Data Science and Big Data in Upper Secondary Schools: What Should Be Discussed From a Perspective of Computer Science Education?

Birte Heinemann, Lea Budde, Carsten Schulte, Rolf Biehler, Daniel Frischemeier, Susanne Podworny and Thomas Wassong

Abstract The domain of data science is a large field, combining statistics, computer science and sociocultural issues. It is an open question which topics and which contents can and should be implemented in school, e.g. from the perspective of computer science education. Within the frame of a design-based research project a pilot course is designed by computer science and statistics educators at the Paderborn University, addressing upper secondary students. In this paper, we concentrate on the second of four modules, in which machine learning and neural networks are adressed. Some individual phases of the module are presented, followed by a metaperspective of the curriculum development that contributes to our project, and further research questions.

Birte Heinemann · Lea Budde · Carsten Schulte
Didaktik der Informatik, Paderborn University, 33102 Paderborn, Germany
✉ birte.heinemann@uni-paderborn.de
✉ lea.budde@uni-paderborn.de
✉ carsten.schulte@uni-paderborn.de

Rolf Biehler · Daniel Frischemeier · Susanne Podworny · Thomas Wassong
Didaktik der Mathematik, Paderborn University, 33102 Paderborn, Germany
✉ biehler@math.upb.de
✉ dafr@math.upb.de
✉ podworny@math.upb.de

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 5, No. 1, 2018

DOI: 10.5445/KSP/1000087327/26
ISSN 2363-9881

1 Introduction and Project Overview

In the present days and due to the increasing availability of big data, data science is becoming more and more relevant in our daily lives, as can be seen e.g. by the visualisation of Google search trends (see Figure 1). What data science really means, however, and what it affords to do and teach in school education, are open topics.

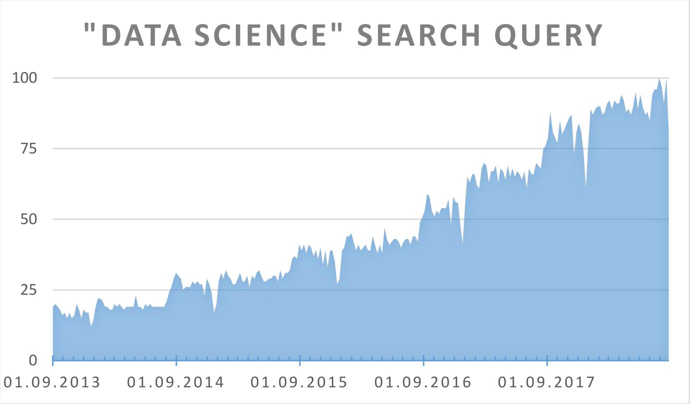

Figure 1: The search query "Data Science" has been gaining popularity for 5 years – represented as a value between 0 and 100. Data source: Google Trends (`https://www.google.com/trends`).

Our project "ProDaBi" (Project: Data science and big data in secondary schools, `https://www.prodabi.de`) searches for answers to the question what, why and how to teach about data science at school. It is a joint project of statistics and computer science education research groups at Paderborn University, which is made possible by the Deutsche Telekom Stiftung (`https://www.telekom-stiftung.de`).

Our goal in the project is the development of a curriculum, which includes the need to figure out core ideas and competencies in the field of big data and data science. The domain, however, is a large area that combines aspects of statistics, computer science and socio-cultural issues. In order to ensure that all these topics lead to a coherent curriculum, we have already discussed our ideas for the project in an interdisciplinary way in some preliminary work. We have organized an international symposium with experts from fields such as computer science education, statistical education, artificial intelligence (computer science), IT industry, and, last but not least cultural sciences (to address sociocultural aspects of Data Science).[1] Results and an overview on several facets of data science can be found in Biehler et al. (2018a).

First ideas and a good overview of the whole project can be found in the paper of Heinemann et al. (2018). A main objective in the first year of the project is the conception, implementation and evaluation of a one-year pilot course (3 hours per week) of class 11–12 at an upper secondary school in Paderborn. The course consists of 19 students, 2 female and 17 male.

Our pilot course on data science and big data in upper secondary schools 2018/2019 (and so the resulting curriculum) is divided into four modules:

1. In the first module, the students are familiarised with basic statistical procedures around big data. This (statistical) module aims to promote data literacy and awareness. A detailed description and a first insight into our reflection can be found in Biehler et al. (2018b).

