Big Data in online education: Who produces value and who reaps the rewards?

Davor Petreski

University of Glasgow

Abstract

From classifying learners to predicting learner behaviour, the application of Big Data in online education has been vast. Besides the potential benefits of Big Data in education, it is necessary to critically engage with some ethical and social challenges that Big Data presents to the field of online learning. The increasing use of big data by large institutional actors and corporations raises questions not only about data privacy and ownership, but whether this data is used to genuinely improve learner and teacher online learning experiences, or primarily for commercial profits and institutional benefits. When addressing ethical concerns regarding the use of Big Data in education, critiques often follow a reasoning that is in line with corporate interests and neoliberal logic of marketization of education. Given the importance of the pursuit for democratic online education, the need for critical perspectives in the field is ever-more essential. This research tries to critically address the role and impact of Big Data on labour relations and economic fairness in online education by examining both corporate and institutional data practices in online learning. The study puts forward a provisional theory of the use of Big Data in two large online learning platforms (Coursera and Blackboard) using critical grounded theory. The core category of Exploitation of the learning community, the three constituent concepts; the Vendor-Institutional Complex, Use of learner generated value for profit, and the Behavioral monitoring and engineering; and the sustaining category, the Magic Trick, were the foundational blocks for developing an emancipatory theory that addressed ethical issues of economic fairness regarding the use of big data in online education.

Keywords

Big Data, online education, data ethics, Coursera, Blackboard

Introduction

The rapid technological advancement in computing in the past three decades has allowed humans to quickly and efficiently gather, access, and process large quantities of information. This revolution or breakthrough in information technology is often referred to as the Big Data Revolution (Kitchin, 2014).

Just like many other industries, sciences, and areas of social life, education too, is under a mass wave of digitization and datafication. Meaning, more and more learning and teaching is done online, using software programs that run on, collect, and process massive amounts of digital data. Thus, the practices and logic of the big data revolution have also penetrated education. The aim of this study is to bring about greater conceptual clarity regarding the ethics of big data practices in education, particularly to propose a conceptual framework regarding issues relating to the use of big data, and economic fairness in online learning. Furthermore, this study aims to critically address economic fairness and labour relations in online education, in light of the big data revolution.

To conduct the study, I carried out a qualitative critical grounded theory (CGT) case study of two of the most prominent online education providers, Coursera and Blackboard. The study resulted in a provisional conceptualization of the economic model of online education in the age of big data and the ethical concerns relating to it.

In line with CGT principles and in order to stay open to emergent questions and concepts from the data, I chose to only pose one preliminary broad question that allowed me to approach the data openly and inquisitively, yet with a clear topic in mind. In the words of Glaser, this preliminary research question let me engage the initial stages of the research with the “abstract wonderment of what is going on” (1992, p.22). The first research question is as follows:

> *How and for what purpose is Big Data used in online education?*

This question allowed me to engage with other emerging questions and problems that came to light throughout the data collection and analysis process. From the emergent questions and problems, one central research question was defined:

> *What is the role and impact of Big Data on labour relations and economic fairness in online education?*

This research question is central due to its synergistic relationship with the data and the study. On one hand, it is informed by the data and was arrived at by analysing and ‘following’ the patterns in the data, and on the other, it served as a guiding tool for further exploration and analysis.

Once, I reached a certain level of theoretical saturation regarding the second research question, I noticed that there were some definite conceptual and explanatory gaps in the emerging theory. More precisely, whereas a conceptualization of the economic model and logic of big data in online education was developed (or discovered), an explanation as to why and how is that model maintained, was missing. This led to the emergence of one last, new research question:

> *How is the economic model of big data in online education maintained?*

Methodology and Methods – Critical Grounded Theory

By aiming to bring about conceptual clarity, this research warrants a methodological approach that is suitable for theory building or development, rather than empirical theory testing. Thus, the grounded theory methodology (GTM) is considered most appropriate research methodology for this study due to three reasons.

1. Firstly, grounded theory seeks to generate new explanatory theories (Corbin & Strauss, 2008).
2. Secondly, the essence of grounded theory is that the theory building process is grounded in the data, and hypothesis testing is avoided (Suddaby, 2006).
3. Lastly, grounded theory is specifically appropriate for “discovery-oriented” research in areas of study that are under-theorized (Burck, 2005, p.244).

