Artificial Intelligence in Higher Education: Applications, Promise and Perils, and Ethical Questions

By Elana Zeide
What is artificial intelligence? In any discussion of artificial intelligence (AI), this is almost always the first question. The subject is highly debated, and I won’t go into the deep technical issues here. But I’m also starting with this question because the numerous myths and misconceptions about what artificial intelligence is, and how it works, make considering its use seem overly complex.

When people think about artificial intelligence, what often comes to mind is The Terminator movies. But today we are far from machines that have the ability to perform the myriad of tasks even babies shift between with ease—although how far away is a matter of considerable debate. Today’s artificial intelligence isn’t general, but narrow. It is task-specific. Consider the computer program that infamously beat the world’s champion in the Chinese game Go. It would be completely befuddled if someone added an extra row to the playing board. Changing a single pixel can throw off image-recognition systems.

Broadly, artificial intelligence is the attempt to create machines that can do things previously possible only through human cognition. Computer scientists have tried many different mechanisms over the years. In the last wave of AI enthusiasm, technologists tried emulate human knowledge by programming extensive rules into computers, a technique called expert systems. Today’s artificial intelligence is based on machine learning. It is about finding patterns in seas of data—correlations that would not be immediately intuitive or comprehensible to humans—and then using those patterns to make decisions. With “predictive analytics,” data scientists use past patterns to guess what is likely to happen, or how an individual will act, in the future.

All of us have been interacting with this type of artificial intelligence for years. Machine learning has been used to create GPS systems, to make translation and voice recognition much more precise, to produce visual digital tools that have facial recognition or filters that create crazy effects on Snapchat or Instagram. Amazon uses artificial intelligence to recommend books, Spotify uses machine learning to recommend songs, and schools use the same techniques to shape students’ academic trajectories.

Fortunately—or not, depending on one’s point of view—we are not at the point where humanoid robot teachers stand at the front of class. The use of artificial intelligence in education today is not embodied, as the roboticists call it. It may have physical components, like internet of things (IoT) visual or audio sensors that can collect sensory data. Primarily, however, educational artificial intelligence is housed in two-dimensional software-processing systems. This is perhaps a little less exciting, but it is infinitely more manageable than the issues that arise with 3-D robots.

In January 2019, the Wall Street Journal published an article with a very provocative title: “Colleges Mine Data on Their Applicants.”

The article discussed how some colleges and universities are using machine learning to infer prospective students’ level of interest in attending their institution. Complex analytic systems calculate individuals’ “demonstrated interest” by tracking their interactions with institutional websites, social media posts, and emails. For example, the schools monitor how quickly recipients open emails and whether they click on included links. Seton Hall University utilizes only about 80 variables. A large software company, in contrast, offers schools dashboards that “summarize thousands of data points on each student.” Colleges and universities use these “enrollment analytics” in determining which students to reach out to, what aspects of campus life they should emphasize, and assessing admissions applications.

**AI Applications**

Figure 1 shows a summary of the different kinds of applications that currently exist for artificial intelligence in higher education. First, as I’ve discussed above, is institutional use. Schools, particularly in higher education, increasingly rely on algorithms for marketing to prospective students, estimating class size, planning curriculum, and allocating resources such as financial aid and facilities.

This leads to another AI application, student support, which is a growing use in higher education institutions. Schools utilize machine learning in student guidance. Some applications help students automatically schedule their course load. Others recommend courses, majors, and career paths—as is traditionally done by guidance counselors or career services offices. These tools make recommendations based on how students with similar data profiles performed in the past. For example, for students who are struggling with chemistry, the tools may steer them away from a pre-med major, or they
Today’s artificial intelligence is based on finding patterns in seas of data—correlations that would not be immediately intuitive or comprehensible to humans—and then using those patterns to make decisions.
The promise of AI applications lies partly in their efficiency and partly in their efficacy. AI systems can capture a much wider array of data, at more granularity, than can humans. And these systems can do so in real time.

may suggest data visualization to a visual artist.