2. The second module, which is described in more detail in this article, is intended to introduce students to the area of artificial intelligence, in particular into machine learning. The learning goal of module 2 is the change from classical algorithmic problem solving to data-driven processes using the subtopic of machine learning. The students learn the basics of Python programming, decision trees and artificial neural networks.

3. The third module gives the opportunity to try out and deepen the learned basics in real "Big Data" projects. The students will work in two smaller groups of 9 students on a real data project and evaluate it. An important element will be the interaction of "stepping in" and "stepping out". The students actively carry out their data project (stepping in) and should cyclically distance themselves from the active process and reflect on further aspects of their project (stepping out). In this way, experiences can be gathered which then can be reflected on in the broader context: What goals, opportunities and changes does my work bring about?

---

[1] Perspectives for data science Education at School Level – Educational Contributions from Statistics, Computer Science and Sociocultural Studies (`http://go.upb.de/SymposiumProDaBi`).

4. In the last and fourth module, social and cultural aspects based on the collected experiences will be dealt with. In this module the chances and risks, as well as the role of the human being in the context of data science will be finally reflected.

In the following Section 2, we will go one step back from the more concrete contents to the question “Why should the students learn data science?”. We want to make it quite clear that educational questions and content in the classroom should not (only) be determined by current scientific topics. Rather, the curriculum and lesson development should be viewed from the pupils’ perspective: What do young people need in the future to be able to lead a mature and self-determined life? Subsequently, we want to concretize the rather abstract thoughts by presenting our concrete implementation in Section 3. Here our second module (artificial intelligence and machine learning) is presented in more detail. The paper ends with an outlook and further steps in connection with our project in Section 4.

2 Identifying the Rationale

Developing a curriculum can be an astonishing complex endeavour. In curriculum theory or curriculum research, several models can be found, one interesting model is presented by Van den Akker (2004). In the curricular spider web, Van den Akker outlined the idea, that the various components of a curriculum, and hence of teaching and learning are interrelated, and mutually dependent on each other (see Figure 2).

Thijs and van den Akker comment: “Our preferential visualisation of the ten components is to arrange them as a spider web [...] not only illustrating its many interconnections, but also underlining its vulnerability” (Van den Akker, 2004, p.5).

The most important component, and consequently placed in the middle of the web, is the rationale, the underlying vision or basic philosophy of the curriculum. The rationale should not be confused with the aims and objectives of the curriculum – it serves more the purpose of linking together the components, providing consistency, and providing basic principles for justification and evaluation of a curriculum.

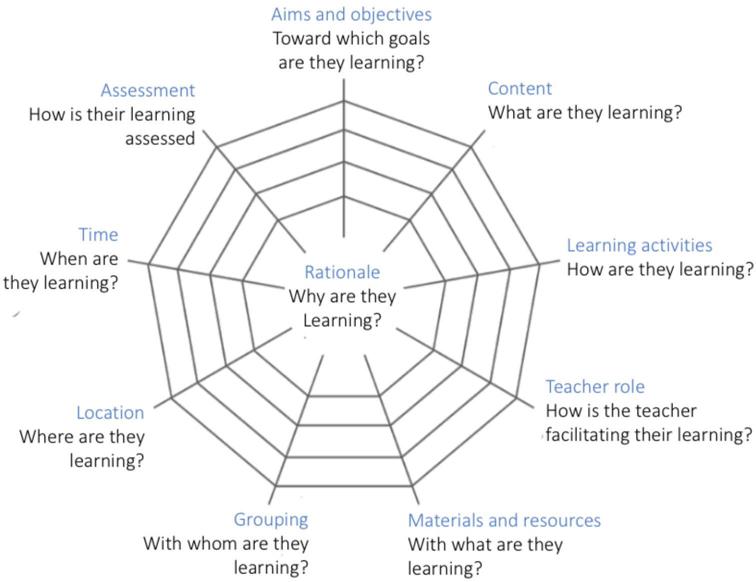

Figure 2: Curricular spider web (compare with Van den Akker (2004), page 6).

The implicit or underlying nature of the rationale often causes it to be mentioned in the introduction only, and to be implicit, and, therefore, often overlooked in the rest of the curriculum. It is, however, probably the most relevant issue in curriculum development. As Van den Akker puts it: “The *rationale* (referring to overall principles or central mission of the plan) serves as major orientation point, and the [...] other components are ideally linked to that rationale and preferably also consistent with each other” (Van den Akker, 2004, p.40).

