Chapter 2

Big Data and the Transformation of Decision Making in Higher Education

Braden J. Hosch

Gosh, you've really got some nice toys here.

– Roy Batty *Blade Runner* (1982) set in 2019

Introduction

In the summer of 1956, luminaries of mathematics and information sciences gathered at Dartmouth College for two months to hold in-depth discussions about what the group organizer John McCarthy termed "artificial intelligence." McCarthy's proposal to the Rockefeller Foundation to support the summer meeting provocatively and presciently asserted: "If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have" (McCarthy, Minsky, Rochester, and Shannon 1955).

Participants in this summer meeting included, among others, Arthur Samuel, who would later in the decade coin the term "machine learning" as he developed a computer program that could win a checkers game against a human being, Ray Solomonoff, who developed algorithmic information theory that machine learning is probabilistic and can be trained on existing data to solve new problems, Marvin Minsky, who would go on to co-found the AI Lab at MIT, and Claude Shannon at Bell Labs, who developed information theory and would later in life quip that

in computer-humans chess games, he was "rooting for the machines" (Shannon 1987). The meeting was significant not just because of the brilliance of its attendees but because of the problems addressed, which included discussion on automatic computers, how computers can be programmed to use a language, neuron nets, theory of the size of a calculation, self-improvement of data programs, abstractions, and randomness and creativity. These problems represented the central challenges of achieving artificial intelligence envisioned by Alan Turing (1950) several years earlier that a machine could mimic human behavior.

Solutions, however, remained elusive until improvements were made to computing power, data storage and management systems, and networking. The resulting developments are changing and will continue to change how organizations, including universities, operate and deliberate. This chapter provides historical context of how the transformation of computing from record keeping and administrative processing into what Agrawal, Gans, and Goldfarb (2018) call "prediction machines" affect how decisions are made and how big data represents a transformational means for colleges and universities to improve if not re-imagine their operations. "Big data," in this respect, is a broad term that includes huge amounts of structured data (such as all the clicks of all students in all online learning materials at a university), but also unstructured data like social media feeds with text, images and video files, as well as a set of non-hypothesis driven analytical techniques applied to existing (smaller) data sets. This chapter asserts that resulting developments in machine learning, artificial intelligence, and the Internet of Things provocatively point toward a future for the higher education sector in which decisions made by students, faculty, and administrators are approached much differently from earlier periods.

The Evolution of Computing Power and University Decision-Making

In the 1950s, colleges and universities were organizations of students, faculty and staff, concentrated on a geographically-defined campus, and who labored to produce voluminous textual material – articles, books, term papers, tests, and memos. World-class science occurred in laboratories, but results were recorded in quadrille notebooks and written up as lab reports, later typed up, submitted to and published in journals that would later be housed on the bookshelves of libraries. The walls of Registrar's offices were obscured by beautiful wooden filing cabinets with reams of student files printed on paper (often handwritten) and stored in actual folders. Decisions about whom to admit or not, whom to hire or let go, what programs to start or retire were made largely on the basis of professional expertise and judgement of experts who had spent entire careers at an institution. As Gladwell (2005) demonstrates, there is real value in the judgement of experts in their fields of expertise, but these judgements are also necessarily bounded by the knowledge of those making them.

Administrative computing became a reality at research universities following its deployment in academic computing in the late 1950s and 1960s, featuring large mainframe computers built by IBM and later Burroughs and Cray running with vacuum tubes adjacent to large cooling facilities. Initial processing power in the mid-1950s was measured in hundreds of instructions per second, increasing to millions of operations per second by the late 1960s (IBM 2003), several orders of magnitude slower than the personal mobile devices of the late 2010s, which record speeds of billions of operations per second (Simonite 2018). Data analysis became easier with the release of statistical software applications still in use today, such as the Statistical Package for the Social Sciences (SPSS) released in 1968 and the Statistical Analysis System (SAS) released in 1971. As faculty circulated through administrative roles, including the

relatively new function of institutional research, these applications became widespread tools of choice among institutional researchers to prepare descriptive statistics informing institutional leadership about the past. This knowledge was invaluable to institutional decision making, and campus planning estimates made use of cohort attrition models for enrollment planning and segmented yield rates for admissions, but forecasting still relied heavily on professional expertise informed by population-level statistics.