Since its initial introduction, even between the original authors, there have been multiple points of contention on how grounded theory should be done. Therefore, multiple branches of GTM have emerged and are often in dispute with one another. Namely, first is the Classical Grounded Theory, which is closest to the original methodology presented in 1967 (Glaser & Strauss, 1967); second, the Straussian model, first introduced by Corbin and Strauss (2008); third, Constructivist grounded theory, developed by Charmaz (2000); and one of the latest variants of grounded theory, Critical Grounded Theory (CGT). For this study, CGT was chosen as the most appropriate approach. Critical grounded theory is divergent from all the other variants of GTM in three aspects. Firstly, ontologically it is neither based on the post-positivist nor the constructivist paradigm, rather it aligns with the critical realist ontology. Secondly, it is concerned with creating critical, emancipatory knowledge

regarding issues such as power, justice, and equality. Thirdly, it introduces retroduction as a mode of critical inquiry. The methodological principles, coding, and data analysis processes were mainly informed by Hadley (2017; 2019).
The methodological principles employed in this study are: Openness, Iteration and the Constant Comparative Method, Theoretical Sampling, Memoing, Theoretical Sampling, and Production of a Substantive Theory. The data analysis process was conducted in four steps: Open Exploration, Focused Investigation, Theoretical Construction, and Transformative Dissemination.

Research Findings and Theory Building

The research findings bring together results from the data analysis, relevant data codes, author memos, and emerging categories in order to construct a conceptual framework of 'what is going on' in the field of big data in online education, specifically relating to the critical issues regarding economic fairness and digital labour. The core category of *Exploitation of the learning community*, the constituent concepts such as; *the Vendor-Institutional Complex*, *Use of learner generated value for profit,* and *the Behavioral monitoring and engineering;* and the sustaining category, *the Magic Trick,* were the foundational findings that will serve as the base for the construction and presentation of the substantive theory.

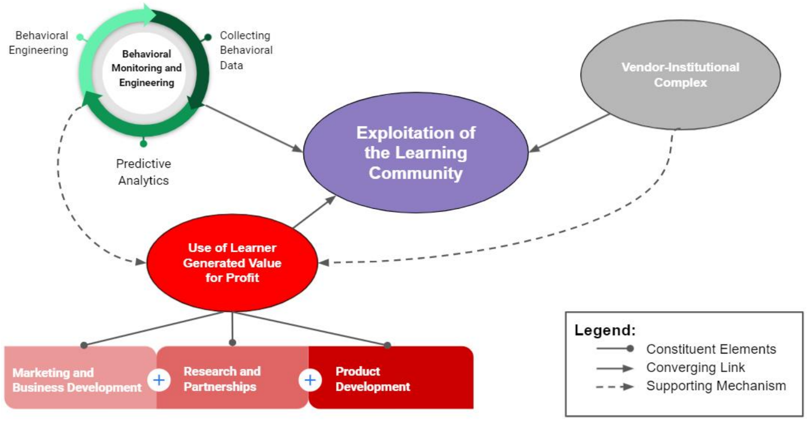

Figure 1: Theory of Exploitation of the Online Learning Community in the era of Big Data

Core Category: Exploitation of the learning community

The core category of *Exploitation of the learning community* materialised as a product of the conceptual relationships drawn between the other three surrounding concepts that emerged from the data. In other words, the Core Category is the aggregation of the three main categories or concepts. Furthermore, it is the central thesis of this research and the basis for the emergent theory.
To more clearly understand the Core Category, it is divided into two main constituent sections: Exploitation and Learning Community.

Exploitation

This section of the category addresses the question 'what is being done?', it focuses on the action or the practice of exploitation in online education. Exploitation can be defined as the action of taking an unfair advantage over someone, for one's own benefit (Stanford Encyclopaedia of Philosophy, 2016) . Since in the field of big data in online education, data is primarily used for financial gain, we can classify the exploitative data practices in online education as primarily of a commercial character. Therefore, this form of exploitation entails extracting value from vulnerable or unaware individuals and groups in an unfair way, and using this value to generate profit.
Through their data practices, policies, and actions, both Coursera and Blackboard engage in such extraction of value to secure financial gains. The extraction of the value is mainly done through the use of *learner generated data for profit.* Furthermore, what makes this extraction unfair is the *behavioural monitoring and engineering* that supports this extraction, and the *magic trick* that maintains the exploited in a state of unawareness and confusion. An example of the blend between using learner generated data for profit, and using behavioural engineering to support extraction of value by Coursera is presented in Code #78.

Table 1: Coursera Code #78

Data Type	Quote	Source
Key Actor: Emily	"For example, our learner-product interest models determine what degrees each user sees in their browser, how degrees are ranked in her	Sands, E. (2020, April). Coursera's

Proceedings for the Thirteenth International Conference on Networked Learning 2022, Edited by: Jaldemark, J., Håkansson Lindqvist, M., Mozelius, P., Öberg, L.M., De Laat, M., Dohn, N.B., Ryberg, T.