The rationale is not only relevant for curriculum designers, but also for teachers. There is often a gap between curricular ideas and the specific teaching implementation. In a generalised form, the problem is that the teachers often have a wrong or missing idea of computer science and they (among other things), therefore, are not clear what individual topics should be taught. Duncan et al. (2017) gives an example of this problem:

> “ *For example, in some of the professional development with the teachers using black and white cards to represent binary values, teachers have asked which colour is used for* 0 *and* 1*. The real concept is that some sort of convention must be agreed on, rather than rote-learning a rule about this particular abstraction.* ” (Duncan et al., 2017)

In addition, the problem goes deeper, because it should not only be possible to draw conclusions about computer science principles from the teaching example, but also about their relevance for everyday life, e.g. in connection with other information systems – but how should this be possible for learners, if even the teachers cannot draw this connection? Tim Bell calls the problem the need for a "big picture" the teachers (and curriculum designers) should be aware of (see Bell, 2018). This is also important for the data science curriculum we want to develop, especially as data science is a dynamic and complex field (see also Biehler et al., 2018a; Arbeitsgruppe Curriculum 4.0, 2018).

Our idea for a big picture or rationale underlying the data science curriculum is that of a hybrid interaction system between humans and technology. The role of a big picture in general is described in more detail in Schulte and Budde (2018).

The interaction between two agents lies at the core of the model: The human being and the technology. This should be denoted by the term "hybrid". The interaction can be understood as a chain of actions that goes back and forth between the two parties: "[...] human action followed by computer action followed by human action [...]". Both agents have their own characteristics and intentions that influence the interaction. The crucial idea is that the actions of both are shaping and at the same time are being shaped by the Hybrid Interaction System (HIS). And, moreover, the agents themselves are also shaping and being shaped by and within the HIS. The roles and nature of one of the agents cannot be fully understood without the framework or context of the HIS which they are part of.

In the end, within this framework we conceptualise the role of a designer or computer scientist not as developing an isolated piece of technology, but instead of designing and developing the HIS (shown in Figure 3).

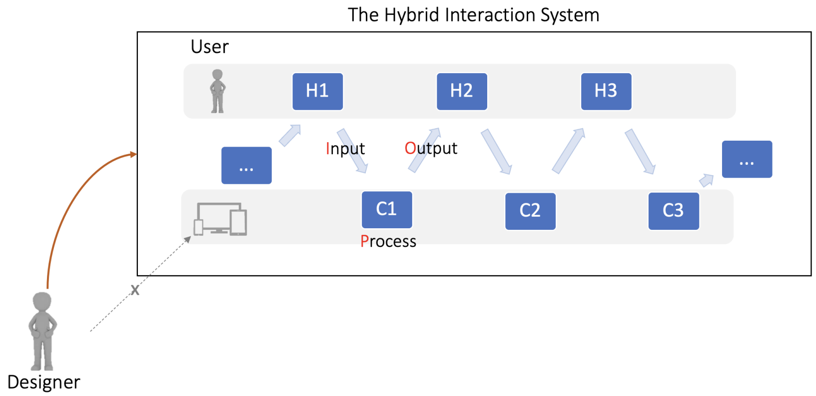

Figure 3: Role of the designer in the context of the Hybrid Interaction System (HIS).

The question now is, is this framework useful in the context of data science, and more precisely, for teaching about artificial intelligence and machine learning? Based on the model of the HIS we can explore the bigger picture of AI to be transported in the teaching module:

1. **The change in problem-solving:** Problem solving is, in general, one of the major rationales why to teach computer science at school. So, when the overall approach to problem solving in computer science changes, that should be reflected properly. We can see that the traditional approach can be conceptualized as focusing on the Input-Process-Output (IPO) model, and even more to producing the P: the way the computer transforms an input into the desired output (see Figure 3). So we can embed the traditional approach to problem solving within a bigger picture. With AI, the designers role can be seen as changing from analyzing input and output in order to find a solution, and having to understand the solution in order to find the algorithm. Now the system learns the solution by data driven methods that automatically generate the connection of input to output. And the system is able to transfer the found solution to new cases. This changes the role of the designer and developer of computational artifacts.