Even though computing power was still relatively limited, the promise of artificial intelligence to transform education was under active exploration, as evidenced by Ellis Page's initiative (Page, Fisher, and Fisher 1968) to grade composition papers using computing power of the day. Project Essay Grade, funded by the U.S. Department of Education (Page and Paulus 1968), investigated the feasibility of automatically analyzing and evaluating student writing using a FORTRAN program for natural language processing after student papers were keyed into mainframes. Project Essay Grade demonstrated that computer programs were about as good as human raters at evaluating student writing, although the methods remained too costly for widespread adoption (Page et al. 1968). Page later revived the project in the 1990s and with the exponential increase in computing power, the widespread use of computer terminals in testing, and the motivation of testing companies to cut costs, the basic infrastructure of Page's project became widespread in the 2000s.

Data storage and processing also evolved markedly during the 1950s and 1960s and the increased capacity to store data had implications for decision-making processes. Data and programs were created and stored on punch cards – technology from the 19th century to automate textile production. Use of magnetic tape to store data was introduced by IBM in 1951 and offered great advantages for increasing speed and volume but still carried limitations of

sequential storage. In the mid-1950s and with marked advancements in the 1960s, hard disks and

allowed for random access to the blocks in which data were stored, providing additional

advances in storage capacity and retrieval speed. Importantly, the technology allowed

development and commercialization of the floppy disk in the late 1960s and early 1970s that

allowed for the transport of data between microcomputers and mainframes. Direct access storage

of data versus sequential storage on tape or a box of punch cards also allowed for development

of data management systems, with the introduction of navigational databases in the 1960s and

the relational database in the next decade by Edgar Codd (1970). These events were followed by

the development of structured query language (SQL) later in the decade (Chamberlin and Boyce

1974) and later to be commercialized by Oracle for release in 1979 and is in widespread use

forty years later.

These intensive mainframe computing resources and data tools, however, were generally

reserved for large research universities, not smaller colleges; it was not until the 1980s, with

proliferation of the microcomputer or personal computer (PC) into faculty and administrative

offices, that computing power became inexpensive enough to become widespread for

management of colleges and universities. Gilbert and Green (1986) describe this era as the

computing revolution, noting that almost half a million microcomputers were operating on

campuses by the middle of the 1980s and over half of entering freshmen reported having

occasionally or frequently written a computer program. Importantly, personal computers

effectively pushed the ability both to generate and access data to every member of the university

community, although the potential of this breakthrough was realized only over the succeeding

decades. Gilbert and Green (1986) offered to campus leaders an overview of the challenges and

opportunities of technology adoption as well as a taxonomy for making decisions about

technology. However, their focus, and indeed the focus of administrative IT of the period, rested on how college and universities could and should manage information technology while remaining silent about how the computer revolution had potential to improve management of colleges and universities.

The Advent of Enterprise Systems

Potential for more widespread application of computing power to manage the higher education enterprise advanced significantly in the 1980s and 1990s with migration from locally developed administrative computing systems to broader adoption commercial Enterprise Resource Planning (ERP) systems like Banner and PeopleSoft. University ERP systems brought together many of the basic business operations of universities into an integrated platform, so that registration and student records, billing, budgeting, and human resources management became entirely digitized processes with data stored in common locations. These systems still notably omitted many mission-level functions of colleges and universities such as management of learning outcomes, teaching effectiveness, use of student services, and research activity and outcomes. In fact, the absence of these features within major higher education ERP systems has been the hobgoblin of efforts to measure and improve institutional effectiveness over the past two decades. Where ERP systems fell short for specific higher education functions, other vendors stepped into the breach with customer relations management (CRM) systems for admissions, learning management systems (LMS) for teaching and learning, assessment management systems for educational outcomes, donor management systems for alumni affairs and advancement functions. (For more information on CRM systems, see chapter 9; for more information on LMS systems as part of learning analytics, see chapter 10.)

Nevertheless, the ERP systems were transformative for decision-making processes within the institution. Significant and insignificant transactional details about students and employees migrated from paper records or siloed spreadsheets to centralized repositories of digital records, every added and dropped course, every salary increase or extra service payment, every purchase and payment was assigned an effective date and stored as a row in a relational database for later retrieval. From these systems, IR offices, finance and budget offices, planning offices, and others extracted material for reporting, analysis, and forecasting. Decision making became reliant upon a culture of reporting that offered answers to questions in close to real time: how many applicants do we have now compared to the same time as last year? Is our spending for the month in each unit above or below what was budgeted? How many grant applications and for how much money do we have this year compared to the same time last year? Armed with this level of information, university leaders have been better able to adjust tactics and strategy to respond to current situations. Processes to access, analyze, and communicate this information to leadership are neither automatic nor systemically available and required human talent to extract data and transform it into information. Decision-makers at institutions with the resources to invest in personnel devoted to analysis received better intelligence than those that did not.