Sands, VP of Data Science, Coursera 2020 Virtual Conference	megamenu and more. These algorithms are built on deep understanding of learners from self reported features like work and education history, to behavioral features like how the learner found Coursera, what she searched for and enrolled in, and how she progressed through her learning experiences. Combined with meta-data on each degree and using as training data the conversion behaviour of the millions of other learners who have been exposed to degrees on Coursera in the past we estimate each learners interest in each program. ...This is leading to a 40% increase in degree applications through browse.”	Product Leadership Presents: Product Innovations [Product Innovation Presentation]. 2020 Coursera Virtual Conference, Mountain View, CA, United States.

Learning Community

Provided that Exploitation is an existing reality in the field of online education, and big data is the enabler, it is crucial to understand who are the exploited, and why. This section particularly addresses these questions. There are different actors in the big data economy of digital learning. Namely, there are the companies such as Coursera and Blackboard, academic institutions such as universities and schools, teachers and content providers, and lastly, the learners. In the cases of Blackboard and Coursera, the companies and the academic institutions are the owners and controllers of the data, and they decide how and why the data is used and collected.

The data that the companies and institutions control is mined from the learners’ activities, content and experiences. Therefore, we arrive at having two groups with a clear and distinctive difference in power and economic benefit. On one hand, we have the companies and institutions as data controllers who extract value and use it for their own benefit, and on the other, we have the learning community which is comprised of learners and instructors, whose data is being collected.

The learning community, just by existing and functioning on learning platforms such as Coursera and Blackboard, is the producer of big amounts of behavioural and learning data. As producers of such data, the learners and instructors are not compensated for the economic value they are producing, and therefore, are engaging in invisible unpaid labour. Moreover, a large learning community is both a key selling point for business partnerships and an essential competitive advantage. Thus, the learning communities are not only the uncompensated producers of data but also the products and commodities of online education platforms. Lastly, due to *behavioural monitoring and engineering*, the learning community are also the subjects in light of big data usage by educational platforms. As such, they are being manipulated, researched about, and experimented on, in order to gain business or product insights, or compel them into paying and producing more data on these platforms.

Concept 1: Use of Learner Generated Value for Profit

The *Use of Learner Generated Value for Profit* is one of the central, and first categories that emerged from the data. In its essence, it is the idea that online education providers such as Coursera and Blackboard use the data produced by the learning community for their own commercial benefit. This benefit can be segregated into three goals: *Marketing and Business Development, Research and Partnerships,* and *Product Development*.

Marketing and Business Development

Blackboard and Coursera, as the controllers of data and online education providers, are able to translate the learner-generated data into profit by extracting valuable insights that fuel their business development and marketing strategies, or in the case of Blackboard, the marketing strategies of their partner institutions. For Coursera, this can range from internal marketing efforts, such as converting non-paying learners on their platform into paying customers for a low cost of acquisition, to external behavioural advertising methods in order to attract more learners to their platform. Code #78 (Table 1), presented in the previous section is a clear example of the former.

Additionally to using user-generated value for marketing purposes, Coursera also uses learner data for the development of its business and exploring new profit-making avenues. For instance, as stated in Coursera Code #48, through learner data powered decision making Coursera informs its business development roadmap.

Table 2: Coursera Code #48

Data Type	Quote	Source
Key Actor: Vinod Bakthavachalam, Data Scientist at Coursera; Website Content; Blog	“At Coursera we use data to power strategic decision making, leveraging a variety of causal inference techniques to inform our product and business roadmaps”	Bakthavachalam, V. (2018, November). Controlled Regression: Quantifying the Impact of Course Quality on Learner Retention. Medium.

Similarly to Coursera, Blackboard also uses learner-generated data for marketing purposes and behavioral targeting and advertising. Furthermore, Blackboard also provides digital marketing data-powered services to academic institutions. The key selling point for this service is the ability to closely track learner behaviour through the enrolment marketing funnel.

Research and Partnerships

Besides marketing and business development, the learner-generated value in the form of data is also being used to conduct experiments and research in the newly established field of online education. For this purpose, learners may be shown different variations of content offerings in their courses. For Coursera, this research is often coupled with building profitable relationships with academic institutions and other business partners. Blackboard shares research data with partner institutions similarly to Coursera. However, Blackboard also uses this learning analytics research for product innovation and development. Thus, synthesizing the general research in online education, and particular research that mostly benefits Blackboard for their product promotion and development.