2. **New role for users:** In the classical view, a user just uses the algorithmic solution designed by the computer scientists, here the user does the same, but only at the first look: By triggering the AI to search for a solution, based on the user's input, the user is not only using a fixed system; instead she is feeding some new data points into it. An AI-system can continue learning, and not only giving a solution based on a re-trained model. In this way the user's role changes because (s)he is part of the process to design or refine the solution. A prominent example is Microsofts chatbot Tay (see Reese, 2016).

3. **New roles of designer and user in the HIS:** The sometimes as strictly separated roles of user vs. designer conceptualized relationship between user and designer is replaced by a continuum of in-between roles (see Schulte and Budde, 2018, for details). It means there are users who become designers even during the interaction. They actively change technology and are not pure users of predefined interaction possibilities. We can see the same with AI: Designers design a "living" system, which is influenced by the user anyway, so why not allow the user to be able to interfere with how the user gets solutions, see e.g the suggestions by Weigend (2017). The user on the other hand should see his own responsibilities for results and their interpretation. See again the example of Microsoft's chat bot, and the discussion on bias in data.

4. **The notion of the human-in-the-loop:** This is sometimes debated in AI (see also Demartini, 2015; Holzinger et al., 2016). It is an important aspect to consider in the teaching module: The human **is shaped** by AI, but the human can also **shape** the AI, that is the general idea in the HIS-concept of shaping and being shaped by technology.

To sum it up: The notion of HIS should foster reflection within the data science course that the roles of the human and of the technological agent can be conceptualised in various different ways, and ultimately our learners' should reflect on these roles of humans. We think for example, that the "Human-in-the-loop" in AI systems has several roles. Especially, machine learning is not problem solving without human interference. Instead, setting up machine learning systems allows and requires human creativity, experience, and professionalism.

In our project we have tried to implement these ideas of the "bigger picture" in first attempts. In module 2, the idea of the "the change in general problem-solving approach (classical approach)" and "the change in general problem solving approach, the new approach" should be made tangible. In the following chapter we will introduce the module on the more content-related level. In this way it will become clear how the ideas can be combined with concrete teaching at the curriculum level.

3 Looking for the Overlap Between Data Science and Computer Science Eduction: The Module on Artificial Intelligence and Machine Learning

The second module, which lasts 7 weeks with overall 21 teaching hours, focuses on the topic of machine learning, a field which connects statistical methods and computational learning. A challenge in this module is the meaningful introduction of the basic concepts of artificial intelligence (AI). We need an appropriate didactical reduction of a topic, which is also treated in whole university courses. That is why we begin with an overview and an introduction to the subject area of AI and machine learning. Our aim is for learners to understand the change towards "learning algorithms".

Since even the field of machine learning is still too large to be dealt with comprehensively, we will focus on decision trees and artificial neural networks as examples for two important paradigms (symbolic AI vs Neural Networks, see Munakata, 2008, p.2).

Module 2 is thus divided into two units. In the first unit pupils will get in touch with the shift from classical problem-solving with self- and handmade Decision Trees to generated Decision Trees and data-driven problem-solving methods.

In a first step the learners generate small trees "manually", and then learn how decision trees can be generated automatically by an algorithm (using probably ID3). By comparing and reflecting on the difference between manually constructed and algorithmically derived decision trees the learners should understand the shift from classical algorithmic problem-solving to machine learning methods, and the associated different role of the human problem solver.

The second unit contains practical experiences with artificial neural networks and first discussions about effects on society, which will be deepened in the following modules in the pilot course. In the next two subsections we will point out further details on both units, giving some activities, design ideas and learning goals.

3.1 Unit 1: Introduction in Machine Learning and Decision Trees

In this section we will not go into detail about the thematic treatment of Decision Trees and will only present the introduction to machine learning. As already mentioned, we will begin to teach the students basics about AI and machine learning and then come to the first of the two more deeply discussed paradigms. Pupils will gain experience with the concept of white-box learning using Decision Trees as an example. More precisely we want our students to ...

... learn about basic ideas of artificial intelligence and machine learning.

... learn two concrete methods, namely decision trees and artificial neural networks.

... be able to describe how machines learn.

... be able to describe limitations of machine learning and to distinguish it from human learning.

... gain deeper knowledge in programming with Python using Jupyter Notebooks (in contrast to module 1, see Biehler et al., 2018b).

... get basic knowledge about validation of models.

