Deep learning goes to school: toward a relational understanding of AI in education

Carlo Perrotta & Neil Selwyn

To cite this article: Carlo Perrotta & Neil Selwyn (2019): Deep learning goes to school: toward a relational understanding of AI in education, Learning, Media and Technology, DOI: 10.1080/17439884.2020.1686017

Deep learning goes to school: toward a relational understanding of AI in education

Carlo Perrotta & Neil Selwyn

Contact: carlo.perrotta@monash.edu

ABSTRACT

In Applied AI, or 'machine learning', methods such as neural networks are used to train computers to perform tasks without human intervention. In this article, we question the applicability of these methods to education. In particular, we consider a case of recent attempts from data scientists to add AI elements to a handful of online learning environments, such as Khan Academy and the ASSISTments intelligent tutoring system. Drawing on Science and Technology Studies (STS), we provide a detailed examination of the scholarly work carried out by several data scientists around the use of 'deep learning' to predict aspects of educational performance. This approach draws attention to relations between various (problematic) units of analysis: flawed data, partially incomprehensible computational methods, narrow forms of educational knowledge baked into the online environments, and a reductionist discourse of data science with evident economic ramifications. These relations can be framed ethnographically as a 'controversy' that casts doubts on AI as an objective scientific endeavour, whilst illuminating the confusions, the disagreements and the economic interests that surround its implementations.

Introduction

In 2015, a research poster presentation titled 'Deep Knowledge Tracing' featured in the twenty-ninth conference on neural information processing systems (Piech et al. 2015). The research was a collaboration between Stanford University, Google Brain (Google's research unit on artificial intelligence), and Khan Academy (the popular online learning platform combining video lectures and algorithmic personalisation).

This study saw the authors apply a state-of-the-art 'deep learning' approach to two 'educational' datasets. One of these datasets derived from Kahn Academy, and another originated from an Intelligent Tutoring System used widely in the US secondary school sector called ASSISTments. Deep learning is a specific method of predictive modelling where computers are able to discover patterns in data using complex recursive operations which, through the use of artificial neural networks, loosely mimic how a biological brain works. Once a deep learning application has learnt how to predict patterns in a 'training dataset', it can be used subsequently to predict the same patterns when it encounters new instances of the same sort of data. Piech and colleagues' proposal for deep knowledge tracing (DKT) revolved around the claim that neural networks can be used to discover unexpected ('deep') features of the data arising from students' use of online learning environments.

While deep learning had been applied before to educational data, the Deep Knowledge Tracing study caused considerable interest among the data science community due to the authors' claims of achieving 85 percent predictive efficiency. Crucially, the authors reasoned that their work could be generalised to other learning environments, this potentially having great implications for anyone working in the areas of personalised learning and adaptive learning. The finding even prompted favourable news media coverage, with headlines such as 'RoboTutor is a Class Act' (Macdonald 2016; Rutkin 2015). Nevertheless, these initial reports also acknowledged doubts over the applicability of the original study's findings for actual educational practice. Neil Heffernan, a computer scientist from Worcester Polytechnic Institute in Massachusetts involved in the original design of the ASSISTments system, put it bluntly: 'What does that mean, to be able to do a much better job at predicting stuff? I wish we could turn that into something that's meaningful.'

This statement was intended not so much as an expression of total disapproval as an invitation to caution. Nevertheless, it began to expose interesting fault-lines in the debate on automation in education. Going back a few years to an interview with the New York Times Magazine, also titled 'The machines are taking over' (Murphy Paul 2012), we learn of the existence of two 'camps' in the educational data mining and Intelligent Tutoring System communities – one (represented in the article by Heffernan) that sees humans and computers interacting in an organic manner, and another that pursues automation in a much more vigorous fashion (reportedly championed by Ken Koedinger, a professor of human–computer interaction and psychology at Carnegie Mellon University). Heffernan is again quoted as saying 'Let computers do what computers are good at, and people do what people are good at.' These excerpts of voice from the mainstream media are valuable in illustrating how scholarly divergences encroach into larger public controversies around automation and human-machine cohabitation. They also point to the distinct ethos of development that underpins the non-for-profit ASSISTments platform, based on open research principles and supported by substantial public funding (https://www.neilheffernan.net/bio/grants). This sets it apart from more corporate educational platforms like Khan Academy or Knewton, which do not tend to share datasets and confidential information about their underlying analytical models.

Against this lively background, the claim for a huge improvement in predictive efficacy made in the DKT study sparked a period of intense research activity amongst a small group of particularly interested data scientists and, within a few years, a number of more formal responses had been published. These follow-up studies introduced other computational methods and, in some cases, other datasets on which the deep learning techniques were trained. In total, six educational datasets (only the ASSISTments one being open) were used during this period of replication and further testing of the original claims. These studies challenged the original result on the basis of several arguments. Most were simply interested in the computational aspects or sought to show that pre-existing methods could perform just as well with careful adjustments (Lalwani and Agrawal 2017; Yeung and Yeung 2018; Wilson et al. 2016; Wang et al. 2017; Zhang et al. 2017). One study developed a theoretical argument against the applicability of automated knowledge discovery to educational data (Khajah, Lindsey, and Mozer 2016), and another identified problems of data integrity in the original 2015 study which, it was argued, contributed to inflated results (Xiong et al. 2016).

By the standards of academic controversies, this sequence of claims and counterclaims was of minor significance – prompting little excitement beyond the confines of the educational data science community. Research on deep learning in education neither 'took off' nor ceased altogether as a result of these contestations, although a more refined consensus was reached about the uses and misuses of neural networks. Instead, the DKT debate is perhaps most significant in highlighting issues that extend beyond the relatively straightforward aspects of empirical replication and scholarly discussion. In one sense, the DKT case is indicative of a cycle of particularly exaggerated initial hype and ensuing backlash that AI seems to attract (Elish and Boyd 2017). Another significant aspect of the DKT episode was the central involvement of for-profit actors (such as Google, Khan Academy, Knewton and Funtoot). These organisations all share considerable economic interests in the educational take-up of AI-powered predictive modelling. Indeed, in general terms outside of education, neural networks are poised to become a mainstream tool through the development of off-the-shelf AI frameworks such as Google's TensorFlow. Without doubt, there is much commercial interest in

applying these developments to educational contexts. In view of this, the present paper explores the 'minor' DKT controversy as a significant bellwether for the likely future of AI in education. Drawing on ideas and methods from the field of Science and Technology Studies (STS), we examine the eight main empirical studies that made up the central sequential 'back and forth' of the controversy. These are approached in an interpretative manner –that is, not simply as empirical reports but as evidence of relationships between key sociotechnical elements. The remainder of this paper therefore approaches the DKT case in terms of three research questions relevant to any application of AI in education:

(a) How does educational AI (AIED) operationally 'act' on datasets? Or, more specifically, how can we gain a sense of (a particular brand of) AIED's underlying logics of knowledge modelling and learning progression?
(b) Why do educational data scientists claim that new forms of learning can be 'discovered' and predicted by machines, without human intervention?
(c) If we explore the cultural assumptions operationalised in AIED in the context of commercial competitive relations, what would this tell us about how these assumptions were formed?

Before exploring questions, we first outline some basic conceptual assumptions and identify contributions from the STS literature that helped us operationalise our approach.

STS and artificial intelligence

While the subject of much discussion, in general terms Science & Technology Studies (STS) is a broad collection of theoretical orientations and empirical approaches that share one overarching goal: to deconstruct 'objective' endeavours in technological and scientific domains, and reconstruct them as complex entanglements of humanity, discourse and materiality. In this sense, STS research sets out to challenge the workings and the assumptions of techno-scientific knowledge and recast them as the result of multiple social influences.

Various strands of STS prioritise different influences, such as the traditional forces of sociological structuration (Winner 1980), gender biases and engrained forms of bodily and social oppression (Haraway 1988), or more fluid and dialectic forms of social construction (Law 2010). Running throughout these different theoretical concerns is a recurrent emphasis on the notion of relation. This refers to the explicit and implicit linkages, tensions and dependencies that work to order things in conditions of uncertainty. Through the detailed description of these relational processes of order- ing and stabilisation – what Annamarie Mol called 'ontological politics' (1999) – STS researchers have produced case studies of how material technologies and forms of elite scientific knowledge are done and redone through practices and human/non-human alliances. We therefore seek to pos- ition the present paper within this tradition.