Product Development

There are multiple examples of Coursera and Blackboard using user-generated data to fuel their product development and improve their products. For instance, Coursera has developed a relevancy-based algorithm for their search engine using data of over 10 million learners. This algorithm allows Coursera to show the courses and degrees that learners are most likely to enrol in and pay for. Emily Sands, VP of Data Science at Coursera explains this in Code #80.

Table 3: Coursera Code #80

Data Type	Quote	Source
Key Actor: Emily Sands, VP of Data Science, Coursera 2020 Virtual Conference	"We also evolved our search engine… to a relevance-based algorithm. Ranking according to what learners searching for that term ultimately went on to enrol in, pay for, and apply to. This enables learners to find the right content from among the base selection on Coursera faster powered by the search and downstream behaviour of the 10s of millions who came before."	Sands, E. (2020, April). Coursera's Product Leadership Presents: Product Innovations [Product Innovation Presentation]. 2020 Coursera Virtual Conference, Mountain View, CA, United States.

With the three constituent parts explained, we can move onto shortly summarising this Concept and integrating it with the existing literature. The main notion of this Concept is that value is being generated by learners in the form of data, which is then used by Coursera and Blackboard for their own profit and benefit. Furthermore, even though these companies continue to reap the financial rewards of the value generated by learners, the learners are not compensated.

This issue is largely overlooked by the scholarly work on big data in online education, nevertheless, several authors in the relevant academic literature raise similar concerns. For instance, Shum and Luckin (2019) argue that tracking and quantifying human behavioural data is a gold mine for marketers and researchers, but little is being done to improve teaching and learning. Furthermore, Williamson (2019) conceptualises the marketisation of Higher Education and the data infrastructure that surrounds it. Drawing on Srnicek (2017), Williamson brings to light the generation of value and profit from learner produced data (2019). Williamson expands on this by examining the market-making practices in digital platforms in Higher Education, particularly the case of Pearson (2021). Lastly, relating to the Research and Partnerships segment of this concept, Marshall (2014), brings up concerns regarding the experimentation on learners using untested pedagogical practices on the EdX online education platform.

Concept 2: Behavioral monitoring and Engineering

If the first concept discussed in this paper addressed the unfair use of data in online education, the concept of *Behavioral Monitoring and Engineering* pertains to the unfair extraction of data. Behavioural data is the data gathered by tracking and monitoring the actions and experiences of learners, such as how long do learners spend on certain pages, where do they click, what actions do they perform before paying for a course, once enrolled in the course, what steps do they take before dropping out or successfully finishing etc. Therefore, behavioural data is of central value for online education providers. Blackboard and Coursera use behavioural data for two distinct purposes. Firstly, behavioural data is used to predict and improve what is deemed to be learner success, and secondly for commercial purposes, such as influencing a learner to pay for a certificate or enrol in a degree.

The Behavioral Monitoring and Engineering process is split into three steps. First, it starts by collecting the behavioural data. Secondly, predictive analytics are used to predict future behaviour, such as the likelihood of dropping out, or not finishing a course. Lastly, it ends by intervening in order to alter unwanted future behaviour for the benefit of the company, institution, or the learner.

Predictive analytics is the practice of using large historic behavioural data sets to train algorithmic models which then predict the behaviour of learners. Put simply, in order to make predictions about a current, individual learner, these models reflect on how similar learners with similar past experiences behaved. For instance, Coursera may use an algorithmic model to predict whether a learner is likely to pay for a certificate at the end of the course based on their performance and behavioural data and the performance and behavioural data of millions of other past learners in that course.

Besides predicting human behaviour, Coursera and Blackboard use big data to alter it by the use of targeted communication and nudges, visual modification and recommendation models, and advertisements. Targeted communications and nudges are automated messages and notifications that aim at intervening in and altering human behaviour. This method has the simplest underlying model of behavioural control, since it largely relies on verbal or textual communication. However, the structure behind when, where and how are these messages and notifications sent, is incredibly complex and based on large amounts of data and computational analytics. Some messages aim at altering behaviour in order to improve learner success and the learning experience. However, others, aim at compelling students towards enrolling and paying for online degrees, certified programs or similar paid content.

Furthermore, another mode of behavioural engineering is the visual modification and recommendation models. Often, this mode is also named 'Personalisation of Content'. I will not use this terminology, since I believe it falsely represents the practice of modifying content for commercial benefit as an attempt for personalisation and improvement of the learner's personal learning journey. Through content modifications and recommendations informed by big data, companies such as Coursera can control what the learners see, and do not see. Consequently, learners might enrol in a degree that is just simply made more visible to them, rather than taking their own, personal learning path.

