · Jun Yuan

# Visualization for Artificial Intelligence

Springer

Shixia Liu
School of Software
Tsinghua University
Beijing, China

Junpeng Wang
Visa Research
Foster City, CA, USA

Weikai Yang
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Jun Yuan
School of Software
Tsinghua University
Beijing, China

ISSN 2159-516X ISSN 2159-5178 (electronic)
Synthesis Lectures on Visualization
ISBN 978-3-031-75339-8 ISBN 978-3-031-75340-4 (eBook)
https://doi.org/10.1007/978-3-031-75340-4



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Acknowledgements

We are deeply grateful for the contributions and support from a wide range of individuals.

First of all, we would like to express our sincere thanks to Prof. Xiting Wang, Prof. Changjian Chen, and Prof. Jing Wu for their insightful discussions and invaluable feedback throughout the development of this book. Their expertise and constructive suggestions have been pivotal in shaping the final manuscript. We also appreciate the thoughtful suggestions provided by the anonymous reviewers, which greatly improved both the content and presentation of this book. Special thanks go to Duan Li for his insightful suggestions and efforts in improving the figures.

We are grateful to Prof. Niklas Elmqvist and Prof. David Ebert for offering us the opportunity to publish this book in their synthesis lecture series on visualization. Niklas' continuous encouragement, feedback, and support have been crucial in bringing this book to fruition. We also extend our appreciation to Christine Kiilerich, Boopalan Renu, and Jayanthi Narayanaswamy for their excellent work in managing and editing this book.

Our research was supported by the National Natural Science Foundation of China under grants U21A20469 and 61936002, whose financial support was essential for this work.

Finally, we would like to acknowledge the unwavering encouragement, patience, and support of our friends and families. Their belief in this book has been a constant source of strength.

November 2024

Shixia Liu
Weikai Yang
Junpeng Wang
Jun Yuan

# Contents


# Introduction 1

Artificial intelligence (AI) refers to the capability of a computer to perform cognitive tasks that are typically associated with human intelligence such as learning, reasoning, and problem-solving. Due to the success of machine learning, especially foundation models (e.g., ChatGPT), the field of AI is currently experiencing rapid growth and has the potential to revolutionize many aspects of our lives, from health care to finance, criminal justice to education. With the increasing deployment of AI systems in these fields, ensuring their ability to generalize to new, unseen data has become crucial for optimal performance and practical utility in real-world applications. Furthermore, it is crucial that the decisions made by these systems are transparent and explainable for greater accountability and trustworthiness.

## 1.1 Generalization and Interpretability of AI

Machine Learning serves as the foundational backbone and a critical component in the field of AI. It centers on the development and refinement of models that learn from large amounts of data, identify patterns, predict outcomes, and make decisions based on input data. Generalization, the ability of an ML model to perform well on new and previously unseen data, is crucial for high performance and real-world applicability. This ability enables an AI system to perform robustly in the presence of changes in the data distribution. When a model lacks strong generalization capabilities, it tends to overfit the training data by capturing the spurious correlations and statistical noise of the training data instead of the underlying patterns and rules. In such cases, the model may perform poorly on new data and result in poor performance and low applicability in real-world scenarios. Therefore, ensuring that AI systems can generalize well is a critical challenge in the development and deployment of these systems.



S. Liu et al., *Visualization for Artificial Intelligence*, Synthesis Lectures on Visualization,
https://doi.org/10.1007/978-3-031-75340-4_1

On the other hand, as AI becomes more ubiquitous in various high-stake tasks such as precision medicine, law enforcement, and financial investment, there is a growing need for transparency and interpretability of ML models and their predictions. This is where explainable artificial intelligence (XAI) comes in [8, 146]. It enables users to understand the inner workings of these models and trust their generated outputs. In the aforementioned areas, the significance of XAI has increased substantially due to the potential risks and consequences related to inaccurate or biased predictions. This makes XAI techniques essential to ensure the reliability, fairness, and accountability of AI systems [132]. First, the increasing reliance on machine learning models, particularly complex ones like large language models, has made the field of XAI indispensable. These models may have billions of parameters or even more, which makes them hard to interpret. This lack of transparency raises trust issues, especially in critical applications. For example, doctors may be reluctant to rely on a deep model for diagnosing medical conditions if it cannot explain its predictions. Second, without understanding how a model works, it can be difficult to diagnose problems when the model fails to perform as expected. For example, if a self-driving car fails to detect an obstacle and causes an accident, it is important to understand why the associated model failed. This understanding is crucial for implementing effective corrections. Third, XAI is essential to ensure fairness and accountability in decision-making. Machine learning models are increasingly used to make decisions that have a significant impact on people's lives, such as determining whether to grant a loan or hire a job candidate. If the model is biased or unfair, it can have a negative impact on certain groups of people. XAI can help ensure that models are fair and unbiased by providing insights into how the models arrive at their decisions. Finally, XAI can help build trust and acceptance of machine learning models in real-world applications. People are often skeptical of models they do not understand, and this can lead to resistance or even rejection of the technology. By providing explanations of how models work, XAI fosters trust and acceptance among users, which leads to more widespread adoption of machine learning technologies.