The implicit aim of this type of work is to highlight the plurality and uncertainty that undergird 'objective' techno-scientific facts, and to study how scientific things become real (i.e., are 'materialised') differently through enactments and practices. As Law (2010, 184) puts it, 'reality is not destiny … if we attend consistently to practices, then we start to discover alternative forms of materialisation'. One specific STS tradition that the specific paper pursues is the interrogation of specialised epistemic communities through engagement with technical aspects of their practical and discursive conventions, including their interactions in advanced laboratories and through scho- larly outputs (Latour and Woolgar 2013). Often leaning towards the sociology of knowledge, and inspired by diverse thinkers such as Foucault, Bourdieu and Kuhn, this approach has generated a wealth of case studies into the workings of various groups of scientists and engineers such as grav- itational wave physics (Collins 2010), high energy physics (Knorr-Cetina 1995), geneticists (Mack- enzie et al. 2013), statisticians (MacKenzie 1978) and so on. The present paper therefore attempts something similar in terms of unpacking the epistemological work of 'educational data scientists'.

More specifically, then, our study is aligned with recent STS-inflected research on algorithms and data. Indeed, this 'algorithmic turn' in the social sciences, spurred by a wave of public and economic interest in big data, automation and Artificial Intelligence, has reinvigorated interest in STS approaches. STS therefore offers a ready means of unpacking the claimed objectivity and operational complexity of algorithmic technologies of enumeration, classification and prediction (e.g., Crawford and Joler 2018; Dourish 2016; Kitchin 2014; Mackenzie 2017; Selbst et al. 2018). This recent work makes use of earlier STS authors like Collins (1993), Agre (1997), Forsythe (2002) and Suchman (1987) who problematised the development of AI and computer science from the 1960s onwards. Thus over the past 30 years, these authors have detailed key controversies about human reason and machine intelligence that now have great bearing on recent developments in the area of AI and education.

The key argument that we develop in this paper is that attempts to predict and automate aspects of educational performance through the application of AI technologies is party to complex forms of reductionism. While this is not in itself an original claim, there is something novel in the way computational methods like neural networks, 'owned' by a small elite of data experts driven by technical mindsets and commercial incentives, superimpose multiple layers of algorithmic complexity on stripped-down (and highly contentious) understandings of human learning. In this sense, we follow on from the work of Harry Collins (1993) who – drawing in turn on Dreyfus (1979) and Wittgenstein (1953) – observed that data representation in AI and computer science begets a paradox: as reality is gradually reduced to its commensurable constituents, the need for complex logical and mathematical abstraction grows stronger. In other words, the more things are simplified, the more complexity they require to remain real. This paradox comes about because expressing social life and culture in terms of algorithmic and statistical rules assumes a 'ceteris paribus' condition, i.e., that everything must stay the same for a computational system to remain internally coherent and capable to operate mechanically. However, this condition can only be upheld through an 'infinite regress' to underlying rules and requirements, in turn suffering from a tendency towards mathematical abstraction.

This point is reprised and further clarified in more recent work (Collins 2018), where Collins critiques bottom-up pattern recognition as the epistemological paradigm that underpins most current forms of applied AI. This paradigm is grounded in an inductivist logic – a distinctive mode of knowing – where predictions and generalisations are derived from past observations. It is also a very reductive paradigm because it relies, as suggested before, on a regressive modus operandi where patterns are assumed to be interpretable in the same standardised way across all cultures and contexts. There exists a different mode of knowing: the 'top-down model of interpretative sociology' (Collins 2018, 111) where knowledge depends on a productive engagement with 'forms of life' (Wittgenstein 1953). If we choose to operate in this mode, we no longer can interrogate mechanical induction (bottom-up pattern recognition in AI), without entertaining a different perspective on how knowledge is first developed and then endowed with cultural relevance. That is, without trying to understand 'how people live their lives in different societies' (Wittgenstein 1953). This alternative epistemic position allowed our empirical framework to emerge as we followed a specific predictive algorithm down the potentially infinite regress of neural activations and 'weightings', to arrive at a point where technical interrogation – a necessary but insufficient first step – became exhausted. In this sense, our falling short in this task is not an indication of limited technical knowledge, but evidence that the only productive way to investigate algorithms and AI is by looking around, rather than inside, increasingly opaque and unknowable black boxes. Hence, our focus shifted towards the analysis of 'forms of life': a range of sociological categories such as the cultural meanings hidden in digital materialisations of student learning (a specific dataset from a personalised learning platform), the economic interests of the predictive industry, and the disciplinary and professional entanglements of data science as a domain replete with controversies and uncertainties.

Methodology

Methodologically, the paper draws on the case study approach favoured in Science and Technology Studies (STS) where particular attention is paid to epistemic 'controversies'. These controversies take the form of disagreements and debates that illustrate how social aspects influence the otherwise seemingly 'objective' process of knowledge production in science and engineering domains (Law 2016). In this paper, we treat the academic debate around the DKT case as a minor controversy that can illuminate the sociotechnical factors involved in the production of knowledge about data science and AI techniques in education. In particular, our paper presents an interpretative analysis of the disciplinary debate itself. This approach therefore follows the work of Kelty and Landecker (2009) by focusing on a corpus of formal knowledge analysed as an ethnographic informant: 'something to be observed and engaged as something alive with concepts and practices not necessarily visible through the lens of single actors' (177). In terms of conventional educational research, this constitutes a relatively experimental and unconventional method. However, we argue that complex phenomena like 'algorithmic education' can only be studied by focusing on their digital and epistemic manifestations, and (it follows) through a pragmatic yet careful use of multiple methods that stretch well beyond a traditional reliance on interviews and other qualitative self-report methods.

In the remainder of the paper, we therefore take the eight articles that comprise the DKT controversy to develop an ethnographic understanding of the following three elements of a 'relational framework':

(1) *The educational data-set and broader digital 'learning' platform.* The first focus is on one of the specific 'educational' datasets involved in the DKT controversy. As mentioned earlier, the DKT debate considered six datasets in total. In this paper, we focus on the most prevalent dataset in the controversy that featured in six of the eight articles. This dataset is from a US-based Intelligent Tutoring system called ASSISTments, designed to teach (mostly) the topic of algebra. For this element of our analysis we examine the technical documentation relating to ASSISTments and treat the actual dataset as a digital-ethnographic artefact.
(2) *The AI method.* Second, we focus on the specific machine learning method implicit in the DKT debate: recurrent neural networks (RNNs). Here we attempt an interpretative reading of the papers, looking beyond their face value as 'objective' empirical reports. The interpretation focuses on tension between the 'new knowledge' that RNNs tried to discover in the data, and the 'existing knowledge' codified in the data environment as a result of an epistemic consensus amongst the educationalists that created it (i.e., pedagogic consensus about how the topic of algebra is learnt, and a pedagogic consensus about learning progression).
(3) *The cultural, discursive and economic aspects of data science in education.* Third, we examine the DKT debate as a specific instance of epistemic discourse. In particular, this involves analysing the patterning of a specific learning-related keyword ('performance') across the papers as indicative of problematic cultural assumptions. We then place this discursive contestation in the context of competitive relations that the DKT studies mediated between universities, corporate entities and the six digital educational datasets.

Weak AI: hype, backlash and the complexities of 'learning from data'

Before examining each of these three elements, it is important to develop a good working understanding of the AI method that is under scrutiny here. In particular, we outline the specific machine learning method implicit in the DKT debate: recurrent neural networks (RNNs), which is one of many methods that can be used in AI. As shall be clear when we go on to consider the three elements, understanding the logic of this method and the way it differs from other machine learning approaches is an important pre-requisite to making sense of the DKT controversy.

In order to understand RNNs, it helps to examine what preceded them. The 60 year old field of AI is characterised by periods of hype followed by backlash, which involved two well-documented 'AI winters' following spells of enthusiasm and vigorous research activity. The first backlash of the 1970s resulted from the failures of so-called 'Good Old Fashioned AI' (GOFAI), also known as 'symbolic AI', which emerged from work in mathematical logic where a number of principles derived from human reasoning were theorised and formalised. During the 1950s and 1960s, many researchers attempted (with limited success) to encode these principles in computational systems to simulate autonomous intelligence. This led to reduced funding and a subsequent sharp decline in research activity. However, the subsequent development of 'expert systems' in the 1980s was seen to herald a resurgence of AI research. These systems aimed to emulate the decisional processes of experts by applying procedural (IF/THEN) logic methods to defined knowledge bases that had been devel- oped through consultation with human domain experts.