Many works in the contemporary scholarly literature deal with behavioural data in online education (Kizilcec et al., 2020; Qiu et al., 2016; Tseng et al., 2016; Wassan, 2015). However, critical perspectives on the use of behavioural data in the field are rare (Regan & Jesse, 2018; Reidenberg & Schaub 2018). Firstly, Reidenberg & Schaub (2018) raise concerns over the increase in learner stress, knowing that their steps are being watched and surveilled. Moreover, they further note the danger of the use of learner behavioural data for manipulation outside of the learning context, for commercial purposes. Similar to the findings in this study, Regan and Jesse (2018) find the ethical issues of nudging problematic in certain circumstances, especially in the field of education. They argue that these nudges must be transparent and promote social welfare, rather than become tools of manipulation for commercial benefit. The findings of this study are further supported by Yeung (2017), arguing that by using nudges companies become "choice architects" and have the power to alter human behaviour in a predictable way.

Moreover, ethical concerns regarding the use of behavioural data are being raised in the broader literature on big data, as well, especially in the fields of information systems and economics (Herschel & Miori, 2017; Newell & Marabelli, 2015; Zuboff, 2015; Zuboff, 2019). For instance, Newell & Marabelli (2015), uncover the falsely portrayed 'free access' to information on the internet, arguing that in fact, large tech companies have control over what we see and access. They further argue that this control over what the user sees leads to a slow and subtle manipulation of the user's worldview.

Concept 3: The Vendor-Industrial Complex

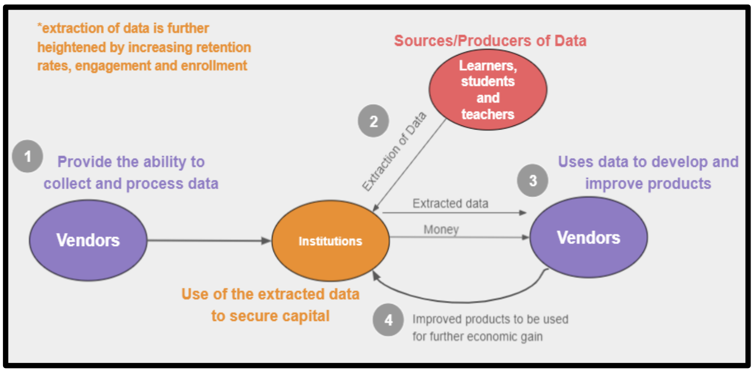

Figure 2: The Vendor Industrial Complex

The title of the Vendor-Institutional Complex concept is partially inspired by the existence of other industrial complexes such as the Military-Industrial complex, or the Prison-Industrial Complex. It captures how institutions, in this case, academic ones, reconstruct their relationship with industrial enterprises in accordance with capitalist and neoliberal models with the aim of financial growth.

In online education, institutions and vendors (such as Blackboard and Coursera), as owners and controllers of the data, have a shared, vested economic interest in extracting data from learners and benefiting from the free labour that the producers of data provide. Therefore, their relationship forms an economic model that is based on and aimed towards the extraction of value from the data students and teachers produce. Even though both Blackboard and Coursera are engaged in the Vendor-Institutional Complex by partnering with universities and other academic institutions, there is much richer data explaining Blackboard's involvement in such relationships with their institutional partners.

In order to more clearly understand the Vendor-Industrial Complex, Figure 2 presents a diagram that I created during the Theoretical Construction stage. I will further explain this diagram by listing the four main steps in the cyclical process of the Vendor-Institutional Complex.

As seen in the diagram above, the Vendor-Institutional Complex has four main stages or steps.

1. Firstly, the vendors, in this case, Blackboard, provide the ability for institutions to collect and process data *en masse*. This practice is called the productization of data collection and processing.
2. Academic Institutions extract data from students and teachers, who are seen as the mere producers or sources of data. Once the data is extracted, institutions use this data to gain value and secure economic gains. The data can be used for commercial purposes such as cutting costs, retaining students, or improving administrative efficiency, or informing digital marketing strategies.
3. Following the extraction of value, institutions share the extracted data with the vendors and provide them with payment for their services. In turn, Vendors use this data and resources to further develop and improve their products.
4. Lastly, these improved or newly developed products and services are sold to academic institutions, which are then used for further data extraction, economic gain, and cost-cutting.