## 1.2 Visualization for AI

Visualization transforms data into graphical forms such as charts, graphs, and maps, and allows users to interact with it. Its interactive nature enables users to engage directly with the data, often in real time. This interaction results in a more dynamic and personalized understanding of the data presented. Visualization is frequently used in data analysis and decision-making activities because it allows users to examine complex and massive data from various angles and gain insights and understanding that may not be obvious through other methods. It has demonstrated effectiveness in providing explanations, facilitating communication, and promoting human-machine collaboration [153]. This makes visualization a suitable choice for fully understanding and analyzing AI systems [259, 270]. Visualization can be particularly useful for understanding the data used to train machine learning models,

the inner workings of these models, and how they arrive at their predictions. Consequently, the area of visualization for artificial intelligence (VIS4AI) has emerged as an exciting area for research and development. It offers many opportunities for advancing the use of visualization techniques that can improve the interpretability and reliability of machine learning models. In addition, VIS4AI fosters human-machine collaboration and paves the way for more reliable and effective AI applications.

VIS4AI methods fully combine the advantages of interactive visualization and machine learning techniques to facilitate the analysis and understanding of key components in the learning process, with the aim of improving performance. For example, research in VIS4AI that focuses on explaining the inner workings of deep neural networks has successfully increased the transparency of deep models and has received growing attention from the research community in recent years [44, 98, 154, 272]. These methods are applicable in all phases of the machine learning lifecycle, from data preparation to model development.

As shown in Fig. 1.1, the machine learning lifecycle consists of two pipelines: a data pipeline and a model pipeline. The **data pipeline** prepares the data for machine learning. It typically includes data collection, data cleaning, data augmentation, and feature engineering. The goal of this pipeline is to ensure that the training data collected is representative, unbiased, and of high quality, while also ensuring that the training data contains the necessary features for training a machine learning model. A well-designed data pipeline guarantees not only the accuracy and robustness of the model, but also its ability to effectively generalize to new datasets. The **model pipeline** involves selecting, training, evaluating, and deploying a machine learning model. This pipeline consists of both the model development and model deployment stages. Model development centers on the creation, training, and optimization of a machine learning model. This includes tasks such as model selection, training, validation, and evaluation, all of which aim to identify the best model for the given task and improve its performance. Model deployment aims to make a trained machine learning model available for use in a production environment. This stage includes the tasks of monitoring the model such as scalability, reliability, security, and fairness, as well as maintaining the model such as addressing concept drift. A well-designed model pipeline ensures that the model is accurate, reliable, and robust and, therefore, capable of delivering effective and trustworthy results.

Both the data pipeline and the model pipeline are important components of machine learning. They are often used in combination to build and deploy effective machine learning models. By carefully designing and implementing each pipeline, machine learning practi-

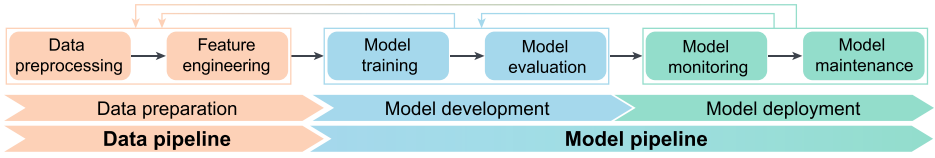


**Fig. 1.1** Machine learning lifecycle consists of a data pipeline and a model pipeline

tioners can ensure that their models are accurate, robust, and effective in solving real-world problems. VIS4AI methods seamlessly integrate with the well-established data pipeline and the model pipeline essential for developing and deploying machine learning models. In the data pipeline, VIS4AI techniques aim to improve the quality of data and features used to train machine learning models. This includes tasks such as cleaning training data and creating interpretable and meaningful features through feature engineering. VIS4AI techniques can effectively identify and mitigate issues such as dataset bias, annotation inconsistency, and outliers, which affect the accuracy and reliability of machine learning models. In the model pipeline, VIS4AI techniques support the development and deployment of machine learning models. During model development, these techniques facilitate model understanding, diagnosis, and steering. They employ visualizations to explore and understand the behavior of the model, identify potential issues, and suggest improvements. Once the model has been developed, the VIS4AI techniques assist in model deployment by enabling decision explanation, model monitoring, and model maintenance. They use interactive visualization techniques to explain the model decisions, monitor model performance in real time, and maintain the performance by tackling the robustness and fairness issues.