While innovative, the field of expert systems stalled after a decade or so, prompting the second AI winter which lasted well into the early 1990s. However, this was followed by a resurgent phase of radically different AI that abandoned the previous emphasis placed on abstract formal logic principles and began instead to rely on methods of statistical inference as well as inductive and abductive reasoning. This phase of so-called 'weak AI' is exemplified by developments in the field of machine learning, where computational methods are used to automate specific tasks of classification and prediction. The societal application of machine learning over the past 20 years have proven numerous – as reflected in developments in online shopping, face recognition, self-driving cars and cancer diagnosis. In terms of the specific focus in the present paper, there has been growing educational interest in the use of machine learning to model student performance and behaviour.

Broadly speaking, machine learning approaches follow a relatively straightforward 'function fitting paradigm' (Hastie, Tibshirani, and Friedman 2009):

$$Y = f(X) + \varepsilon.$$

Described in plain English, this equation posits how an outcome variable 'Y' (e.g., in terms of the DKT controversy, a prediction of student success in completing an algebra problem) is the result of a function 'f' (i.e., a specific type of mathematical operation such as addition or division) applied to a predictor variable 'X' (e.g., the number of times the student seeks help while trying to solve the algebra problem). The equation takes also into account the likelihood that the model will have errors ('ε').

In terms of this function fitting paradigm, then, the purpose of machine 'learning' is to figure out what f actually does. This can be a complex process that stretches well beyond simple arithmetic. In order to learn *f,* a human agent must assemble a 'training set' of observations. In so-called supervised learning, the data scientist will use a pre-defined system to label (and therefore categorise) the inputs and the outputs associated with the phenomenon that is being modelled. For example, the process of learning algebra can be modelled according to the established consensus in maths education and psychology. This model can be broken down in various codified steps which refer to specific inputs and outputs. Once created, this training set can be fed to an algorithm (a computer program) which 'learns' the various interactions and permutations between the inputs and outputs included in the data set. If the approach succeeds, the algorithm will figure out what *f* does, and will then be able to predict an outcome whenever new, unseen instances of the same types of data are encountered.

In unsupervised machine learning, algorithms operate without pre-defined labels and, according to one of the most popular technical textbooks currently available, 'experience a dataset containing many features, then learn useful properties of the structure of this dataset' (Goodfellow, Bengio, and Courville 2016, 105). The notion of an algorithm 'experiencing' something is an anthropomorphism not uncommon in the machine learning literature; it evokes a process of independent knowledge discovery whereby meaningful categories can be constructed in an 'agentic' fashion. It is precisely in this arena – unsupervised knowledge discovery – that neural networks began outperforming other

machine learning methods. Indeed, unsupervised learning played a key role in the 'renaissance' of deep learning in the mid-2000s, when data scientists discovered that neural networks could be pre-trained in an unsupervised fashion, thus making them more effective in supervised tasks (Goodfellow, Bengio, and Courville 2016, 528).

The operational aspects of neural networks are loosely based on an abstracted understanding of human learning as a bottom-up (inductive) process that relies on observation, experimentation and the dynamic adaptation to the new information extracted from the data. This process aims to approximate a model of the biological brain, where signals from multiple inputs are combined, 'weighted' and then trigger parallel neural activations once they pass a certain threshold. Following this approximate model, the standard deep learning model involves the construction of artificial neural networks that consist of layers of sparsely connected units through which data and the associated errors circulate, while the predictive or classificatory task is learnt. These can be seen in terms of three distinct layers: (i) an input layer, (ii) a middle hidden layer where intermediate computations take place, and (iii) an output layer (see Figure 1).

At this point, many readers' attention might understandably be piqued by the notion of the 'middle hidden layer'. In brief, the hidden layer provides a buffer where the errors that inevitably emerge during the training process can be sent back and propagated through the network. Hastie, Tibshirani, and Friedman (2009, 395) describe this as a 'forward and backward sweep over the network' in which errors are first spread out and then recombined to compute the output layer. This optimisation technique is known as 'Stochastic Gradient Descent', essentially showing how the patterning of these errors change in tandem with changes to the weighting in the network. This is an incremental process that minimises 'loss', i.e., a measure of how effective the model is at predicting a single case (e.g., an input-output pairing). The aim of the optimisation technique is therefore to stabilise on a set of weights that, on average, have low levels of loss across the entire dataset. The process concludes when a human agent decides that satisfactory results have been obtained.

One particular type of neural network used in the DKT controversy is the 'recurrent neural network' (RNN). This extends even further the functionalities of other so-called vanilla networks by allowing algorithms to operate with entire sequences of observations (rather than individual observations) that are mapped over long periods across the input and output layers. The recursive, temporal nature of this procedure aims to capture the regularities through which past observations shape

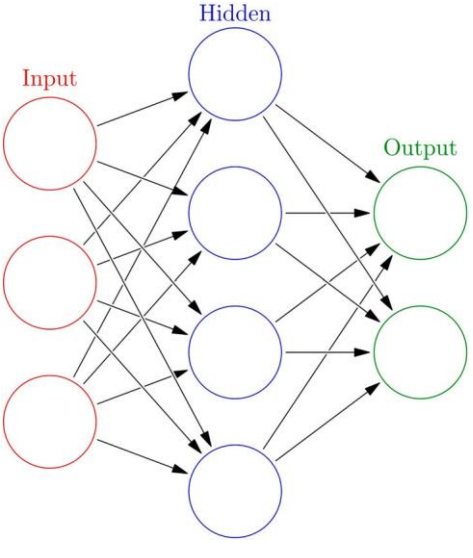

Figure 1. An artificial neural network. Source: wikipedia.org

the likelihood of future events. Recurrent Neural Networks are therefore described in the original DKT study as 'a family of flexible dynamic models which connect artificial neurons over time' (Piech et al. 2015, 3). In the context of the example used before of learning algebra, the advantage of a RNN is the ability to 'remember' past performance from a student and use this information to produce predictions at a 'much later point in time' (Piech et al. 2015).

Recurrent neural networks (RNNs) have generated a significant amount of interest amongst the AI community in recent years. Formalised decades ago to assist complex computational tasks such as speech recognition (e.g., Rumelhart, Hinton, and Williams 1986), they have gained popularity as appropriately large datasets became readily available alongside powerful computational resources. In particular, RNNs have been extensively researched within DeepMind: the flagship Google research program on AI (Graves and Jaitly 2014).

One final point to make before we go on to consider the development of RNNs in the context of the DKT controversy relates to explainability (or lack of it). Like vanilla networks, RNNs achieve their outcomes through the back-propagation of errors – i.e., updating the values in the model as new observations come in. However, unlike vanilla networks, RNNs can 'remember' past information given the fact that connections between hidden units often exhibit a time delay. This enables RNNs to 'discover temporal correlations between events that are far away from each other in the data' (Pascanu, Mikolov, and Bengio 2013, 1310). For non-experts, the recursive and temporal nature of the process is difficult to grasp, and it remains largely opaque even in the specialist literature (Zeiler and Fergus 2014), where it is generally accepted that it is nigh-on impossible to interpret weights and neural activations. As a result, RNNs are deemed capable of discovering unexpected features of the data that appear confusing, other-worldly and/or 'hallucinatory' (Perez 2018), given the approximate understanding of how they were achieved.

The disagreements and confusions of educational data science

Having provided a broad introduction of the AI method and RNNs, we will now examine relationally its implication in the DKT case study, starting by the relations between the online tutoring system, its underlying assumptions about knowledge modelling in algebra, and the resulting data.

(i) the digital environment and its dataset

Intelligent Tutoring Systems (ITSs) have a long history as educational technologies, with the first developments dating back to the 1960s and 1970s (e.g., Carbonell, Michalski, and Mitchell 1983). ITSs have evolved significantly over the past decades and, nowadays, they resemble online software platforms that collect large amounts of student data. One of the most successful ITSs in recent years, widely adopted in the US education sector, is called ASSISTments. This system was developed by the Worcester Polytechnic Institute and is available to teachers free of charge. It is officially described as a 'collaborative ecosystem' (Heffernan and Heffernan 2014) which has involved teachers, researchers and computer scientists working together to produce collections of problem sets and scaffolding materials to teach high school level maths, alongside the modelling of other subject domains such as physics, chemistry and English grammar.