The data warehouse also came of age in the 1990s as a response to the proliferation of data from transactional systems, which often yielded conflicting reports to senior officials because of issues of timing, differing and siloed analyst expertise, and imprecision in how questions were formulated. Bill Inmon (1992) offered the vision that a data warehouse could provide an organization with a "single version of the truth," and the star schema for warehousing introduced by Kimball and Merz (2000) became a standard still widely in use. From these systems, business intelligence (BI) units emerged on many campuses to provide data for decision

support. BI units have generally been housed in university IT departments and typically provide a data and reporting infrastructure for client units across campus (Drake and Walz 2018). In some colleges and universities, this function is fulfilled by institutional research, in others institutional research is a client of the BI unit, and in some instances IR and BI units compete in providing information to other constituencies. In recent years, one approach has been to combine IR and BI units, and as Childers (2016) observes in an organizational and anthropological case study of such a merger at the University of Arizona, opportunities for synergy can be counterbalanced by cultural and disciplinary differences among personnel and even unit missions.

Setting the Stage for AI

Three subsequent advances led to the explosion of data in the last two decades that have set the stage for aggressive and increasingly prevalent use of machine learning and AI: Near universal internet coverage, ubiquitous handheld devices, and the use of these devices for social media, internet access, and mobile applications. In the 1990s, transportation of data was accomplished through hard-wire connections on-campus, and at times via floppy disk and slower dial-up connections across campuses. In the following decade, extensive deployment of high-speed optic cable and high-speed internet access made sharing of larger data files convenient and cost effective, especially as creation of application programming interfaces (APIs) became standard practice among system developers. Satellite networks and mobile towers also contributed to increased connectivity to support the second key advance: the advent of the smartphone. Since the launch of the iPhone in 2007, which extended the email functionality of the BlackBerry to full internet and web browsing access, an estimated 5.1 billion unique mobile

users were active in 2018 with over 4 billion of them accessing the internet (Kemp 2018). The astonishing magnitude of this number of users becomes dwarfed when considering the amount of data each user generates as he or she browses web sites, accepting tracking cookies, allowing data sharing among organizations, and providing data through "private" forms, transactions, and public posts. Effectively, every interaction even down to the click and keystroke is digitized and becomes data AI needs to construct models and make predictions. It is this revolutionary social and transactional feature of the internet, enabled by Google, Facebook, and Twitter, that opened the world of big data to global corporate giants as an avenue to generate profits. And on a smaller scale, the university through its administrative systems, LMS, and web site, collect data on students, faculty, staff, and visitors – data that are now available to identify patterns and use them to predict future outcomes. It is important to recognize that data collection is more prevalent than user-system interaction. Large stores of passive data are also being collected if not yet substantively used, including digital video files from hundreds if not thousands of video security cameras, location and time tracking from the nodes of the wireless network, and repositories of license plates photographed with time stamps of vehicles that enter and exit parking facilities.

This amount of data exceeds the capacity and design of the traditionally structured data warehouse, and now approaching 2020, college and university officials find themselves at the cusp of moving to more flexible data environments. Web 2.0 companies like Google, Facebook, and Amazon shifted away from data warehouses with the snowflake or star schema to data environments allowing distributed storage of disparate data types. These platforms, commercially available as products like Hadoop, SANA, and Amazon Web Services (AWS) offer a No-SQL environment in a non-relational database, allowing for storage of unstructured

data (e.g. Twitter feeds, video files, course assignments from the LMS) alongside structured data. The organization reflects an environment sometimes termed a "data lake" where data from all sources flow into a single location and then when needed are extracted, loaded and transformed (ELT) to fit the analytical use case. This model for a university was outlined by computer scientists at Clemson University earlier in the decade (Ngo et al. 2012) and is in implementation at institutions like Arizona State University (Wishon and Rome, 2016), Virginia Tech (Campbell, Smith and Kumar 2018), and the California State University System (Aldrich 2018). Kellen (2019) observes that data modeling and design operate under new rules in the data lake environment: entity-based relationship modeling in star schema is replaced by streams of data in narrow tables with millions if not billions of records that represent a "replayable log." Enabled by super-fast computing power, data architecture can represent maximum sematic complexity and be designed to capture everything, not simply data elements assumed *a priori* to hold operational importance. This amount of data becomes critical to machine learning, which requires tremendous amounts of "big data" to make predictions, and it is with the advent of storage of this sort that will enable effective AI to begin to make predictions to guide university operations.

