KI - Künstliche Intelligenz (2020) 34:1–3
https://doi.org/10.1007/s13218-020-00639-w

EDITORIAL

Explaining AI: Are We Ready For It?

Britta Wrede[1]

Dear readers,

How often have you tried to explain to a friend or an uncle why his computer crashed or why the navigation system suggests the same route again and again even though he has taken a better one last time? How often have you succeeded? In my experience, my explanations are seldom as successful as I wish them to be and do not lead to a more appropriate behaviour next time. Yet, we expect that AI should do exactly that: explain itself to non-expert users so that they can use it better next time.

This expectation is based on the assumption that if only the explanation is good enough, people will understand. But what if the reason for non-understanding does not lie in the explanation but in the recipient? Understanding in a general way how a computer, a navigation system or a recommendation system works so that certain predictions about their capabilities, behaviour and reliability can be made requires some basic concepts underlying computing.

One attempt to capture basic concepts necessary to solve problems with a computer has been made with the theory of computational thinking which focuses on a user's ability for abstraction, automation and analysis. While these are important abilities for solving problems with a computer they may not be the ones necessary for understanding and appropriately interacting with systems that are based on any kind of "artificial intelligence". We do not yet know what the relevant concepts for interacting with such systems in an informed and mature way are. It may be necessary that people are able to see the difference in system behaviour that is based on learning from observed data alone versus behaviour from human intelligence which is based on situated cognition and experience. More research in this direction is needed.

One may argue that people should not have to learn new concepts in order to be able to use everyday technology. But humans even have to learn how to interact with other humans as well: while babies are born with the capability for empathy and perspective taking, i.e. to interpret others' thoughts and feelings, they still need to learn to take the other's perspective into account in years of training.

The more artificial intelligence is entering our daily life the more we need to provide people with the ability to use and control it in an adequate way. There is a lot of ongoing discussion how to adapt our eduction systems to the challenge of digitalisation. This aspect of AI thinking should be added to it. However, we also need to educate adults many of whom are already working with AI systems in their jobs. While there is a lot of information about how AI can be used for business little is known about how AI is currently being controlled and evaluated in the field and how users are trained to work with it. Public media is providing little to this, albeit interesting formats exist in terms of youtube videos (e.g. from the Year of Artificial Intelligence 2019[1]) or online discussions touching on the impacts of AI on our society (e.g. the Debatten Podcast by Sascha Lobo[2]).

Yet, a more strategic approach needs to be taken—we as a community need to actively engage in this discussion on how to educate people to use AI as this will become one of the most pressing issues in the upcoming decade.

Britta Wrede

1 Forthcoming Special Issues

1.1 Challenges in Interactive Machine Learning

Guest Editors: Stefano Teso (KU Leuven, Belgium) and Oliver Hinz (Goethe University Frankfurt, Germany)

Machine learning is revolutionizing our ability to leverage data and tackle challenging applications such as natural language and image understanding, recommender systems,

✉ Britta Wrede
bwrede@techfak.uni-bielefeld.de

[1] Applied Informatics, Bielefeld University, Bielefeld, Germany

[1] https://www.plattform-lernende-systeme.de/videos.html

[2] https://www.spiegel.de/thema/podcast_lobo/

fraud detection, and medical diagnosis. Interactive learning supports these advances by casting learning as a dialogue between a model and one or more users, who may play the role of teachers, targets, and judges of the model being learned, and who can also learn as being part of the teaching?learning loop.

Designing successful interactive learning schemes requires to solve a number of key challenges:

- minimizing the cognitive cost for the user while optimizing query informativeness,
- devising effective interaction protocols based on different types of queries (membership, ranking, search, explanation, etc.),
- producing optimal questions by explicitly and efficiently capturing the uncertainty of the model,
- distributing the load of query answering across multiple teachers with heterogeneous abilities,
- designing or estimating realistic models of user behavior, increasing tolerance to noise and actively guiding the user toward providing better and more robust supervision, and, more generally,
- automatically discovering the user?s expertise level and adapting the interaction accordingly. Such an interaction is likely to help make such systems more transparent and the results more explainable.