ASSISTments was launched in 2003 in order to automate the remedial instruction for middle and high school students preparing for high-stakes State examinations. The developers wanted to build 'an online system where students would practice the released MCAS (Massachusetts Comprehensive Assessment System) items, with tutoring on how to work out problems offered to students who got the problem wrong' (Heffernan and Heffernan 2014, 475). This emphasis on the 'Pass/Fail' binary typical of high-stake exams is computationally reinforced by ASSISTment's underlying model which is termed 'knowledge tracing' (Corbett and Anderson 1994). This approach assumes a simple two-state model of human knowledge, where student performance is observed in order to estimate (in a binary fashion) the presence or absence of knowledge on a predefined skill. Students are deemed

to have learnt a skill (such as adding and subtracting integers) when they get three answers right in a row with no help and no mistakes. The exercises and the supporting materials are presented to students through a traditional software interface (see Figure 2), as they progress along a pre-defined trajectory towards increased mastery of a range of skills.

Data gathered from ASSISTments in the school year 2009−2010 has been made available online on a free-to-use basis (https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data). Crucially, this dataset was used as a train/test split in the DKT studies (i.e., one half was used to train the algorithms, the other half for testing purposes). It is downloadable as a CSV file and can be imported in various analysis tools and programming environments for statistical computing, such as Microsoft Excel, SPSS and R. Once opened (Figure 3), the dataset exhibits the typical charac- teristics of statistical tabulation with a total of 401,756 rows, each indicating an assignment done by a student, with all assignments in chronological order and each student tagged with a specific ID.

This dataset is an insightful digital materialisation in its own right, demonstrating the extent of data capture performed in ASSISTments and other similar systems. It contains detailed information about the school in which the task was performed, the ID of the teacher who assigned the problem, the skills associated with each problem, the number of student attempts on a problem, the number of times help was accessed and several other rows about individual performance aspects.

When examined closely, there are numerous learning-related 'stories' in these data: trajectories of educational achievement and struggle, signs of student fragility and maths-related anxiety, as well as school-level factors relating to the cultural contexts where these exercises were performed. We can see for example, student #70363 attempting 14 times to complete an exercise about box and whisker plotting in descriptive statistics. We can also see student #79781 making 66 attempts to answer the same task and then eventually answer 'I have no idea'. Then again, are students #78897, #88129, #78415, #88127 and #88129 − all tackling various exercises ranging from fraction conversion to

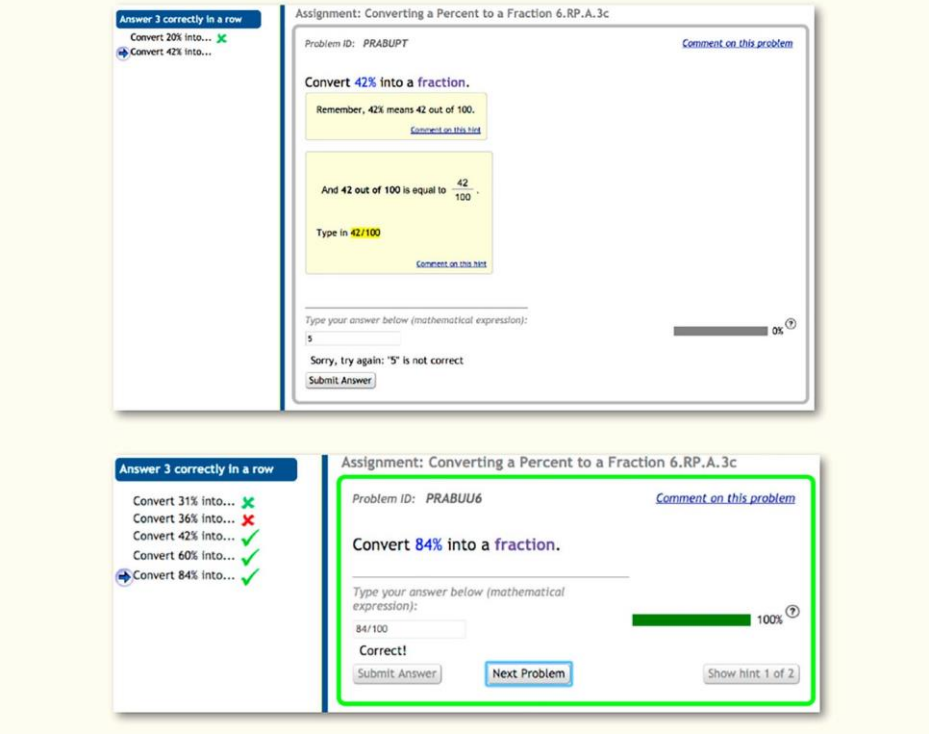

Figure 2. The ASSISTments user interface.

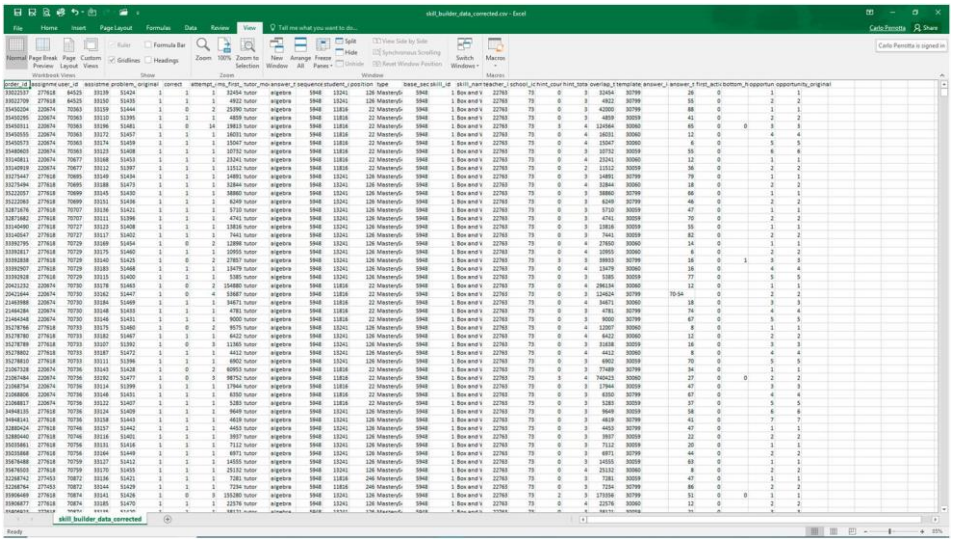

Figure 3. A digital-ethnographic materialisation: the ASSISTments dataset.

the addition and subtraction of decimals, and eventually all independently agreeing on the same final response: 'Your Mom'.

In contrast, the work of abstraction carried out by predictive modelling and data science is not interested in these cultural and emotional aspects of the dataset, dismissing them as background noise interfering with the primary task of predicting 'knowledge states', following an underpinning model that relies on the involvement of human experts who manually label the skills required for a given exercise. Thus, in terms of how the ASSISTments dataset is used in the DKT study, the labelling of skills and their encoding into the model preceded the algorithmic process of machine learning, while the role of human agency in this process reminds us how machine learning is:

> Both a form of automated knowledge production and also one shaped by people working in specific labour conditions, within institutional frameworks, according to professional commitments, worldviews and disciplinary theories about the ways in which the world works. (Williamson 2017, 116)

Aside from these issues of the human labour that underpins these seemingly automated processes, the knowledge model that frames (and curtails) the ASSISTments dataset is also notable in its representation of 'learning'. In particular, the roots of ASSISTments as a tool designed with summative testing in mind become visible in this entanglement between automation and human agency. This results in a problematic educational assumption 'baked' into the data: knowledge about algebra is based on a 'all or none' (Lindsey, Khajah, and Mozer 2014) learning binary. The key point is that this binary construction of knowledge is a distinct design choice which reinforces a pre-existing educational philosophy that can be traced back to the tool's origins in the pass/fail mentality of high stakes testing. Crucially, this design choice also results from a process of computational performance enacted by knowledge tracing's method of choice: hidden Markov model, a probabilistic approach that predicts knowledge states according to a base-2 logic of 0 (knowledge is present) or 1 (knowledge is absent). It is important to consider the compromised and partial nature of this entanglement, which creates the very conditions in which 'traditional' knowledge tracing can successfully model progression in closed software environments. At least, this was the case until the 'mini-debate in the educational data-mining world' (private communication with one of the authors involved the DKT studies) that followed the application of recurrent neural networks.

