TECHNISCHE UNIVERSITÄT DRESDEN

HS SS-2018

Prof. Dr. Frank J. Furrer

Managing the Risk of Intelligent Algorithms

Summary

*Artificial intelligence based on machine-learning leads to a new type of algorithm: The **intelligent algorithm**. Intelligent algorithms have a tremendous positive potential in many application fields, some already alive today. On the other hand, intelligent algorithms introduce new risks, with potentially very harmful outcomes. Managing, i.e. understanding, assessing, and mitigating those risks becomes an important part when developing, implementing and operating intelligent algorithms.*

Context

With the successful rise of *machine-learning*, a new type of algorithm became possible: The *intelligent algorithm*. Intelligent algorithms have the capability to learn from their environment, from their actions, from big data, and from other sources.

Intelligent algorithms:

- Have a high degree of *autonomy*, due to their capability of self-learning;
- Take *decisions* based on the state of the environment and their behaviour, and improve by self-learning;
- Often control cyber-physical systems, i.e. interact with the physical world.

Today, intelligent algorithms can be found in a number of *successful applications*, such as autonomous cars or trains, robots, financial investment advisors, medical diagnostics, energy distribution, traffic management, … and many more. A growing number of products and services is based on intelligent algorithms.

The behaviour of intelligent algorithms is already more successful in a number of areas than their human counterpart. Thus, the rise of intelligent algorithms is assured. Civilization can expect higher levels of quality of life, safety, and efficiency through the use of intelligent algorithms.

However, the capabilities of intelligent algorithms also introduce new, sometimes significant, *risk*. Because they often autonomously control cyber-physical systems, physical damage to life or property is possible. Due to their decision capabilities based on big data, they may influence decisions on human health, employment chances, investment opportunities, traffic control, … or other areas important for humans or for society.

Intelligent algorithms generate:

- *Technical* risks (Malfunction, loss of service, incorrect decisions, …);
- *Legal* risks (Assignment of responsibility, liability, legal conformance, …);
- *Ethical* risks (Moral questions, boundary between artificial and human intelligence, supervision, …).

As is known with certainty from a long history of technology, *zero risk* is not attainable in technical systems. As the *technology maturity level* raises, the risk is incrementally reduced, but never fully eliminated.

Our task as engineers is to find and contribute to the *acceptable balance* between benefits and risks. This is especially true in the field of intelligent algorithms and leaves us with a high responsibility.

Seminar Work

This seminar will work on the central theme: *How can we manage – i.e. recognize, assess, and mitigate – the risks of intelligent algorithms*?

Each participant choses one of the 3 fields:

F1: Which are possible ***technical*** risks when developing and operating intelligent algorithms? How can we manage these (or a subset)?

F2: Which are possible ***legal*** risks when developing and operating intelligent algorithms? How can we manage these (or a subset)?

F3: Which are possible ***ethical*** risks when developing and operating intelligent algorithms? How can we manage these (or a subset)?

The Hauptseminar has 3 seminar days (see separate work program, dates below):

- An introduction day: ***Intelligent Algorithms*** will be introduced in a lecture by Professor Dr. Frank J. Furrer, and the parts of the Hauptseminar (Paper, presentation) will be defined,
- Individual, guided research in the selected area and authoring of a scientific paper. Feedback from peer reviewers,
- A first seminar day: The participants will present their results and receive feedback from the audience,
- Improvement of the paper and the presentation, based on the peer feedback,
- A second seminar day: The participants will present their improved results and receive feedback from the audience,
- Delivery of the final paper.

Learning Outcome

The participants will learn: (a) to do focused research in a specific area ("Intelligent Algorithms"), (b) to author a scientific paper, (c) to improve their LATex expertise, (d) to experience the peer-review process and (e) to hold convincing presentations, and (f) to benefit from a considerable broading of their perspective in the field of technology, software, and applications.

Seminar language is English. Three seminar days will be held and 3 ECTS credits are awarded for the successful participation.

Audience is limited to 7 participants. Please register in advance.

Mandatory Reading

[1] ***Managing algorithmic risks – Safeguarding the use of complex algorithms and machine learning***. 2017 Deloitte Development LLC., New York, USA. Downloadable from: https://www2.deloitte.com/us/en/pages/risk/articles/algorithmic-machine-learning-risk-management.html?id=us:2em:3na:consedgeIT:awa:cons:081717 [last accessed: 7.2.2018]

[2] ***Demystifying artificial intelligence – What business leaders need to know about cognitive technologies***. 2014 Deloitte Development LLC., New York, USA. Downloadable from: https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/what-is-cognitive-technology.html [last accessed: 7.2.2018]

[3] James Barrat: **Our Final Invention – *Artificial Intelligence and the End of the Human Era***. Griffin Publishing, 2015. ISBN 978-1-2500-5878-2

Seminar Schedule:

Kick-Off Meeting (Introduction): Wednesday, **April 25, 2018** / 11:10 – 12:40 in APB/INF 2101

Seminar Day 1: Wednesday, **June 13, 2018** / 09:20 – 10:50 & 11:10 – 12:40 in APB/INF 2101

Seminar Day 2: Wednesday, **July 11, 20187** / 09:20 – 10:50 & 11:10 – 12:40 in APB/INF 2101