Another area for AI use in student support is just-in-time financial aid. Higher education institutions can use data about students to give them microloans or advances at the last minute if they need the money, for example, to get to the end of the semester and not drop out. Finally, one of the most prominent ways that predictive analytics is being used in student support is for early warning systems, analyzing a wide array of data—academic, nonacademic, operational—to identify students who are at risk of failing or dropping out or having mental health issues. This particular use shows some of the real advantages of artificial intelligence—big data can give educators more holistic insight into students’ status. Traditionally, an institution might use a couple of blunt factors—for example, GPA or attendance—to assess whether a student is at risk. AI software systems can use much more granular patterns of information and student behavior for real-time, up-to-the-minute assessment of student risk. Some even incorporate information such as when a student stops going to the cafeteria for lunch. They can include data on whether students visit the library or a gym and when they use school services. Yet while these systems may help streamline success, they also raise important concerns about student privacy and autonomy, as I discuss below.

Lastly, colleges and universities can apply artificial intelligence in instruction. This involves creating systems that respond to individual users’ pace and progress. Educational software assesses students’ progress and recommends, or automatically delivers, specific parts of a course for students to review or additional resources to consult. There are often called “personalized learning” platforms.
The promise of AI applications lies partly in the efficiency and partly in their efficacy. AI systems offer excellent observations and inferences very quickly and at minimal cost. These efficiencies will hopefully lead to increased efficacy—to more effective teaching, learning, institutional decisions, and guidance. So this is one promise of AI: that it will show educators things they can’t assess or even envision given the limitations of human cognition and the difficulty of dealing with many different variables and a wide array of students.

Given these possible benefits, the use of artificial intelligence is also being framed as a potential boom to equality. With the improved efficacy of systems that may or may not require as much assistance from humans or necessitate that students be in the same geographical location, more students will gain access to better-quality educational opportunities and will perhaps be able to network with peers in a way that will close some of the achievement gaps that continue to exist in education. Lastly is the promise of a more macrolevel use of artificial intelligence in higher education to make gains in pedagogy, to see what is most effective for a particular student and for learning in general.

The use of artificial intelligence in higher education also involves perils, of course. One is the peril of adverse outcomes. Despite the intention of the people who develop and use these systems, there will be unintended consequences that are negative or that can even backfire. To avoid these adverse outcomes, you should take into account several different factors. One of the first to consider is the data that these tools draw upon. That data can vary in quality. It may be old and outdated. Or it may be focused on and drawn from a subset of the population that may not align with the students being targeted. For example, AI learning systems that have been trained on students in a particular kind of college or university in California may not have the same outcomes or reflect the same accuracy for students in another part of the country. Or an AI system that was based on Generation X students may not have the same efficacy for native digital learners.

Another data aspect concerns comprehensiveness. Does the data include information about a variety of students? There has been much discussion about this recently in terms of facial recognition. Scholars looking at the use of facial recognition by companies such as Google, IBM, Microsoft, and Face++ have shown that in many cases, these tools have been developed using proprietary data or internal data based on employees. The tools are much more accurate for light-skinned men than light-skinned women or darker-skinned men. In one study, the facial recognition tools had nearly 100 percent accuracy for light-skinned men but only 65 percent accuracy for dark-skinned women. Joy Buolamwini, a co-researcher of this study, created her own, much more accurate tool simply by drawing from a broader array of complexion in the training data she used.

Next to consider are the models that are created using this data. This again raises the issue of accuracy. Models are based on correlation; they are not reflective of causation. And as the Spurious Correlations website hilariously demonstrates, there are some wild correlations out there. Some correlations do seem to make intuitive sense, for example that people who buy furniture protectors are better credit risks, perhaps because they are more cautious. But the point of AI tools and models is to show less intuitive, more attenuated correlations and patterns. Separating which correlations and patterns are accurate and which are simply noise can be quite difficult.