... understand how to use Python-packages for data science, e.g. sklearn.

... analyze data with Decision Trees, (ID3 algorithm).

In the introduction to this learning module, pupils will get a first idea of the concept "Learning" by performing an unplugged activity, which is called Sweet Learning Computer McOwan and Curzon (2018). Our work in this context was further differentiated and published in the Science Year 2019 (an initiative of the german ministry of Education and Research). Further information can be found on the website: `https://www.wissenschaftsjahr.de/2019/`. It is a game that is played like a small version of chess called Hexapawn. Moves are represented with the help of (the colour of) sweets (see Figure 4). Moves that have not led to success are eaten by the pupils, and so the learning of the Sweet Learning Computer takes place. Initial random moves are improved by this form of training, where eating up of sweets is connected to deletion of bad moves. In short, the computer learns in the unplugged game by deleting bad moves from the repertoire of all possible moves. So the machine will not make that move again. This way the machine learns from mistakes. Aspects that should be discussed afterwards include the fact that learning depends on the quality of the data (how the students play). Another crucial point for learning is also the topology. In the game the learners learn that the machine learns by making mistakes. Another possibility would be to reinforce good moves. These and other possibilities can be discussed in the reflection of the game played. The knowledge about the created trees is used in the following sessions to introduce decision trees and to meet the other learning objectives mentioned above.

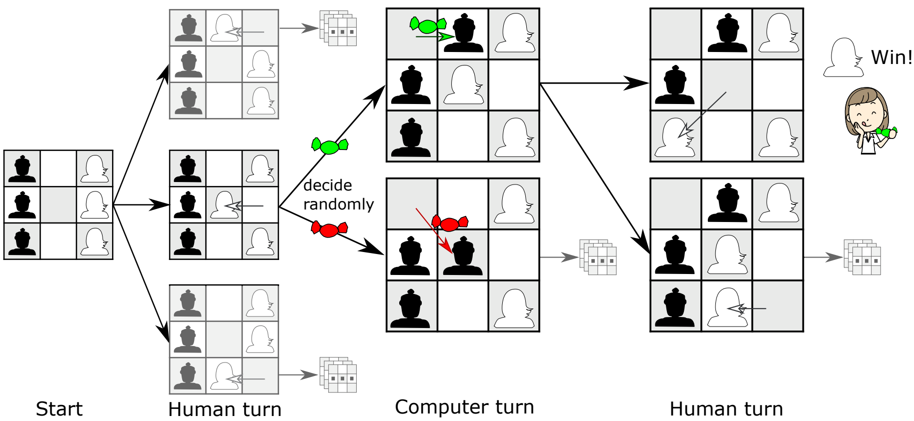

Figure 4: If the human wins, the last move of the computer is deleted. Deletion means the human eats the sweets and thus the move that caused the loss no longer occurs.[2]

[2] Built upon the image series "Mmmmm Delicious" by Oksmith, published on `http://openclipart.org/`. All of the images on this site are in the public domain i.e. CC0.

3.2 Unit 2: Artificial Neural Networks (ANN)

The second unit introduces artificial neural networks and will cover the training phase. Why the black box-like solution process of ANN is difficult to explain (see Benítez et al., 1997). Benítez et al. (1997) show a way (with university complexity) to "obtain an understandable interpretation of neural nets". More precisely we want our students to ...

- ... know how numbers and characters are automatically recognized.
- ... understand that black box models such as ANN do not require any (human-like) prior knowledge, for example, to classify numbers successfully.
- ... understand connections between layers in ANN and what the meaning of training is.
- ... understand the general principle of backpropagation and modification of weights.
- ... know about avoidance of over-adaptation.
- ... have knowledge about "deep" learning.
- ... know basic terms such as activation function, input layer, output layer, hidden layer, epochs, etc.
- ... analyze, interpret, process and visualize information with machine learning methods.
- ... understand what a bias is and to know about societal challenges and ethical questions.

The introduction to the topic will again take place via an unplugged activity, which will familiarize the students with the tasks and functions of a neural network (see McOwan and Curzon, 2016). An illustration of this activity can be found in Figure 5. The activity is called brain-in-a-bag, because everything you need to build an artificial brain fits in a shopping bag. The artificial brain, in which each student represents a neuron, plays the simple card game Snap. If enough students take part, two groups representing two brains can play against each other.