How the New Data Ecology Impacts Decision-Making

These developments have prompted numerous institutional transformations that impact decision making. Data have become "democratized" in that they are widely available and comprehensible for campus constituencies to analyze and make their own decisions. Basic activities of departments and offices have transformed as they adapt to process, understand, and use these data. Individual students are assigned probability scores for "risk" to predict likelihood

they will need support, and individual faculty are assigned productivity scores to assist with personnel decisions and unit development. Detailed data from sensors in buildings, from every online interaction or chat, even from metadata associated with stored files can provide the raw material for AI to assist with monitoring operations and decision-making. Digitization of virtually everything has systems explosion of data beyond the capacity to curate them.

Business Intelligence

At a number of higher education institutions in the US, business intelligence units have pushed data to academic and administrative units so that they can directly access operational data about themselves. Some of this is simple: real-time budget information, real-time class rosters, and lists of advisees. Lists of individuals or expenditures represent some of the simplest information, but they can also be aggregated into reports than can assist with management, including grade distributions and analysis of courses where students struggle, student-faculty ratios and dashboards of faculty assignments to monitor educational quality and equitable work load, and spending per credit hour or FTE to monitor efficiency. Placed in the hands of unit leaders, such information assists with solid decision-making, but it also challenges the "one version of the truth" vision from the centralized data warehouse. For example, Patti Barney, the Vice President for Information Technology at Broward County Community College reported that implementation of a business intelligence system, a robust employee training initiative, and a cultural shift to value data-driven decision-making significantly improved effectiveness of decision-making:

> We no longer have management meetings where there are arguments over whose version of the data is correct, whose spreadsheet is correct, or whose report is correct … Now we

see the metrics generated from the business intelligence applications and we can focus on

the business issues we are facing, and what's preventing us from meeting key objectives,

and what's the real impact on meeting the needs and success of the students (Halligan

2009, 15).

Even if these data emerge from centralized sources in pre-programmed reports, however, the

potential for the data to be re-combined, filtered, analyzed over different time periods, and linked

to other data sources leads to the eventuality that the unit will look at the data in a different way

that others in the organization view the same data. Thus, localized decision making can assist

with stronger management in the unit, but can also add to confusion and opacity from a

centralized perspective.

How Units Are Changing to Adopt AI

Additionally, the advances in data storage, computing power, and analytics have

transformed university departments and business units from reliance upon paper file clerks to

essential operations conducted by data professionals. Registrar functions and financial aid

processing are the most salient examples, with the oaken file cabinets of the 1950s replaced with

computer terminals for running batch processing for schedule preparation, prerequisite checks,

and degree audits. Similarly, financial aid officials extract data from federal systems and run

numerous programmed procedures to construct and optimize aid packaging, which is digitally

communicated to the student and the bursar. Academic affairs officers use digital systems to

review promotion and tenure; research administration offices maintain digital faculty profiles to

identify optimal teams to pursue funding opportunities harvested from digital sources; and

facilities and operations divisions monitor the physical plant in real time for energy usage, work

order completion, and video footage. Because university personnel increasingly interact with the processes of the university as digital representations, these employees require increased data literacy if not competency. The less obvious issues might include how big data will force the transformation of campus IT and data management infrastructure, how third party software systems will disrupt the data management ecology on campus, how people who are more accustomed to working in silos must learn to collaborate more in the new digital environment, and what senior leaders need to know about how rapid changes in technology and digital content management process will mean to campus decision making.

Making Decisions about Individuals Based on a Score

The advent of machine learning and AI on top of big data and unit-level distribution of information is the introduction of the personalized "risk score" for educational progress or perhaps better phrased as predicted need for support. Much like credit scores that were introduced in the late 1980s by Fair, Isaac, and Company (Kaufman 2018) to help financial institutions evaluate credit worthiness of borrowers summarizing all of their credit history into a single number, the 2010s have seen the introduction of individualized scores for students signaling their propensity to perform well and remain enrolled. Companies like EAB (what does this stand for), which has its roots in the health care industry and population health care management through data analysis, and Civitas, which evolved more directly from higher education, offer services to colleges and universities to use their data to establish individual-level predictions of success, and various institutions have also approached this problem using their own resources (Arnold 2010). These approaches make use of demographic data, course performance data, and how students interact with the learning management systems. In some

instances, non-academic data are brought to bear, such as Purdue's Forecast App which uses wireless node data tracking students' cell phones to locate when students are physically present in academic buildings and how this relates to their success (Blumenstyk 2018).