Algorithmic bias plays a role here. This is a real concern because it is something that can occur in the absence of discriminatory intent and even despite efforts to not have different impacts for different groups. Excluding a problematic or protected class of information from algorithms is not a good solution because there are so many proxies for things like race and gender in our society that it is almost impossible to remove patterns that will break down along these lines. For example, zip code often indicates race or ethnicity. Also, because artificial intelligence draws from existing patterns, it reflects the unequal access of some of today’s current systems. A recent example is Amazon’s hiring algorithm, which was criticized for being sexist. There is no evidence that Amazon had any intention of being discriminatory. Amazon used artificial intelligence to detect those characteristics that were most indicative of a successful employee, incorporated those characteristics into its algorithm, and then applied the algorithm to applicants. However, many of Amazon’s successful employees, currently
and in the past, were men. So even without any explicit programming, simply the fact that more men had been successful created a model skewed toward replicating those results.1

An additional, often overlooked factor in adverse outcomes is output. Developers’ decisions shape how the insights that AI systems offer are instructed and interpreted. Some provide detailed information on various elements of students’ learning or behavior that instructors and administrators can act on. Other observations are not as useful in informing interventions. For example, one predictive analytics tool estimated that 80 percent of the students in an organic chemistry class would not complete the semester. This was not news to the professors, who still wondered what to do. So it is important to understand in advance what you want to do with the information these tools provide.

A final factor to consider in avoiding the peril of adverse outcomes is implementation, which is also not always covered in the AI debates in the news or among computer scientists. To use these systems responsibly, teachers and staff must understand not only their benefits but also their limitations. At the same time, schools need to create very clear protocols for what employees should do when algorithmic evaluations or recommendations do not align with their professional judgment. They must have clear criteria about when it is appropriate to follow or override computer insights to prevent unfair inconsistences. Consider the use of predictive analytics to support decisions about when caseworkers should investigate child welfare complaints. On the one hand, caseworkers may understand the complex and highly contextualized facts better than the machine. On the other, they may override the system in ways that may reflect implicit bias or have disparate outcomes. The people using these systems must know enough to trust—or question—the algorithmic output. Otherwise, they will simply dismiss the tools out of hand, especially if they are worried that they do not align with their professional judgment.

Ethical Questions

Ethical questions revolve around consequences in terms of different groups and subgroups, educational values, and how AI systems might alter those values.

The Black Box

Unpacking what is occurring within AI systems is very difficult because they are dealing with so many variables at such a complex level. The whole point is to have computers do things that are not possible for human cognition. So trying to break that down ends up creating very crude explanations of what is happening and why.

Invisible Infrastructure

By choosing the variables to be fed into admissions systems or financial aid systems or student information systems, these AI tools are creating rules about what matters in higher education. This leads to an invisible infrastructure. None of this is explicitly considered by the people implementing the infrastructure. The best example is when learning software specifies particular learning outcomes—choices that are at the core of educational and institutional policy. But educators often overlook that fact when they adopt technology, not understanding that doing so is in some ways the equivalent of imposing an entirely different rubric, instead of standards, in the academic attainment.2

Authority Shifts

The entity doing the data collection and visualization is often a private company. That company is thus in charge of many decisions that will have an important impact and that will alter core values of systems in a way that is, again, not always visible. These private companies may be less directly accountable to stakeholders of the educational institutions—in particular, stakeholders such as students. It is important to consider this authority shift, and the shift in incentives, when using these technologies.

Narrowly Defined Goals

Applications that are based on data often promote narrowly defined goals. That is because in order to work, these systems must literally codify the results that are deemed optimal. This leaves less flexibility than is currently the case with human interactions in classrooms and on campuses. An example is acquiring an education broadly versus learning more narrowly. Optimizing learning outcomes—for example, skills acquisitions, better grades, or increased retention—may crowd out more abstract educational goals promoting citizens capable of self-governance or nurturing creativity. The latter are aspects that one could technically, perhaps, represent in data, but doing so involves crude proxies at best. As a result, they may not be measured or prioritized.