There are three types of neurons that are embodied in the neural network by students. The respective types of neurons always represent one layer of the network. The first layer is responsible for input. The learners playing neurons of this input layer receive input and react accordingly. This means that if they see a red card, for example, they give a roll of paper to the next student. The student who receives this roll is a neuron of the next layer. The so-called processing layer. They represent the brain as interneurones. These students can get a roll of paper from two different places. If they do, they “switch”, passing the roll of paper one layer at a time. The third and final layer is the output. If a paper roll arrives here, the student “switches” and outputs “snap”. So we have three layers in total:

1. Input as vision processing,
2. Processing as brain,
3. Output as speech processing.

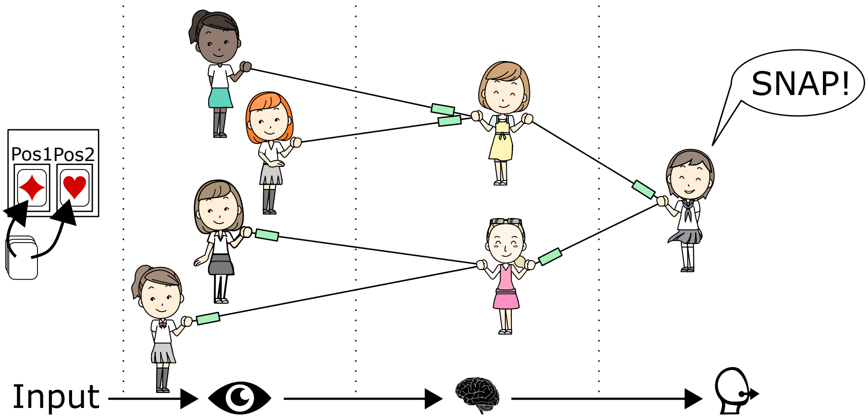

Figure 5: The procedure of the game Snap in an abstracted brain with artificial neurons. Ropes represent connections between neurons, tubes represent messages between neurons. [3]

[3] Built on a picture from the series “Mmmmm Delicious” by Oksmith, published on `http://openclipart.org/`. All of the images on this site are in the public domain i.e. CC0.

These layers are marked in the figure and carry a respective symbol eye, brain, output (in the sense of speaking). The students get only their description of their task. If we now let the whole network act, we can observe how the interaction creates a system, which recognizes cards of the same color.

An example to illustrate these concepts, ways of thinking, approaches and algorithms in the next part of the unit is the handwritten digit recognition task (using the MNIST Database, see LeCun et al., 1998). Students experiment with the influence of parameters like e.g. the number of neurons and layers, the learning rate, the activation function, the data representation used for input (features), and so on using the interactive visualisation provided by the tensor flow playground (`https://playground.tensorflow.org/`). Later they will experiment in the same way with an implemented neural network in a Jupyter notebook, similar to the example provided by Tariq Rashid's course: "Make Your Own Neural Network" (Rashid, 2016) or "Deep Learning with Python" by Chollet (2017).

In addition, we plan to introduce the mathematics of back-propagation at least to some degree and probably let them calculate a small example, similar e.g. to calculating the numbers by hand on squared paper, like it is done by Schaal (2017).

These examples as a whole should provide the pupils with an understanding of the role of parameters, and hence the human designer of a machine learning setup, as well as some ideas how the black box works. In addition, the shown techniques should be understood and mastered to a degree that allows to use them (with probably some additional help by the supervisors) in the project phase of the course (that is, in module three).

A discussion about the risks and about possible biases serves as an introduction to the societal challenges, ethical questions and the role of humans in human-computer interaction. Possible questions are:

- How can you combine the different types of learning and intelligence to leverage the strengths of both?
- How do man and machine act as interaction partners in order to represent a uniform system?

4 Summary and Further Steps

Within the framework of our project, there are a multitude of exciting questions that still need to be considered. These have already been intensively analysed in certain steps in our planning and integrated into the plan.

An exciting question concerns the field of learning behaviour of students, competences and (possible) misconceptions. Since data science and big data is still a very young and broad field, there is little research about beliefs and (mis-)conceptions of learners (and teachers). Also, models for competence development and learning trajectories in the field of data science and big data have not yet been developed and researched in detail.