(ii) The AI method meets the data: black boxes, tensions and glitches

The distinctive accomplishment of RNNs applied to ASSISTment data is that they deal with skills that have been simultaneously organised in temporal sequences, rather than individual skills. The process in figure 4 is a simplified visual representation of the RNN architecture applied to the ASSISTments data, based on the original 2015 study, as well as the subsequent ones which provided counterevidence and responses. The most crucial part of the process is the hidden layer where one is, quite literally, forced to imagine the existence of a transformative process through which the tabulated data in Figure 3 is 'exploded in a multidimensional vector-space, to the point that is no longer representable diagrammatically' (Mackenzie 2017, 73) (Figure 4).

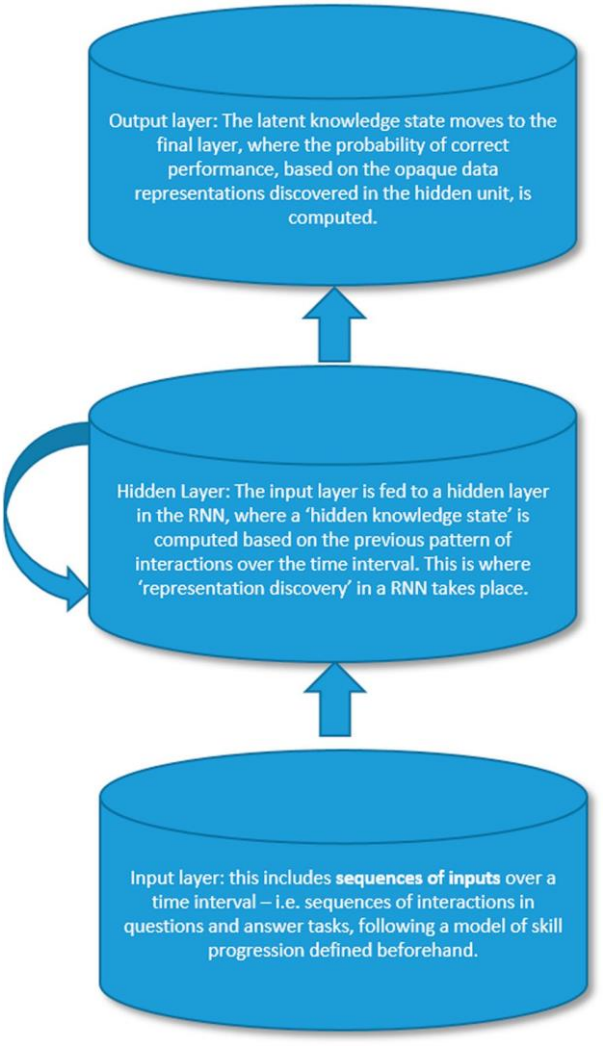

Figure 4. What the RNN did to the ASSISTments data.

The hidden neurons in the middle layer are connected in a recursive manner, that is, information is propagated over time in such a way that each hidden unit connects back to all other hidden units. Such process does not require pre-existing data representations, since the algorithm is able to discover patterns through recombination and optimisation. As mentioned earlier, this discovery process is the most distinctive element of deep learning, unsurprisingly generating a great deal of dis-

cussion in academia and the public sphere. While it may appear 'magical' and it is often described as such in the mainstream media (Naughton 2018), this mechanism relies in fact on a probabilistic logic that targets increasing degrees of plausibility. Plausibility is a domain-dependent concept that may be beneficial in areas like image processing and recognition, where 'hallucinated outcomes' are expected by-products of the iterative approximation process of deep learning, but it is more problematic when educational constructs are involved.

In this sense, an important tension at the heart of the DKT controversy became manifest in one of the response papers (Khajah, Lindsey, and Mozer 2016), when the authors noted that the application of deep learning to the ASSISTments knowledge model discounts the hand-crafted features defined by human experts on the basis of domain expertise, to favour instead the discovery of unintelligible (but 'plausible', not unlike hallucinations) representations through the recursive propagation of information in the hidden layer of a neural network.

> The features discovered by deep learning exhibit a complexity and subtlety that make them difficult to analyze and understand (…) no human engineer could wire up a solution as thorough and accurate as solutions discovered by deep learning. (…) (it) discards hand-crafted features in favor of representation learning, and often ignores domain knowledge and structure in favor of massive data sets and general architectural constraints on models. (Khajah, Lindsey, and Mozer 2016:94)

This observation signals another instance of the DKT scholarly discussion overflowing into the broader public controversy about automation and human-machine equilibrium that was mentioned in the introduction, thus bringing into view the fundamentally non-human nature of automated mechanical induction. Indeed, representation discovery makes sense in a perceptual, sub-symbolic image classification task where the relationships between the foundational elements of a digital image (pixels) can be decomposed, learnt and then recombined in manifold ways without human intervention. RNNs have proved quite capable to discover the underlying principles that govern this re-combinatorial flexibility. The academic discipline and the educational practice of human learning do not exhibit the same degree of flexibility, as their constructs are largely symbolic abstractions that reflect an empirical and discursive consensus among human experts about cognition and domain knowledge.

We are dealing here with a crucial tension between theoretical and operational 'versions' of learning, both problematic from an educational point of view. On one side, the representation discovery of neural networks wants to proceed inductively from the data, leading to the paradoxical conclusion that deep learning might discover 'unknown', yet plausible, algebra skills that, presumably, might even surprise mathematicians and maths educators. On the other side, we have a highly structured process of learning and mastering a form of knowledge, which wants to proceed deductively from a binary and narrow knowledge model inspired by a desire to help students succeed in high stakes tests. Unfortunately, not much was made of this 'philosophical' tension between two forms of reductionism, as the DKT debate became absorbed in the infinite regress of challenging prediction scores and tweaking computational models. In fact, the main 'plot twist' in the debate could not be any further removed from these theoretical and philosophical considerations. In 2016, Xiong et al. (2016) discovered that the original study's inflated prediction scores were, in part at least, the result of flaws in the ASSISTments open data, determined by the 'unclear transformational rules' (Xiong et al. 2016, 550) of RNNs, which exhibited a strange tendency to randomly duplicate rows of data as part of its recursive, unintelligible work of 'discovery'. While this dealt a serious blow to the orig- inal 85% claim, it also prompted the authors to issue a reminder about the importance of basic data hygiene when using such increasingly opaque methods: 'while we advance new algorithms and fine tune their parameters, we should also consider (and, if possible, report on) the robustness of the algorithms to common data glitches' (Xiong et al. 2016, 550).

(iii) Cultural, discursive and economic aspects of data science in education

Thirdly, then, we consider the cultural and economic underpinnings of the educational data science community involved in the DKT controversy and the field of predictive AI research and development in general. These are areas where methodological proficiency and analytic power have great currency, and where professional reputations and careers are founded on the capacity to subject radically different forms of data to comparable and scalable forms of computational analysis. As Adrian Mackenzie notes, the career trajectories of 'machine learners' regularly imbue a cultural motif of data as a computational challenge (Mackenzie 2013). In this sense, online education is just one domain among many where this challenge can be engaged with. Here we can see, then, how the DKT controversy was driven (in part) by data scientists seeking a fertile context where they might successfully apply general principles of predictive modelling. As they moved into the DKT controversy, the data scientists who deployed RNNs onto educational datasets brought two qualities into the work that are distinct from what might be conventionally considered 'educational' concerns. First was a functional reliance on a 'black box' approach to mathematical modelling that treats prediction as an optimisation task based on opaque (deep) recursive mechanics. Second was a preoccupation with predictive performance as an indicator of their own professional accomplishment.