This special issue aims at surveying established research in interactive learning, as well as recent advances on algorithms, models and effective process design around humans in the loop. If you are interested in contributing (technical reports, system descriptions, project reports, survey articles, discussions and dissertation abstracts) to this special issue, please contact one of the guest editors before the submission deadline.

1.2 Special Issue on Ontologies and Data Management

Guest editors: Thomas Schneider (University of Bremen, Germany) and Mantas Simkus (TU Wien, Austria)

This special issue focuses on the theory and practice of applying ontologies in data management (ODM), which is a research topic of significant interest in Knowledge Representation and Reasoning (KR&R) and Database Theory. Modern data-centric software systems need to handle data that is often heterogeneous, sensitive, very large, and even incomplete or inconsistent. Moreover, it often has very complex structures. Thus, the development of proper tools and techniques to handle this complexity is a pressing task. Ontologies in combination with automated reasoning are acknowledged as a promising tool to address some of these challenges, and thus are receiving significant attention both among researchers and industry. For instance, the prominent data integration paradigm called ontology-based data access (OBDA) suggests the use of ontologies to provide a conceptual view of a problem domain, where various possibly heterogeneous data sources can be linked to the same ontology using mappings, enabling users to pose queries using the ontology vocabulary. For query answering in OBDA, automated reasoning is used to compile information from the sources, possibly employing the domain knowledge in the ontology to infer new information.

The special issue welcomes contributions on all aspects of ontologies and data management, including but not limited to:

- Query answering: standard semantics, bag semantics, inconsistency-tolerant semantics
- Further inference tasks in the presence of data: learning, materialization, non-monotonic reasoning
- Decidability and complexity analyses
- Ontology languages and extensions: description logics (DLs), rule-based languages, first-order logic
- Combinations of ontology languages with other formalisms such as temporal logic, probabilities, action formalisms
- Applications related to ODM
- Systems and tools related to ODM

The KI Journal, published and indexed by Springer, supports a variety of formats including technical articles, project descriptions, survey articles, discussions, dissertation abstracts, conference reports, and book reviews.

1.3 Special Issue on Developmental Robotics

Guest Editors: Manfred Eppe (University of Hamburg), Verena V. Hafner (HU Berlin), Yukie Nagai (University of Tokyo), Stefan Wermter (University of Hamburg)

Submission deadline: March 15, 2020

Human intelligence develops through experience, robot intelligence is engineered—is it? At least in the mainstream approaches based on classical Artificial Intelligence (AI) and Machine Learning (ML) the robotic engineering approach is pursued and data- or knowledge-based algorithms are designed to improve a robot's problem-solving performance. Based on this engineering perspective of classical AI/ML approaches plenty of valuable application-specific impact has been achieved. Yet, the achievements are often subject to restrictions that involve domain knowledge as well as constraints concerning application domains and computational hardware.

Developmental Robotics seeks to extend this constrained perspective of engineered artificial robotic cognition, by building on inspiration from biological developmental

processes to design robots that learn in an open-ended continuous fashion. Developmental Robotics considers cognitive domains that involve problem-solving, self-perception, developmental disorders and embodied cognition.

This perspective helps to improve the performance of intelligent robotic agents, and it has already led to significant contributions that inspired cutting-edge application-oriented Machine Learning technology. In addition, Developmental Robotics also provides functional computational models that help to understand and to investigate embodied cognitive processes.

For this special issue, we welcome contributions that include, but are not limited to the following topics:

Robotic self-perception and body representation; Typical development and developmental disorders; Neural foundations of development and learning; Continual learning; Transfer learning; Embodied cognition; Problem-solving; Predictive models; Intrinsic motivation; Language learning.

The KI Journal, published and indexed by Springer, supports a variety of formats including technical articles, project descriptions, survey articles, discussions, dissertation abstracts, conference reports, and book reviews.

Interested authors are encouraged to contact the guest editors at their earliest convenience.

If you are interested in contributing to this special issue, please contact the guest editors:

Acknowledgements Open Access funding provided by Projekt DEAL.