## 1.3 The Development of VIS4AI

Over the past two decades, there has been a growing interest in the development of VIS4AI techniques. The goal of these techniques is to enhance the generalizability, interpretability, trustworthiness, and reliability of machine learning models by using visualization techniques. This has become increasingly important as machine learning continues to be used in various applications. To achieve this goal, many VIS4AI methods have been developed to advance the understanding of how an ML model works and how it arrives at its predictions. Figure 1.2 summarizes the evolution of VIS4AI methods over time. The lower part outlines the VIS4AI methods related to the data pipeline, and the upper part outlines the VIS4AI methods related to the model pipeline. The VIS4AI methods on the model pipeline can be classified into two categories based on their target users: model development tailored for model developers and model deployment designed for model consumers.

Initial attempts in VIS4AI for the data pipeline focus on feature engineering, which involves interactive feature selection [124] and feature creation [24]. Later, there are increasing advocates for improving the quality of training data, including improving crowdsourced annotations [157], correction of mislabeled data [260], and the detection of out-of-distribution samples [38]. The researchers also proposed to use unannotated data [37] and multi-modal data [39] to improve model performance. Recently, efforts have been made to reweight training samples in order to address data bias, including noisy labels and imbalanced class distributions in training datasets [267]. In addition, how to detect and analyze data heterogeneity in federated learning is also studied [251].

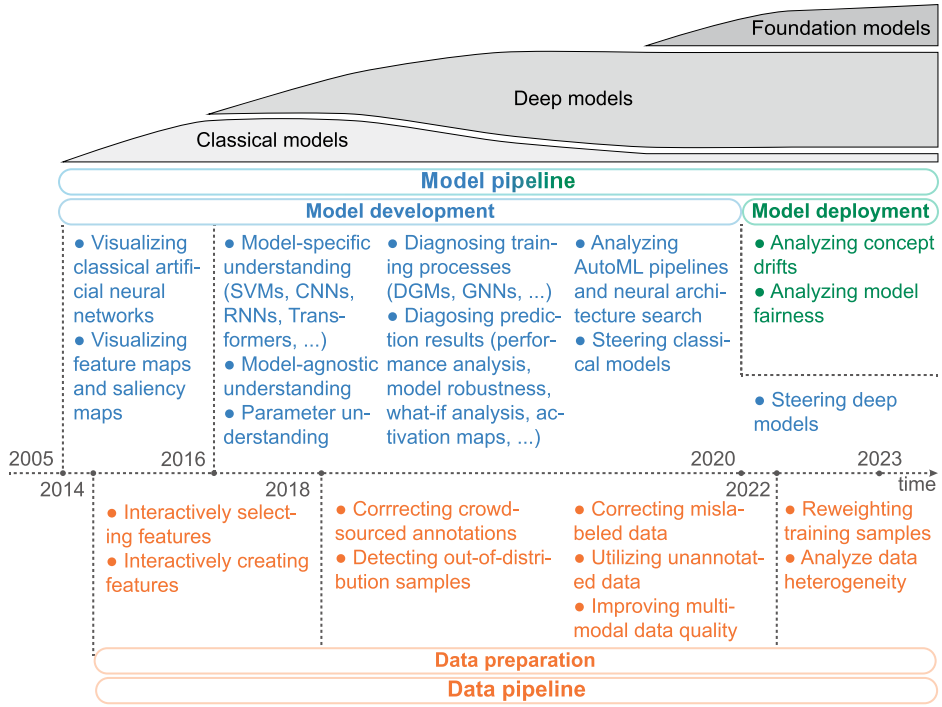


**Fig. 1.2** The evolution of VIS4AI methods over time

Initial efforts in model development focus on utilizing visualization techniques to facilitate the understanding of classical models such as neural networks [234], decision trees [232], and regression models [178]. With the development of deep models, efforts have been shifted to understanding various deep models such as convolutional neural networks (CNNs) [154], recurrent neural networks (RNNs) [172], deep generative models (DGMs) [115], and transformer-based models [53, 145]. Hereafter, researchers seek to diagnose the training process of machine learning models [151] and the prediction results [27]. Recent efforts focus on analyzing AutoML pipelines [186] and neural architecture search [271], and steering classical models [264] and deep models [175]. In the model deployment stage, the researchers focused on analyzing concept drifts [263] and model fairness [26].

## 1.4 Conceptual Framework and Method Overview

The core principle of VIS4AI is based on the well-established mantra of visual analytics, which advocates the integration of interactive visualization and data analysis techniques to facilitate human reasoning and decision-making processes. Keim et al. [119] defined visual