Finally, it is important to note the cyclical and reproductive nature of the Vendor-Industrial Complex. The increased efficiency of data practices and improved retention and enrolment rates lead to further data extraction from a larger pool of learners and teachers, or in other words producers of data.

The Vendor-Institutional Complex is a novel conceptualisation and to my best knowledge, does not relate to any of the previous literature. For instance, Reyes (2015), completely excludes vendors and online education platforms as stakeholders that benefit from big data in online education. Furthermore, Selwyn (2014) provides a critical perspective of the 'digital university', arguing the emphasis on neoliberal logic by educational key actors such as policymakers and influencers. However, Selwyn does not explore the role of vendors and private companies in the process of building the digital university (2014). Therefore, critical scholarly work focusing on the relationship between academic institutions and commercial vendors is quite limited, and further work exploring the vendor-institutional complex is needed.

Sustaining Category: The Magic Trick

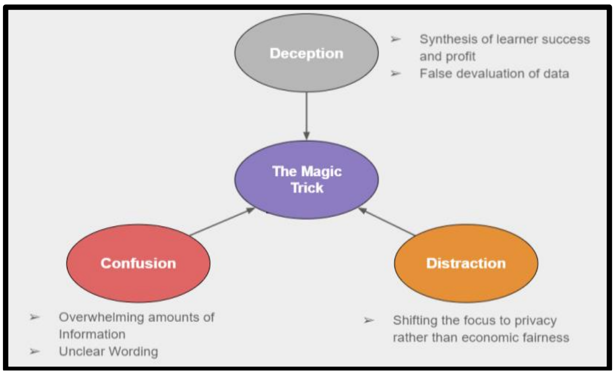

Figure 3: Sustaining Category – the Magic Trick

The Magic Trick category emerged by asking "what must be true for the exploitation of the learning community to be taking place, and how is this model maintained?". One of the reasons for this inquiry was because I was confused by the fact that individuals and groups within the learning community are not massively protesting this exploitation.

Upon further data collection and analysis, I arrived at two emerging possibilities. One, the learning community is comfortable with and consensual to the big data practices and the logic underlying them. Two, there is a lack of informed knowledge about the exploitative practices, and these are being hidden from their awareness. The first possibility has some minimal supporting evidence, that suggests that students were comfortable with being contacted based on the use of learning analytics. However, the students were not informed about what information was collected and how they were tracked, and the research was conducted by Blackboard. Furthermore, supporting evidence for the second possibility is overwhelmingly more voluminous.

The name of the concept, Magic Trick, comes from the three different methods used to conceal the exploitative practices of Blackboard and Coursera. In other words, the learning community is tricked into unawareness. Any good magician uses three basic methods to pull off a magic trick; confusion, distraction, and deception. Similarly, these practices are also present in the Magic Trick that big data based online education vendors are playing on the learning community.

Confusion

When magicians perform a trick, they might employ a tactic of overwhelming the subjects with too much information or simply performing a plethora of movements and actions so that the subject is left confused. Confusing and overwhelming the audience is one way of covering what the magician is really doing. Coursera and Blackboard, virtually employ that same tactic of confusion, by presenting the audience with overwhelming amounts of information that is often unclear, and that is incredibly difficult to navigate. For instance, Memo #29, presents an observation made about the time and effort it takes to go through all the information needed for one to understand how their data is used. Namely, one user needs to go over approximately 100 pages of highly technical text.

Table 4: Memo #29

Memo #29
Through my data collection and analysis work, I have come to realize how much time and effort is actually needed to clearly understand how Blackboard and Coursera are using learner-generated data. For instance, in the case of Blackboard, one must go through over 50,000 words of text (privacy statements, terms and conditions of use, third party statements etc.), and that is not including the privacy policies and statements of the academic institutions and some smaller third-party partners, who also use user-generated data on Blackboard.

Moreover, it’s not only that the amount of text is overwhelming, but the wording in the privacy statements and documents is often incomplete and unclear, leaving open possibilities for further exploitation. This unclear wording, when communicating the collection of data from learners often includes phrases such as “among other things” or, “any other data that is generated by you”, setting no boundaries to what data can be collected and for what purposes.

Additionally to the overwhelming and unclear information provided to learners by Coursera and Blackboard, learners must also go through the data policies of third-party partners and policies based on local laws and regulations. For instance, one of Blackboard and Coursera’s largest partners is Amazon Web Services (AWS), they use learner-generated data on Blackboard and Coursera to train their machine learning algorithms (e.g. algorithms for natural language processing, facial recognition etc.).