Generally, models that create individualized scores use data mining techniques (including decision trees, clustering, and other methods discussed in chapter 3 of this book) to predict student success (discussed in chapter 8) or even their likelihood to enroll at all (discussed in chapter 7). Such models are extensible to the advancement function and faculty recruitment and development. These sorts of predictions, of course, have been made in the admissions and hiring process very often, but with machine learning more data are brought to bear, ostensibly with a better chance of being correct more often than the use of more limited data combined with human judgement. They also bring with them potential for bias and replicating social inequality, and so require some attention to what data are included and how the output of the algorithms treat various populations (discussed in chapter 4). The end result, however, is that personnel in university units (such as academic advisors, advancement officers, department chairs and hiring committees) have specific and actionable information to work with individual students, donors, and faculty, so that the decisions they make can be tailored to individual needs. That said, university professionals will need to learn how to work with these new sources of data and how to treat probabilities and propensities assigned to individuals by machine learning algorithms, and universities will need to systemically learn how to make use of these without inadvertently causing harm.

Using AI to Improve Teaching and Learning

Chapter 2-14

The digital enhancement of teaching and learning through extensive use of learning management systems (LMS) like BlackBoard and the complete digitization and delivery of courses online has also generated prolific amounts of data. When these data are harnessed and integrated with AI, they can offer insight into how students learn, what pedagogical methods will be effective and with whom, and how to adjust the teaching and learning process in real time (Popenici and Kerr 2017). For example, members of the Unizin consortium founded in 2014, which includes almost a million students across eleven member institutions, share data across a common learning platform called Canvas to become a massive learning laboratory. Partner institutions realize savings by sharing resources and also have access to anonymized data that allows for identification of general trends as well as class and assignment-level data for their own students. By 2018, the system had exhibited capacity to predict course-level success for some students within half a letter grade (Kafka 2019). The full-scale application of AI to identify markers for student success has potential to individualize instruction for each learner in much the way the medical field is being transformed by personalized medicine. Other examples include Georgia Tech's transformation of some its graduate programs to incorporate massively online open courses (MOOCs), which include experimentation with computerized teaching assistants that appear to pass the Turing test, at least as far as student-teaching assistant interactions are concerned (Gose 2016). In another instance, the Mandarin Project at Rennselaer Polytechnic Institute teaches students Chinese in an interactive virtual world in which students can interact with simulated AI speaking in Mandarin and providing feedback about their performance (McKenzie 2018a). These uses of AI hold significant promise for improving the teaching and learning process but also raise ethical questions, and the use of algorithms always carries potential for bias and unintended consequences (O'Neil 2016). As AI is considered for student

learning on campus, a number of important questions can be asked, including: Will student-level predictions of struggling in a course incentivize institutions to limit access for some students? Will they create self-fulfilling prophecies? Will AI teaching assistants neglect some groups of students who need extra help or help in different ways? Will virtual learning environments limit access to educational opportunities for students with disabilities? What happens when a large data breach occurs releasing terabytes of educational records and detailed information about how individuals struggled in class? Additional discussion on learning analytics is included in chapter 10 of this volume.

Using AI to Improve Campus Operations

Machine learning and AI are by no means limited to predicting individual success. Example are already prevalent of how AI has allowed universities to replace or re-purpose human decision-making by substituting AI and using the Internet of Things (IoT). For instance, the University of Texas at Austin has successfully used AI to use extensive sensor networks and climate data to operate its sprinkler system and the University of Iowa has connected elaborate sensor networks in its buildings to detect potential maintenance and failure issues before they become problems (Gardner 2018). At Georgia State University, and discussed further in chapter 8 of this book, the university officials have successfully deployed automated admissions assistants (Gardner 2018). Community colleges in North Carolina rely on AI to connect, structure, and curate their digital content to improve real-time delivery of organizational learning across campuses (Schwartz 2019). In many of these instances, the technology and manner of use is novel enough to warrant news coverage and appear like the "nice toys" that Roy Batty references in *Blade Runner*, but caution is warranted to assume widespread organizational

transformation. It falls within institutional interests to publish and promote success stories and bury failures. Systematic investigation of the return on investment for AI in higher education has yet to be done, and even in the private sector, the technology consulting firm Gartner predicts that four out of five AI projects "will remain alchemy, run by wizards whose talents will not scale in the organization" (White 2019). Nevertheless, these deployments of AI and machine learning can be expected to increase exponentially, with some of the successful and scalable implementations resulting in significant monetization.