These underpinning qualities are exemplified by one of the papers written in response to Piech et al. (2015), where the authors stated that they were 'driven by both noble goals (testing the reproducibility of scientific findings) and some selfish ones (how did they [Piech and colleagues] do so much better at predicting student performance)?!' (Xiong et al. 2016, 545). Indeed, if we trace the use of this key word ('performance') across all eight published studies then some important cultural assumptions implicit in the DKT debate are highlighted. For example, the term 'performance' (alongside various stemmed words: perform, performed, outperform, performing, performs) appears 213 times in the texts across the eight empirical articles. Examining the use of these words across the eight DKT articles reveals a rhetorical relationship between two themes: 'performance' of the algorithmic model (how good it is at predicting) and the 'performance' of students (what is being constantly monitored and automatically predicted). This is apparent in the following examples:

> In Deep Knowledge Tracing a recurrent neural network was trained to predict student responses and was shown to outperform the best published results (…) We found that IRT-based methods consistently matched or outperformed DKT. (Wilson et al. 2016, 539)

> Recently, with a surge of interest in deep learning models, DKT [12], which models student's knowledge state based on an RNN, has been shown to outperform the traditional models, such as BKT and PFA. (Yeung and Yeung 2018, 1)

> When we replicated simulations (… .) we obtained significantly better performance: an AUC of 0.73 versus 0.67 on ASSISTments. (Khajah, Lindsey, and Mozer 2016, 98)

> A student who performed poorly on the last trial because they were distracted is likely to perform poorly on the current trial. (Khajah, Lindsey, and Mozer 2016, 97).

> The system continuously monitors the student's performance, updates the knowledge states and based on that takes further decisions. (Lalwani and Agrawal 2017, 448)

As these excerpts illustrate, the eight DKT articles exhibit a tendency to extend the instrumental notion of predictive 'performance' as computational challenge into an associated understanding of the educational 'performance' of students. In other words, student 'learning' quickly gets conflated with student 'performance', which itself is positioned as a matter of algorithmic tractability based on temporal sequences of inputs and outputs. This discursive entanglement of predictive performance as an attribute of the model, and learning performance as an attribute of the students eventually stabilises around a 'granular' view of education as a score-driven dynamic and collection of machine-readable signals:

> Given three exercises each of skills A and B, presenting the exercises in the interleaved order A1−B1−A2−B2−A3−B3 yields superior performance relative to presenting the exercises in the blocked order A1−A2−A3−B1−B2−B3. (Khajah et al: 97).

> The task of knowledge tracing can be formalized as: given observations of interactions $x_0 \ldots x_t$ taken by a student on a particular learning task, predict aspects of their next interaction x_t+1 [5] (…). The authors found that RNNs can robustly predict whether or not a student will solve a particular problem correctly given their performance on prior problems (Wang et al. 2017, 325).

This distinctive motif of learner performance as computational performance can be considered as the main 'discursive work' enacted across the DKT studies, as their authors competed over prediction scores and contested the 'stunning performance advantage' (Khajah et al: 94) of deep learning. As such, the theme of performance was an important linguistic/ideological device of 'translation' that created competitive convergences (Callon 1980, 211) between the following actors:

(a) the datasets from a handful of online learning environments: ASSISTments (open), Khan Academy (proprietary), Knewton (proprietary), Funtoot (proprietary), Hour of Code (open) and the Carnegie Tutor Geometry dataset (open).
(b) a small group of academic institutions: Stanford University, University of Colorado Boulder, Hong Kong UST, Worcester Polytechnic Institute;
(c) the corporate entities who directly employed some of the data scientists involved in the debate: Knewton, Funtoot, Google.

The relationships between these analytic entities/actors as evident from the eight DKT articles is illustrated in Figure 5. In this visualisation, the ASSISTments dataset and the study that 'triggered' the discussion (Piech et al. 2015) occupy a central position due to having the largest number of connections. The study from Lalwani and Agrawal (2017) is also distinct as the sole study that relied exclusively on a proprietary dataset from Funtoot, a popular Bangalore-based education technology company. Figure 5 therefore illustrates how applied AI became an 'educational thing'. This occurred through the competitive relations that the eight DKT studies mediated between academia, the corporate sector and, crucially, a handful of digital educational datasets that shared similar assumptions about knowledge modelling and originated from platforms competing for market share in the K-12 EdTech sector.

The ability to map these mediated relations in this manner therefore adds depth to our understanding of the DKT controversy. While ASSISTments and Khan Academy are classified as non-for-profit entities, the involvement of large, for profit companies like Knewton and Funtoot points to their strategic research interests in predictive modelling, as a potentially integral part of their portfolios of personalised, adaptive and 'intelligent' educational products. The involvement of Google Brain (a leading deep learning research unit) is also significant in flagging Google's intention to shape the deployment of applied AI in education as one of its various domains of competitive activity.

Of course, these relations evident within the DKT articles are only the 'tip of the iceberg' of deep learning as a much larger phenomenon shaped by market forces. The techniques developed through the DKT studies are highly portable and scalable across various domains of society. Thus this map of educational AI is likely to be replicated in health, criminal justice, and multiple similar cases where the same deep learning methods act as connective tissue between ensembles of academia and economic interests. Indeed, the past few years have witnessed the rise of general-purpose predictive infrastructures with large technology companies developing various cloud-based or distributed AI/deep learning frameworks. The most notable development in this regard is Google's Tensorflow – released under an Open Source licence in 2015 and rapidly established as a market leader. Indeed, the DKT case study suggests that Google's expertise was instrumental in enabling their particular brand of deep learning into the education domain, with Tensorflow chosen as the framework to build and train the models in two of the papers involved in the DKT controversy (Xiong et al. 2016; Zhang et al. 2017).

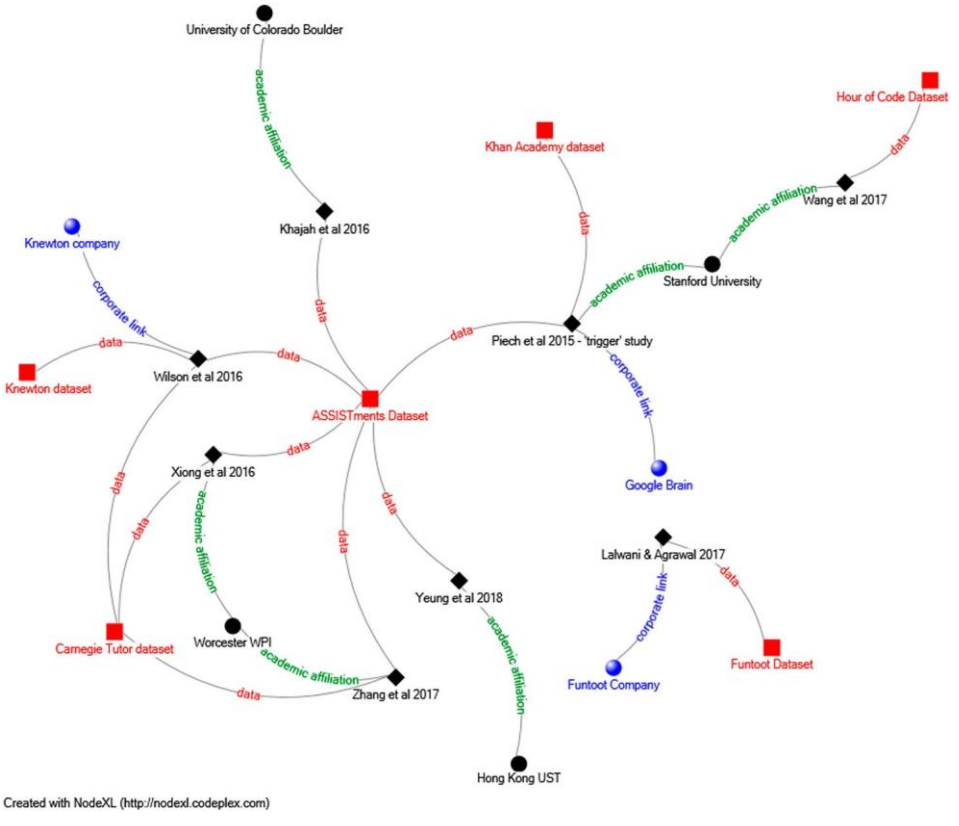

Figure 5. Mapping the Deep Knowledge Tracing controversy – relations between datasets, data scientists, academia and corporate tech. The main 'nodes' are the individual studies and the datasets used as part of the experiments.