Distraction

Often, a magician will want to shift their subject’s focus away from what is really important, the trick. They do this by distracting the audience by presenting a dummy point of attention, or a decoy. Unlike economic fairness, safeguarding privacy does not challenge the logic behind the commercial value generation from big data in online education. After all, the financial gains in online education are not made by monetising personally identifiable information, but by productizing and marketizing big data sets and data analysis tools. Therefore, shifting the focus of ethical concern away from the economic exploitation in the field is achieved by paying and driving special attention to privacy. In this study, I have recorded twelve codes where privacy concerns have been addressed by Coursera and Blackboard, however, none addressing concerns over economic fairness and data exploitation. This overwhelming focus on privacy is also translated in the academic literature, where most of the work on big data ethics in online education is focused on privacy issues (Chen & Liu, 2015; Fischer et al., 2020; Johnson, 2014; Prinsloo & Slade, 2017; Reidenberg & Schaub, 2018; Wang, 2016; Williamson, 2017).

Deception

The last and the most central step of any magic trick is *Deception.* It is the act of leading someone to accept a false truth, or in other words, the act of hiding the truth under a veil of falsehood. Relating to this phenomenon, a peculiar category emerged from the data I collected on Blackboard and Coursera; the synthesis of learner success and commercial gain. Namely, both companies marry learner success with their financial success and the financial success of their partners. This way, Blackboard and Coursera can exploit learner’s data by falsely claiming that it is the learner’s success that they have in mind, not profit.

An additional category that might be relevant to the phenomenon of *Deception* is the devaluation of data. By arguing that data must first be analysed, refined and cleaned before it is valuable, key actors at Blackboard and Coursera are assigning no value to the raw data that is generated by the learning community. This way, the extraction of learner data will not be perceived as economically unfair, or as exploitation, since these companies are not extracting anything of direct commercial value.

Summary, limitations and final remarks

The findings in this study presented the emergence of a core category, three main concepts and one sustaining category. These main elements and the relationships between them compose the Theory of Exploitation of the Online Learning Community in the era of Big Data. The theory explains the purpose and role of Big Data in creating and maintaining the economic model and labour relationships in the field of online education. Being critical in nature, the theory particularly raises concerns regarding economic fairness and labour exploitation. Furthermore, by incorporating the sustaining category of the Magic Trick, the theory further explains how the economic model and exploitation are maintained. Figure 1 illustrates an overview of the theory. The sustaining category of the Magic Trick is not included in the illustration since the Magic Trick is the invisible background on which the relationships between the concepts and the core category play out.

By looking at the legend in Figure 1, one can notice that there are three types of relationships; constituent elements, converging links, and supporting mechanisms. The first one relates to the elements that constitute a certain concept. For instance, predictive analytics is a constituent element of the concept Behavioral Monitoring and Engineering. The converging links represent the convergence of the concepts into the Core Category. In other words, when the main three concepts are united, the core category of Exploitation of the Learning Community emerges. Each of the three concepts play a part in explaining how and why the learning community is unfairly exploited. Lastly, the supportive mechanism links represent relationships where one concept supports the existence of the processes in another. For example, the Collection and Engineering of Behavioural Data supports the Use of Learner Generated Value for marketing and business development purposes. These constituent elements of Concept 2 provide the necessary mechanisms for the materialisation of the processes in Concept 1.

Implications for Networked Learning (NL)

The findings in this paper also have some important implications and insights for the field of NL. Namely, the research focused on platforms that enable “connections between individuals, learning materials, and learning community” (Rodríguez-Illera et al., 2021), and digital spaces in which networked learning happens. Therefore, this paper provides critical insights into the platform commercialisation and commodification of these connections by extracting, storing and analysing valuable data from learning networks in these spaces. In the field of NL, more recently, there have been growing number of concerns expressed regarding the ethical implications of using commercial platforms as spaces for learning (Rodríguez-Illera et al., 2021). Expanding on these, this paper presents a critical investigation into the economical fairness of the use of big data practices in some of the most current and popular online spaces and platforms for networked learning.

Limitations

Besides some methodological limitations of CGT, there are other, theoretical limitations that demand consideration. Firstly, including only two case studies as the focus for the study, the knowledge and the theory that emerged is local and narrow in context. Therefore, the emergent theory is not, and it does not aim or claim to be generalizable, limiting the applicability of the theory to different contexts. However, as previously mentioned, the theory is modifiable and open for adaptations and comparisons with contexts, different from the one studied.

Secondly, other than my personal experiences and observations, the theory does not include the experiences and knowledge from main actors in the field such as learners, teachers, and employees in online education companies. Dealing with particular themes such as exploitation and deception, these perspectives are crucial for the development of a holistic theory.