Data Proliferation and Data Governance

As functional areas of the university evolve and adopt their own systems that digitize more and more interactions, too much data are created to curate. Driven by the question, "do you trust your data?" colleges and universities have rushed to advance their systems for data governance and data quality over the past decade, in part because the advent of inexpensive and widespread analytics on top of all data sources has exposed areas of invalid and incomplete data. These efforts are clearly needed, but as data continue to proliferate, there is a real possibility that some data used by university officials remain outside of a governance structure, or are allowed to remain in a state that accepts higher levels of error because the errors can be accounted for through AI prediction algorithms. As data proliferate and grow into big data and bigger data, the question "do you trust your data?" will likely be supplanted with the question "do you trust your algorithms?" leading to an entire additional set of higher-order control activities for monitoring the function of machine learning and AI. Data lake solutions adopted by institutions like Virginia Tech and the Cal State System attempt to address some of these issues, and while these solutions have clearly added value in the private sector, successful scale-based applications in higher

education are still too early in development as of the writing of this chapter to demonstrate return on investment.

Given the resources and expertise required to design such systems, some institutions have turned to third party vendors to provide predictive modeling, data integration, and even data lake environments. For instance, Rapid Insight was founded in 2002 to provide easy to use predictive analytics for financial aid modeling. Later in the 2000s, the Education Advisory Board (now simply "EAB") grew out of a health care analytics operation to apply similar tools to student success, including an integration of service and usage data, with a host of competitors such as Civitas, Blue Canary, and ZogoTech. More recently, firms like HelioCampus and Snowflake, offer colleges and universities data ecosystems ranging from traditional data warehousing solutions to full-blown data lakes integrating as many as 30 additional products with the ERP, with the data models ostensibly pre-built. Especially where campuses lack maturity in data governance and management, such solutions appear attractive, although adoption of such products without substantial attention to organizational culture and process reform risks simply pushing poor-quality data from one system to another. Even where the garbage-in, garbage-out problem is overcome, institutional leaders should not be naïve to think that cutting a check to a technology company will substitute for the hard work of organizational transformation.

The Road Ahead for AI in Higher Education

The implications of widespread AI and machine learning in higher education certainly deserve some consideration. If AI can drive a car, then can AI run an institutional research office? Push-button compliance reporting for IPEDS and other requirements has been a promise from many ERP systems, though rarely delivered because of institutional customizations. Such

automation seems feasible with improved data governance and data quality, but successful

application may still be far into the future. It also seems feasible that AI could be programmed to

automate evaluation and optimize student success, course scheduling, faculty work load, and a

range of other functions that extend beyond compliance reporting. Scores of vendors have

supplied and continue to supply solutions of this sort to universities, but none of them run

completely automatically, and when they work (and they do not always fully deliver on lofty

promises) they often require additional, not fewer, personnel to supply data, evaluate output, and

ensure modifications are programmed into the applications. Further, because AI functions as a

"prediction machine" that uses a set of historical data to predict the future, it is particularly ill-

equipped to handle new situations and smaller groups of students. Questions like "is this new

ranking of our university something to which we should pay attention or publicize?" "How will

this proposed regulatory change affect our operations?" and "How will addition of this programs

position us in the marketplace?" are complex questions for human analysts to tackle but they lie

beyond the scope of today's AI to address.

However, assume for a moment that AI is able to substantially replace an institutional

research office and a business intelligence unit – could it replace other administrative functions

like a registrar's office or a financial aid office? Indeed, vendors already have solutions to

optimize financial aid distribution for recruitment, student success, etc. What advancements are

needed to begin replacing personnel so that students interact with virtual financial aid counselors

in adjusting financial aid packages? Can the decisions that a financial aid counselor make be

replaced by AI with perhaps a summary review from an associate director or director level

position? Can the financial aid director be replaced and dispute resolution simply be outsourced?