Another aspect we want to include is what difference it makes for students to use different (categories of) tools. By way of example, we will use two tools to investigate the influence tool use has on the use of What You See Is What You Get (WYSIWYG) tools or tools that require programming skills. One subquestion, we want to adress, is part of the discussion about teaching and learning with digital media. In German politics, the focus is often on "learning with media", "learning about media" and how technology works is rarely taken into account. How could we implement a sustainable and meaningful media education not only in data science education, but also in computer science education and in the long term and, especially, through a focus on the "Allgemeinbildung" (general education) in all subjects in school? A good overview of the current state of the German debate can be found in Bastian (2017).

Another field for research will be the implementation of concrete teaching. It should be observed, which contents and methods affect the learners and how. On the one hand, the work in projects will be methodically implemented with points of reference to agile software development. How can the evaluated method (for example SCRUM) be adapted in teaching and what effect does this methodical design of teaching have? A further aspect which can be analysed in more detail is the implemented data cycle. Within the teaching units of the different modules, learning activities are oriented towards the data cycle. For our pilot course we have so far been strongly oriented towards the PPDAC cycle (see Wild and Pfannkuch, 1999). Where the acronym PPDAC discribes initials of the names of the five steps of the cycle: Problem, Plan, Data, Analysis, Conclusion. The following questions are interesting: How can this cycle be integrated into teaching and how can learners experience the processes in the data cycle?

The research questions outlined here are intended to provide an insight into further work. In addition to the questions and aspects described here, there are certainly numerous other fields which are of interest. A more detailed and differentiated description of these research questions combined with the further steps can be found in Heinemann et al. (2018).

In this paper we have tried to give a first impression of our project and our work. On the one hand, we focused on the description and contents of the second module. On the other hand, we have tried to clarify our view of the rationale. We are of the opinion that one must always start from the learner when developing and designing a curriculum and learning processes. Based on the considerations about why learners have to learn something, one can think concretely about possible contents and learning processes. We hope that this attitude became clear in our paper and also inspires others to think about educational theory.

The research questions outlined in the previous section made it clear that the project still has potential and further work to do. Within the framework of our project we are excited to see how our planned learning units work for the learners and what we can observe. Based on the design-based-research principle, we collect a lot of data and observations to document the learning process as accurately as possible (video recordings, audio recordings, collection of student results and observations using observation sheets). On the basis of the data we will then analyse the lessons and draw conclusions about our planning. The acquired knowledge will then flow into our curriculum development and be tried again in the next cycle in 2019/2020. Furthermore, we are pursuing the idea of designing and implementing teacher training courses and offers for interested people. We hope our work will be broad-based and that we will receive differentiated and varied feedback from other teachers and researchers.

Acknowledgements Our programme was made possible by the Deutsche Telekom Stiftung. The authors would like to thank the participants of "Perspectives for data science education at school level – Educational contributions from statistics, computer science and sociocultural studies" – symposium for their support and for their comments, which help to develop a curriculum draft.

References

Van den Akker J (2004) Curriculum Perspectives: An Introduction. In: Van den Akker J, Kuiper W, Hameyer U (eds.), Curriculum Landscapes and Trends. Springer, Dordrecht, pp. 1–10. ISBN: 978-9-048165-11-7.

Arbeitsgruppe Curriculum 4.0 (2018) Curriculumentwicklung und Kompetenzen für das digitale Zeitalter: Thesen und Empfehlungen der AG Curriculum 4.0 des Hochschulforum Digitalisierung. Arbeitspapier Nr. 39. Hochschulforum Digitalisierung, Berlin. DOI: 10.5281/zenodo.2602541.

Bastian J (2017) Lernen mit Medien – Lernen über Medien? Eine Bestandsaufnahme zu aktuellen Schwerpunktsetzungen. DDS - Die Deutsche Schule 109(2):146–162.

Bell T (2018) Computer Science in K-12 Education: The Big Picture. Olympiads in Informatics 12:3–11. DOI: 10.15388/ioi.2018.01.

Benítez JM, Castro JL, Requena I (1997) Are Artificial Neural Networks Black Boxes? IEEE Transactions on neural networks 8(5):1156–1164. DOI: 10.1109/72.623216.

Biehler R, Budde L, Frischemeier D, Heinemann B, Podworny S, Schulte C, Wassong T (2018a) Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts. Universitätsbibliothek Paderborn, Paderborn. DOI: 10.17619/UNIPB/1-374.