Conclusion

While an initial broad-brush encounter with deep learning techniques in education, this article brings a number of different perspectives and points of analysis to the study of AI in education. In particular, our descriptions highlight a series of important relations. This includes relations (and tensions) between 'versions' of learning, relations between models and datasets, relations between academic and corporate entities, and relations between epistemic cultures and economic interests. The analysis of these relations suggests that the trajectory of this particular predictive mod- elling approach in education was a complex sociotechnical affair. This is consistent with the STS approach that informed the study.

In the spirit of this research tradition, the paper has raised the following observations about the DKT controversy:

- Some aspects of this DKT controversy are best seen as epistemic in a traditional sense, related to how the fields of data science, educational data mining and learning analytics interact with each other to develop knowledge through research, publications and conferences.
- Some aspects of the DKT controversy could be described as semiotic, as the meaning of a specific predictive paradigm was negotiated across a number of studies, in relation to (and in tension with) aspects of human cognition and learning.
- Yet other aspects of the DKT controversy are distinctly sociocultural, as the overriding concern for predictive power in the deep knowledge tracing debate is consistent with the values, the discourses and the economic interests of the data science and AI community/industry where the authors of these 'educational' studies originated.

- Above all, the DKT debate exposed the problems and uncertainties surrounding the functional complexity of recurrent neural networks and the applicability of automated representation discovery to the educational domain.

As AI continues to develop different predictive paradigms (including variations of the deep learning techniques described in this paper), the challenge facing education research over the 2020s will be to interrogate their assumptions and unresolved tensions in a critical way, avoiding a wholesale, unproductive closure. These are no wholly bad or incorrect developments to be bringing to bear on education. For example, while there is value in critiquing ASSISTments' original knowledge tracing model as limited and constrained, it is also worth acknowledging that it is based on a sociologically 'real' form of knowledge about cognition and learning. By real we mean it is the result of meaningful patterns and 'scientific generalizations, created through socially agreed choices about what was to count as sound observation and what as unsound' (Collins 2018, 112).

Ultimately, this is where our own analysis departs from extreme forms of 'relational ontology' in STS and turns an eye toward critical realism (Archer 2010). While we fully accept that all entities (including theories of cognition and learning) result from a process of becoming and are not simply endowed with 'substances', once they stabilise in accordance with politics *as well as* criteria of empirical credibility, they should be ascribed a distinctive, more robust status in the multifaceted social debate over truth and knowledge.

This latter point is particularly pertinent when interrogating the process of knowledge discovery in deep learning. In particular, it can be mobilised to support a critical argument against automated data representation in education, i.e., the fact that theories of learning cannot, after all, be 'discovered' by algorithms. Once we accept that the input/output dynamic of proven computational models of cognition can have a stable basis (i.e., they are not just plausible, but 'real' in a sociological way, i.e., as stable forms of enculturation and socially shaped knowledge), the problem becomes about the constrained nature of an algorithmic framing, rather than the framing *per se*. It becomes, in other words, a matter of 'heterogeneous engineering': a debate on how we can 'redraw the boundaries of abstraction to include people and social systems as well, such as local incentives and reward structures, institutional environments, decision making cultures and regulatory systems' (Selbst et al. 2018, 9).

In the case of 'regular' knowledge tracing, this process of boundary redrawing is possible – but this will require a productive dialogue between the fields of educational assessment, computer science and critical educational research. Such dialogue might help fortify the field of education research against adopting a problematic inductivist position where the prospect of something along the lines of 'hallucinated school algebra' can be contemplated as an educational possibility (at least in theory). Indeed, this is a possible end-point that 'deep' knowledge tracing entertains through its reliance on bottom-up pattern recognition and representation discovery. Having followed closely the DKT debate, we can safely conclude that the achievability of such 'discoveries' in the education domain remains highly contestable (if not something that deserves to be rejected outright). There- fore, we must remain vigilant against politically and commercially motivated attempts to downplay these contentious aspects and trade upon the mysterious and other-worldly connotations of AI- based speculation.

In conclusion, this article provides an account of how AIED was 'assembled' through a minor empirical debate about the use of deep learning with educational datasets. It also provides some methodological suggestions as to how social scientists can go about studying and critiquing similar episodes of unsettled knowledge-making.

While we do not claim that our relational framework covers all possible lines of enquiry, we believe it represents a starting-point for further research by highlighting three important units of analysis: (i) the AI methods themselves, (ii) the digital platforms that produce the educational datasets used to train algorithms, and (iii) the 'social life' of computation in education as a site of epistemic, ideological and economic contestation. The DKT debate therefore provided us with an

opening through which the sociotechnical nature of predictive modelling was manifest in a form that could be meaningfully investigated.

As is the case with critical research on data-driven education in general, the primary focus of this paper lies in: (i) questioning the attributes of scientific objectivity and neutrality ascribed to these technological systems, and then (ii) exploring ways in which sociality and diversity can be reinjected in them. In this sense, one of the main themes highlighted by our analysis is the value of combining a precise examination of algorithms with a plural and experimental use of digital-ethnographic methods. Another concluding suggestion relates to the importance for critical education research to engage in an interdisciplinary dialogue with cognitive science and data science. Such a dialogue must build on the acknowledgement that the models of cognition and learning encoded in digital learning environments are meaningful analytic entities that cannot be glossed over in the pursuit of social science theorising.

As a final observation, we wish to remind the reader that we have examined a particular instance of AIED in the form of the use of computational methods to categorise and predict performance in structured learning environments. This specific version of AIED is a reflection of the so-called 'personalised learning' trend in education, which has been abundantly critiqued as an individualistic discourse that overemphasises market-inspired logics and is shaped by the interests of technology companies through metrics, automation and the pervasive collection of data. While not disagreeing with these existing critiques, the relational analysis developed in this paper suggests that the way in which these phenomena actually 'come together' is a nuanced process open, in theory, to alternate social shapings. As such, the continued application of AI methods to education is not something to be rejected outright, but something that is well worth engaging with on its own terms and contesting. In this sense, AI in education needs to be talked about more often in controversial and circumspect terms, rather than accepted as a computational fait accompli.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Carlo Perrotta is senior lecturer in digital literacies in the Faculty of Education at Monash University. His background is in sociology and social psychology. Carlo has published on a variety of topics related to digital technology in education, including the social and political accountability of algorithms in education, the ethical use of video games in schools and socio-material analyses of digital education. His research as PI and Co-I has been funded by leading international bodies such as the European Commission, the ESRC, the Society for Educational Studies, as well as private donors (e.g. Microsoft and Cisco Systems).

Neil Selwyn is a professor in the Faculty of Education, Monash University. His research and teaching focuses on the place of digital media in everyday life, and the sociology of technology (non)use in educational settings. Neil has written extensively on a number of issues, including digital exclusion, education technology policymaking and the student experience of technology-based learning. He has carried out funded research on digital technology, society and education for the Australian Research Council (ARC), Economic and Social Research Council (ESRC), British Academy, the Swedish Research Council for Health, Working Life and Welfare (FORTE), the BBC, Nuffield Foundation, the Spencer Foundation, Gates Foundation, Microsoft Partners in Learning, Becta, Australian Government Office of Learning and Teaching (OLT), Australian Communications Consumer Action Network (ACCAN), Save The Children, Centre for Distance Education, the Welsh Office, National Assembly of Wales and various local authorities in the UK.

ORCID

Carlo Perrotta http://orcid.org/0000-0003-3572-0844
Neil Selwyn http://orcid.org/0000-0001-9489-2692

References

Agre, P. 1997. “Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI.” In *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, edited by Geoffery C. Bowker, Susan Leigh Star, Les Gasser, and William Turner, 131–157. Mahwah, NJ: Erlbaum.

Archer, M. 2010. “Critical Realism and Relational Sociology: Complementarity and Synergy.” *Journal of Critical Realism* 9 (2): 199–207.

Callon, M. 1980. “Struggles and Negotiations to Define What is Problematic and What is Not.” In *The Social Process of Scientific Investigation*, 197–219. Dordrecht: Springer.

Carbonell, J. G., R. S. Michalski, and T. M. Mitchell. 1983. “An Overview of Machine Learning.” In *Machine Learning, Volume I*, 3–23. Burlington, MA: Morgan Kaufmann.

Collins, H. M. 1993. *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: MIT Press.

Collins, H. 2010. *Gravity’s Shadow: The Search for Gravitational Waves*. Chicago: University of Chicago Press.