Data-Dependent Assessment

Data-dependent assessment raises similar issues. Tools that collect information, particularly based on online interactions, don’t always grasp the nuances that teachers might see in person. Consider the case where a student answers a question incorrectly. A machine will record a wrong answer. An instructor, however, may discount the error if she notices, for example, that the student clearly has a bad cold.

Divergent Interests

A divergent interest is sometimes between technology developers and institutions and sometimes between institutions and students. In the first instance, technology developers have an incentive to develop systems that use more and more data to get results that the developers can claim are more and more accurate. This allows them to show that their systems are making a difference. That may sometimes result in a rush to market or an emphasis on scale—which may not mean...
that the best-quality platforms are being used or that their efficacy is being assessed in any meaningful terms. This is certainly not true for all technology developers, but it is important to note.

More significant, and less obvious, is the divergent interest between institutions and students. The use of predictive analytics and early warning systems is often touted as a way to promote student retention by drawing attention to struggling or at-risk students. That is fine if the college or university is then going to try to ameliorate or prevent that outcome. But doing so is not always in the institution’s administrative interest. In a famous example from a couple years ago, the president of Mount St. Mary’s University, in Maryland, administered a predictive analytics test to see which students were most at risk of failing. The idea was to encourage them to drop out before the university was required to report its enrollment numbers to the federal government, thereby creating better retention numbers and improving its rankings. According to the president, his plan promoted student retention by drawing attention to struggling or at-risk students. That is fine, so is not always in the institution’s administrative interest.

Elements to Consider and Questions to Ask
Several elements need to be considered to ensure that the implementation of AI tools is optimal and equitable:

- **Procurement.** Pay close attention to the technologies and companies that will be most applicable to your particular student body in terms of the contractual obligations to provide data about your students. Make sure that if problems arise, you have contracted with a company that will be responsive to your problems.

- **Training.** Prepare those people who are going to implement and use these tools, and train them in the benefits and shortcomings of the tools.

- **Overnight.** Put in place a continuous process of examining whether the tools are working, whether they are more effective for particular groups of students, and whether they may be giving better numbers but not better outcomes. This is something that is difficult to do but is very important, because these tools can get outdated quickly.

- **Policies and Principles.** Create institutional policies surrounding the implementation of tools that rely on analytics, and cultivate principles that translate those policies into operational steps and actions.

- **Participation.** Get students’ and faculty members’ input about their concerns and what they would like to see from these systems. This step is often overlooked because it is messy and can lead to some controversy, but it generally creates a better result in the long run.

Regarding policies and principles, some of the best I have seen were developed in 2015 by the University of California Educational Technology Leadership Committee. The committee listed six principles, elaborating on each: ownership; ethical use; transparency; freedom of expression; protection; and access/control. In addition, the committee recommended learning data privacy practices that security providers can implement in the areas of ownership, usage right, opt-in, interoperable data, data without fees, transparency, service provider security, and campus security.

Finally, to be successful, anyone considering an AI implementation within higher education should ask six essential questions:

1. What functions does the data perform? You can’t just see a red, green, and yellow light about student success and take that at face value, at least not if you are the one implementing the systems and you want to do so responsibly.

2. What decisions don’t people see? These are decisions not just about the computer processing but also about the categorization and the visualization.

3. Who controls the content? Is it you, or is it the technology provider? How comfortable are you with that? How comfortable are your professors with that?

4. How do you check outcomes in terms of efficacy, in terms of distribution, and in terms of positive and negative outcomes?

5. What gets lost with datafication? I use this word to describe doing these things based on data as opposed to on interpersonal or bureaucratic systems.

6. What—and whose—interests do you prioritize?

There are no easy answers, but asking these questions will give you a template for considering the less obvious aspects of these systems.

**Conclusion**
My final message? Do not surrender to the robot overlords just yet. Keep in mind that for all the hype and buzz, these AI tools are just computer systems. They can go wrong. They are created by humans. Their values are shaped by companies and institutions. Their data is not neutral but is defined by the historical patterns. Be cautious and thoughtful about what you are doing with artificial intelligence, and remember: it’s not magic.

**Notes**