Biehler R, Frischemeier D, Podworny S, Wassong T, Budde L, Heinemann B, Schulte C (2018b) Data Science and Big Data in Upper Secondary Schools: A Module to Build up First Components of Statistical Thinking in a Data Science Curriculum. Archives of Data Science, Series A 5(1). DOI: 10.5445/KSP/1000087327/28.

Chollet F (2017) Deep Learning with Python, 1st edn. Manning Publications Co., Greenwich, CT, USA.

Demartini G (2015) Hybrid Human-Machine Information Systems: Challenges and Opportunities. Computer Networks 90:5–13. DOI: 10.1016/j.comnet.2015.05.018.

Duncan C, Bell T, Atlas J (2017) What Do the Teachers Think? Introducing Computational Thinking in the Primary School Curriculum. In: Proceedings of the Nineteenth Australasian Computing Education Conference, ACE '17. Association for Computing Machinery, New York, NY, USA, pp. 65–74. DOI: 10.1145/3013499.3013506.

Heinemann B, Opel S, Budde L, Schulte C, Frischemeier D, Biehler R, Podworny S, Wassong T (2018) Drafting a Data Science Curriculum for Secondary Schools. In: Proceedings of the 18th Koli Calling International Conference on Computing Education Research, Koli, Finland, November 22-25, 2018, ACM, New York, NY, USA, pp. 12–1–12:10. DOI: 10.1145/3279720.3279737.

Holzinger A, Valdez AC, Ziefle M (2016) Towards Interactive Recommender Systems with the Doctor-in-the-Loop. In: Weyers B, Dittmar A (eds.), Mensch und Computer 2016 – Workshopband, Gesellschaft für Informatik e.V., Aachen, pp. 1–9. DOI: 10.18420/muc2016-ws11-0001.

LeCun Y, Cortes C, Burges CJ (1998) MNIST handwritten digit database. Published via: http://yann.lecun.com/exdb/mnist/. Accessed: 2018-06-15.

McOwan P, Curzon P (2016) Computer Science with a Sense of Fun: Brain-in-a-Bag V2.2 (21 November 2014). Published via: Teaching London Computing: A Resource Hub from CAS London & CS4FN. Website. Online: https://web.archive.org/web/20200219214614/, https://teachinglondoncomputing.org/resources/inspiring-unplugged-classroom-activities/the-brain-in-a-bag-activity/.

McOwan P, Curzon P (2018) Computer Science with a Sense of Fun: Sweet Computer Guide V2.1 (10 Jun 2016), Based on the Matchbox Computers of Donald Michie and Martin Gardner. Published via: Teaching London Computing: A Resource Hub from CAS London & CS4FN. Website. Online: https://web.archive.org/web/20200219222348/, https://teachinglondoncomputing.org/the-sweet-learning-computer/.

Munakata T (2008) Fundamentals of the New Artificial Intelligence: Neural, Evolutionary, Fuzzy and More. Texts in Computer Science, Springer, London. ISBN: 978-1-846288-39-5, URL: https://books.google.de/books?id=lei-Zt8UGSQC.

Rashid T (2016) Make Your Own Neural Network, 1st edn. CreateSpace Independent Publishing Platform, USA. ISBN: 978-1-530826-60-5, URL: https://dl.acm.org/doi/book/10.5555/3027816.

Reese H (2016) Why Microsoft's 'Tay' AI Bot Went Wrong. Published via: TechRepublic. Website. Online: https://web.archive.org/web/20200219214311/, https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/.

Schaal J (2017) Neuronale Netze – Backpropagation – Forwardpass. URL: https://youtu.be/YIqYBxpv53A. Accessed Okt. 30, 2018.

Schulte C, Budde L (2018) A Framework for Computing Education: Hybrid Interaction System: The Need for a Bigger Picture in Computing Education. In: Joy M, Ihantola P (eds.), Proceedings of the 18th Koli Calling International Conference on Computing Education Research, Koli, Finland, November 22-25, 2018, ACM, New York, NY, USA, pp. 12–1–12:10. DOI: 10.1145/3279720.3279733.

Weigend A (2017) Data for the People: How to Make Our Post-Privacy Economy Work for You. Hachette UK, London. ISBN: 978-0-465096-53-4.

Wild CJ, Pfannkuch M (1999) Statistical Thinking in Empirical Enquiry. International Statistical Review 67(3):223–248. DOI: 10.1111/j.1751-5823.1999.tb00442.x