More provocatively, could AI replace executive functions – such as a provost or a president? In some respects, the AI presidency would carry a number of advantages. It should exhibit fewer ethical lapses (barring those of its programmers), fewer instances of sexual or personal misconduct, and fewer propensities to make decisions based on bias or personal preference. Conversely, AI could account for far more information about the organization than a college president could ever absorb and use this information to make decisions. All the sentiment from Twitter, Instagram, etc. could be monitored real-time to understand reaction and decisions made accordingly to optimize faculty work load, student enrollment, staffing levels, and the list goes on *ad naseum*. But again, when confronted with new situations, the AI of today struggles because when new situations arise, it simply does not have historical data to provide reliable predictions about the future. There is also the very real challenge that an organization of human beings would resent or resist a set of computer algorithms leading their mission-driven educational activities (even more than human leadership of such activities is resisted). Nevertheless, it seems well within the realm of possibility that a governing board could set up an AI system to function as a check or regulating counter-balance to the president, or a president could have machine learning algorithms to monitor decisions of a provost or other vice president, offering independent predictions for what the institution could or should do. Such a working arrangement would again present challenges with an AI system apparently second-guessing institutional leadership, but in some ways this dynamic exists today, just relying principally on human judgement rather than AI predictions.

Despite the directionality of some of these questions, the continuing adoption of AI on campus likely does not point toward a dystopian world where the machines are in control or their potential self-awareness is perceived as a threat to be eradicated, as suggested by the *Blade*

Runner (1982) reference to technology in 2019 that frames this chapter. In part, evolution of educational systems and structures is by nature iterative and slow, and even the example of machine grading illustrates the slow and uneven adoption of the technology from Project Essay Grade in 1968 into educational environments, and when the computer-based systems make mistakes, they make the news and stir controversy (McKenzie 2018b). Such occurrences suggest that much like with acceptance of self-driving cars by society at-large, adoption of AI in higher education may be more of a social science problem than a technological problem. The way that higher education professionals make decisions will undoubtedly change to incorporate these innovations, and in general should continue to improve the quality and effectiveness of higher education, even as students are challenged to master an increasing body of skills, knowledge, values, and dispositions.

References

Agrawal, A, J. Gans, and A. Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press.

Aldrich, B. 2018. "Lead the Charge: CSU's Transformational Data Program." Paper presented at Educause Annual Conference, Denver, CO, 2018. Accessed Nov. 28, 2018. https://events.educause.edu/annual-conference/2018/agenda/lead-the-charge-transformational-data-leadership.

Arnold, K. 2010. "Signals: Applying Academic Analytics." *Educause Review* 3. Accessed January 29, 2019. https://er.educause.edu/articles/2010/3/signals-applying-academic-analytics.

Blumenstyk, G. 2018. "Big Data is Getting Bigger. So are the Privacy and Ethical Questions." *The Chronicle of Higher Education*. July 31, 2018. https://www.chronicle.com/article/Big-Data-Is-Getting-Bigger-So/244099.

Campbell, J, K. Smith, and T. Kumar. 2018. "Building an Analytics Infrastructure In-House." Paper presented at the Association of Public and Land-Grant Universities Annual Meeting, New Orleans, LA, 2018. Accessed November 24, 2018. http://www.aplu.org/members/commissions/information-measurement-analysis/cima-presentations-2018/In-House%20data%20infastructure.pdf.

Chamberlin, D and R. Boyce. 1974. "SEQUEL: A Sructured English Query Language." *Proceedings of the 1974 ACM SIGFIDET Workshop on Data Description, Access and Control*, 249–64. New York, NY: Association for Computing Machinery. Accessed December 29, 2018. http://www.joakimdalby.dk/HTM/sequel.pdf.

Childers, H. 2016. "Can Business Intelligence and IR Teams be Combined?" Paper presented at

the Annual Forum of the Association for Institutional Research. New Orleans, LA.

Codd, E. 1970. "A Relational Model of Data for Large Shared Data Banks." *Comm. Association*

for Computing Machinery 13, no. 6: 377-87.

Drake, B. and A. Walz. 2018. "Evolving Business Intelligence and Data Analytics in Higher

Education." In *IR in the Digital Era. New Directions for Institutional Research* 178, 85-

97, edited by C. Mathies & C. Ferland. Boston: Wiley.

Gardner, L. 2018. "How A.I is Infiltrating Every Corner of Campus." *The Chronicle of Higher*

Education. April 08, 2018. https://www.chronicle.com/article/How-AI-Is-Infiltrating-

Every/243022.