Lastly, besides offering a conceptual map that provides opportunities for social change and anti-hegemonic action, the theory and the study itself do not present viable solutions and potential avenues for action. Therefore, this is only a preliminary study and further work is needed.

Concluding Remarks

Before finishing this work, I would like to present one last remark regarding the tone and intent of the study. When reading this work, and interpreting the emergent theory, one might falsely assume that I am criticizing the use of big data in online education as a whole, or that I am advocating against the use of these technologies. However, what I aim at critiquing and advocating against is the underlying, exploitative logic behind the use of big data in online education. Big data, learning analytics and artificial intelligence as technologies have huge potential to be beneficial for both learners, teachers, and educational institutions. However, for these benefits to materialize, the priority when using them should be the wellbeing and flourishment of learners and improving the learning experience, not commercial goals and financial gains.

References

Burck, C. (2005). Comparing qualitative research methodologies for systemic research: The use of grounded theory, discourse analysis and narrative analysis. Journal of family therapy, 27(3), 237-262.

Charmaz, K. (2000). Constructivist and objectivist grounded theory. Handbook of qualitative research, 2, 509-535.

Chen, X., & Liu, C. Y. (2015). Big data ethics in education: Connecting practices and ethical awareness. *Journal of Educational Technology Development and Exchange (JETDE)*, *8*(2), 5.

Corbin, J. & Strauss, A. (2008). Basics of qualitative research (3rd ed.): Techniques and procedures for developing grounded theory. SAGE Publications, Inc., https://www.doi.org/10.4135/9781452230153

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., ... & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130-160.

Glaser, B. G. (1992). Basics of Grounded Theory Analysis: Emergence Vs. Forcing. Sociology Press.

Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Adline De Gruyter.

Hadley, G. (2017). Grounded theory in applied linguistics research: A practical guide. Routledge.

Hadley, G. (2019). Critical grounded theory. The Sage handbook of current developments in grounded theory, 564-592.

Herschel, R., & Miori, V. M. (2017). Ethics & big data. Technology in Society, 49, 31-36.

Johnson, J. A. (2014). The ethics of big data in higher education. *The International Review of Information Ethics*, *21*, 3-10.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., ... & Tingley, D. (2020). Scaling up behavioral science interventions in online education. Proceedings of the National Academy of Sciences, 117(26), 14900-14905.

Marshall, S. (2014). Exploring the ethical implications of MOOCs. Distance Education, 35(2), 250-262.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datafication'. The Journal of Strategic Information Systems, 24(1), 3-14.

Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In *Big data and learning analytics in higher education* (pp. 109-124). Springer, Cham.

Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016). Modeling and predicting learning behavior in MOOCs. In Proceedings of the ninth ACM international conference on web search and data mining (pp. 93-102).

Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. Ethics and Information Technology, 21(3), 167-179.

Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. Theory and Research in Education, 16(3), 263-279.

Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. TechTrends, 59(2), 75-80.

Rodríguez-Illera, J. L., Gachago, D., Hansen10, S. B., Themelis, C., Matthews16, A., Thibaut18, P., ... & Czerniewicz32, L. (2021). Networked Learning in 2021: A Community Definition.

Selwyn, N. (2014). Digital technology and the contemporary university: Degrees of digitization. Routledge.

Shum, S. J. B., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. British journal of educational technology, 50(6), 2785-2793.

Srnicek, N. (2017). Platform capitalism. John Wiley & Sons.

Stanford Encyclopedia of Philosophy. (2016). Exploitation (Stanford Encyclopedia of Philosophy). Https://Plato.Stanford.Edu. https://plato.stanford.edu/entries/exploitation/

Suddaby, R. (2006). From the editors: What grounded theory is not.

Tseng, S. F., Tsao, Y. W., Yu, L. C., Chan, C. L., & Lai, K. R. (2016). Who will pass? Analyzing learner behaviors in MOOCs. Research and Practice in Technology Enhanced Learning, 11(1), 1-11.

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends*, *60*(4), 381-384.

Wassan, J. T. (2015). Discovering big data modelling for educational world. Procedia-Social and Behavioral Sciences, 176, 642-649.

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.

Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. British Journal of Educational Technology, 50(6), 2794-2809.

Williamson, B. (2021). Making markets through digital platforms: Pearson, edu-business, and the (e) valuation of higher education. Critical Studies in Education, 62(1), 50-66.

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. Information, Communication & Society, 20(1), 118-136.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. Journal of information technology, 30(1), 75-89.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (1st ed.) [E-book]. PublicAffairs.