Collins, H. 2018. *Artifictional Intelligence: Against Humanity’s Surrender to Computers*. Cambridge: Polity Press.

Corbett, A. T., and J. R. Anderson. 1994. “Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge.” *User Modeling and User-Adapted Interaction* 4 (4): 253–278.

Crawford, K., and V. Joler. 2018. "Anatomy of an AI System." Accessed 19 February 2019. https://anatomyof.ai

Dourish, P. 2016. “Algorithms and Their Others: Algorithmic Culture in Context.” *Big Data & Society* 3 (2): 1–11.

Dreyfus, H. L. 1979. *What Computers Can’t Do: The Limits of Artificial Intelligence*. Vol. 1972. New York: Harper & Row.

Elish, M. C., and D. Boyd. 2017. “Situating Methods in the Magic of Big Data and AI.” *Communication Monographs* 85 (1): 57–80.

Forsythe, D. 2002. *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Redwood, CA: Stanford University Press.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. Cambridge, MA: MIT press.

Graves, A., and N. Jaitly. 2014. Towards End-to-end Speech Recognition with Recurrent Neural Networks. *In Proceedings of the 31st International Conference on Machine Learning* (ICML 2014), Beijing, China, June 21–June 26, 2014, pp. 1764–1772.

Haraway, D. 1988. “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14 (3): 575–599.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* – 2nd ed. New York: Springer.

Heffernan, N. T., and C. L. Heffernan. 2014. “The ASSISTments Ecosystem: Building a Platform That Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching.” *International Journal of Artificial Intelligence in Education* 24 (4): 470–497.

Kelty, C., and H. Landecker. 2009. “Ten Thousand Journal Articles Later: Ethnography of ‘The Literature’ in Science.” *Empiria: Revista de Metodología de Ciencias Sociales* 18: 173–192. Accessed 14 February 2019. http://dialnet.unirioja.es/servlet/articulo?codigo=3130617.

Khajah, M., R. V. Lindsey, and M. C. Mozer. 2016. How Deep is Knowledge Tracing? *Proceedings of the 9th International Conference on Educational Data Mining*. June 29–July 2, 2016, Raleigh, North Carolina, USA, pp. 94–101.

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: SAGE.

Knorr-Cetina, K. 1995. “How Superorganisms Change: Consensus Formation and the Social Ontology of High-Energy Physics Experiments.” *Social Studies of Science* 25 (1): 119–147.

Lalwani, A., and S. Agrawal. 2017. Few Hundred Parameters Outperform Few Hundred Thousand. *Proceedings of the 10th International Conference on Educational Data Mining*. June 25–28, 2017, Wuhan, Hubei, China, pp. 448–453.

Latour, B., and S. Woolgar. 2013. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Law, J. 2010. “The Materials of STS.” In *The Oxford Handbook of Material Culture Studies*, edited by Dan Hicks and Mary C. Beaudry, 173–188. Oxford: Oxford University Press.

Law, J. 2016. “STS as Method.” In *The Handbook of Science and Technology Studies (No. 3rd)*, edited by E. J. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, 31–55. Cambridge, MA: The MIT Press.

Lindsey, R. V., M. Khajah, and M. C. Mozer. 2014. “Automatic Discovery of Cognitive Skills to Improve the Prediction of Student Learning.” Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada, pp. 1386–1394.

Macdonald, C. 2016. “The End of Exams? Algorithm Can Predict How Students Will Answer Questions, and Even Explain Why They Would Get Questions Wrong.” *Daily Mail*. Accessed 19 February 2019. https://www.dailymail.co.uk/sciencetech/article-3380374/The-end-exams-Algorithm-predict-students-answer-questions-explain-questions-wrong.html.

MacKenzie, D. 1978. “Statistical Theory and Social Interests: A Case-Study.” *Social Studies of Science* 8 (1): 35–83.

Mackenzie, A. 2013. “Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity.” *Subjectivity* 6 (4): 391–405.
Mackenzie, A. 2017. *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press.
Mackenzie, A., C. Waterton, R. Ellis, E. K. Frow, R. McNally, L. Busch, and B. Wynne. 2013. “Classifying, Constructing, and Identifying Life: Standards as Transformations of “The Biological”.” *Science, Technology, & Human Values* 38 (5): 701–722.
Murphy Paul, A. 2012. “The Machines are Taking Over.” *The New York Times Magazine*. Accessed 9 October 2019. https://www.nytimes.com/2012/09/16/magazine/how-computerized-tutors-are-learning-to-teach-humans.html.
Naughton, J. 2018. “Magical Thinking About Machine Learning Won’t Bring the Reality of AI any Closer.” *The Guardian*. Accessed 9 October 2019. https://www.theguardian.com/commentisfree/2018/aug/05/magical-thinking-about-machine-learning-will-not-bring-artificial-intelligence-any-closer.
Pascanu, R., T. Mikolov, and Y. Bengio. 2013. On the Difficulty of Training Recurrent Neural Networks. In *Proceedings of the International Conference on Machine Learning*, 16–21 June 2013 Atlanta, pp. 1310–1318, February.
Perez, C. 2018. “The Emergence of Inside Out Architectures in Deep Learning.” *Medium*. Accessed 19 February 2019. https://medium.com/intuitionmachine/controlled-hallucinations-in-deep-learning-architecture-fd617150d677.
Piech, C., J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. 2015. “Deep Knowledge Tracing.” In *Proceedings of the 29th Conference on Advances in Neural Information Processing Systems*. 7th–12th December 2015. Montreal, Canada. pp. 505–513.
Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. “Learning Representations by Backpropagating Errors.” *Nature* 323 (6088): 533–536.
Rutkin, A. 2015. “RoboTutor is a Class Act.” *New Scientist*. Available online as “Hate Exams? Now a Computer Can Grade You by Watching You Learn.” Accessed 19 February 2019. https://www.newscientist.com/article/mg22930542-500-hate-exams-now-a-computercan-grade-you-by-watchingyoulearn/.
Selbst, A. D., S. Friedler, S. Venkatasubramanian, and J. Vertesi. 2018. “Fairness and Abstraction in Sociotechnical Systems.” *ACM Conference on Fairness, Accountability, and Transparency (FAT*)* 1 (1): 1–17.
Suchman, L. 1987. *Plans and Situated Actions*. Cambridge: Cambridge University Press.
Wang, L., A. Sy, L. Liu, and C. Piech. 2017. “Learning to Represent Student Knowledge on Programming Exercises Using Deep Learning.” In *Proceedings of the 10th International Conference on Educational Data Mining*, Wuhan, China, pp. 324–329.
Williamson, B. 2017. “Who Owns Educational Theory? Big Data, Algorithms and the Expert Power of Education Data Science.” *E-Learning and Digital Media* 14 (3): 105–122.
Wilson, K. H., Y. Karklin, B. Han, and C. Ekanadham. 2016. “Back to the Basics: Bayesian Extensions of IRT Outperform Neural Networks for Proficiency Estimation.” *Proceedings of the 9th International Conference on Educational Data Mining*, June 29–July 2, 2016, Raleigh, North Carolina, USA, pp. 539–544.
Winner, L. 1980. “Do Artifacts have Politics?” *Daedalus* 109 (1): 121–136.
Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.
Xiong, X., S. Zhao, E. Van Inwegen, and J. Beck. 2016. “Going Deeper with Deep Knowledge Tracing.” In *Proceedings of the 9th International Conference on Educational Data Mining*, June 29, 2016–July 2, 2016, Raleigh, NC, USA. pp. 545–550.
Yeung, C. K., and D. Y. Yeung. 2018. “Addressing Two Problems in Deep Knowledge Tracing via Prediction-Consistent Regularization.” *Online Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. London, United Kingdom — June 26–28, 2018, doi: 10.1145/3231644.3231647.
Zeiler, M. D., and R. Fergus. 2014. “Visualizing and Understanding Convolutional Networks.” In *Computer Vision, ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, Proceedings, Part I*, edited by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, 818–833. Cham, Switzerland: Springer.
Zhang, L., X. Xiong, S. Zhao, A. Botelho, and N. T. Heffernan. 2017. “Incorporating Rich Features into Deep Knowledge Tracing.” In *Proceedings of the Fourth ACM Conference on Learning@ Scale*. ACM. pp. 169–172, April.