Gilbert, S. and Green, K. 1986. "New Computing in Higher Education." *Information Technology*

Quarterly 5, no. 2: 10-23.

Gladwell, M. 2005. *Blink: The Power of Thinking without Thinking*. New York: Little Brown &

Co.

Gose, B. 2016. "When the Teaching Assistant is a Robot." *The Chronicle of Higher Education.*

October 23, 2016. https://www.chronicle.com/article/When-the-Teaching-Assistant-

Is/238114.

Halligan, T. 2009. "Depending on Data: Business Intelligence Systems Drive Reform."

Community College Journal 80, no. 3: 14-15.

IBM. 2003. "System/360 Model 91." Accessed January 24, 2019.

https://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP2091.html.

Inmon, W.H. 1992. *Building the Data Warehouse*. New York: John Wiley & Sons.

Kafka, A.C. 2019. "Colleges are Banding Together Digitally to Help Students Succeed. Here's How." *Chronicle of Higher Education*. March 31, 2019.
https://www.chronicle.com/article/Colleges-Are-Banding-Together/246002.

Kaufman, R. 2018. "The History of the FICO® Score." Blog post. Accessed April 7, 2019.
https://blog.myfico.com/history-of-the-fico-score/.

Kemp, S. 2018. "Global Digital Report." Blog post. Accessed January 29, 2018.
https://wearesocial.com/us/blog/2018/01/global-digital-report-2018.

Kellen, V. 2019. "21st Century Analytics: New Technology and New Rules." *EDUCAUSE Review* 54, no. 2: 31-39.

Kimball, R. and R. Merz. 2000. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. New York: Wiley and Sons.

McCarthy. J., M.L. Misnky, N. Rochester, and C.E. Shannon. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." Accessed December 18, 2018. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.

McKenzie, L. 2018a. "Pushing the Boundaries of Learning with AI." *Inside Higher Ed.*
September 26, 2018. https://www.insidehighered.com/digital-learning/article/2018/09/26/academics-push-expand-use-ai-higher-ed-teaching-and-learning.

McKenzie L. 2018b. "Autograding System Goes Awry, Students Fume." *Inside Higher Ed.*
November 30, 2018. https://www.insidehighered.com/news/2018/11/30/autograder-issues-upset-students-berkeley.

Ngo, L.B., A.W. Apon, P. Xuan, K. Ferguson, and C. Marshall. 2012. "An Infrastructure to Support Data Integration and Curation for Higher Educational Research." TigerPrints

Publications 5. Accessed January 15, 2019.

https://tigerprints.clemson.edu/computing_pubs/5.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequalities and Threatens Democracy.* New York: Penguin/Random House.

Page, E.B., G.A. Fisher, and M.A. Fisher. 1968. "Project Essay Grade: A FORTRAN Program for Statistical Analysis of Prose." *British Journal of Mathematical and Statistical Psychology* 21: 139.

Page, E.B. and D.H. Paulus. 1968. *The Analysis of Essays by Computer. Final Report.* Washington, DC: U.S. Office of Education, Bureau of Research.

Paulus, D.H., J.F. McManus, and E.B. Page. 1969. "Some Applications of Natural Language Computing to Computer Assisted Instruction." *Contemporary Education* 40: 280-285.

Popenici, S. and S. Kerr. 2017. "Exploring the Impact of Artificial Intelligence on Teaching and Learning in Higher Education. *Research and Practice in Technology Enhanced Learning* 12, no. 22: 1-13.

Schwartz, N. 2019. "How Colleges are Using AI to Save Time on Operations." Education Dive. February 27, 2019. https://www.educationdive.com/news/how-colleges-are-using-ai-to-save-time-on-operations/549277/.

Shannon, C. 1987. "Interview by Anthony Liversidge." *Omni Magazine* (August): 61-66.

Simonite, T. 2018. "Apple's latest iPhone's are Packed with AI Smarts." *Wired*. September 12, 2018. https://www.wired.com/story/apples-latest-iphones-packed-with-ai-smarts/.

Turing, A.M. 1950. Computing Machinery and Intelligence, *Mind* 59, no. 236: 1, 433-460.

White, A. 2019. "Our Top Data and Analytics Predicts for 2019. Blog post. January 3, 2019.

      https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-

      for-2019/.

Wishon, G. and J. Rome. 2016. "Institutional Analytics and the Data Tsunami." *Educause*

      Review. December 12, 2016. https://er.educause.edu/articles/2016/12/institutional-

      analytics-and-the-data-tsunami.