Why Studying the Humanities Is Essential for Designing Artificial Intelligence Systems

By Tom Porter

Four faculty from various disciplines have collaborated to teach a groundbreaking class: a deep dive into the world of artificial intelligence (AI) and the role that ethics plays as this emerging technology transforms nearly every aspect of our lives.

This is the second story in a summer series about the humanities exploring why Bowdoin faculty are committed to teaching literature, languages, the arts, history, philosophy, and religion—and how students benefit.

L-r: Eric Chown, Allison Cooper, Michael Franz, Fernando Nascimento

Ethics in the Age of Artificial Intelligence (DCS 2475), taught last fall, was part of a nationwide initiative by the National Humanities Center to develop new curriculum in response to increasing concern among academics over the ethical implications of AI.

The four faculty members involved are Eric Chown and Fernando Nascimento, from the Digital and Computational Studies Program; Allison Cooper, who teaches film studies and literature; and political scientist Michael Franz. The class consisted of five modules:

Module One, A General Introduction to AI

The aim here was to strip away some of the hype around AI and concentrate on what really matters, says Chown, the Sarah and James Bowdoin Professor of Digital and Computational Studies. A computer scientist whose specializations include cognitive science and robotics, Chown teaches the students how to make their own AI model ("it's surprisingly easy") and explains the strengths and limitations of the technology.

Core to the functioning of AI is classification. "Classification is the thing that every AI model does in one way or another, and we really wanted students to understand the basics of that before we go any further. The essence of AI is you give it input and it produces output by classifying things—maybe by differentiating blue squares from pink circles."

Cute but not real. This baby peacock was generated by AI.

When it comes to systems like generative AI, which creates its own content through programs like ChatGPT and Bard, the technology learns to predict outcomes by using this classification technique, says Chown. "By searching through the internet, the system finds patterns in order to write text or generate an image, for example. This is also how AI systems figure out what videos or other content to recommend to people on social media. The problem is, this is not a perfect system, as it relies on what it can scrape off the 'net," he explains, "so you can end up with bunch of AI-generated junk." (If you want an example of how AI can get stuff wrong, says Chown, look up an image of a baby peacock on Google!)

"It's important for students to be aware of these kinds of limitations, and of the extent to which AI can be used to manipulate our behavior," he stresses, "which is why a humanities-based approach to the subject is so important."

Module Two, Narratives: The Stories We Tell to Make Sense of AI

Taught by Associate Professor of Romance Languages and Literatures and Cinema Studies Allison Cooper, this module is an opportunity for students to explore AI in a broader sense by looking at certain texts and films dealing with the subject. "It was a welcome opportunity to share some of my favorite AI-themed films with the class," says Cooper.

"The touchstone for me was the idea that science fiction stories are more interesting for what they tell us about our current moment, as opposed to some future age when none of us will be here. So, I looked for stories that students would find most interesting because they relate to technologies they're experiencing right now," she explains.

KEANU REEVES LAURENCE FISHBURNE
MATRIX
ON APRIL 2ND THE FIGHT FOR THE FUTURE BEGINS
JOAQUIN PHOENIX AMY ADAMS ROONEY MARA
OLIVIA WILDE SCARLETT JOHANSSON
her
A SPIKE JONZE LOVE STORY
COMING SOON
ff
KLARA AND THE SUN
KAZUO ISHIGURO
WINNER OF THE NOBEL PRIZE IN LITERATURE

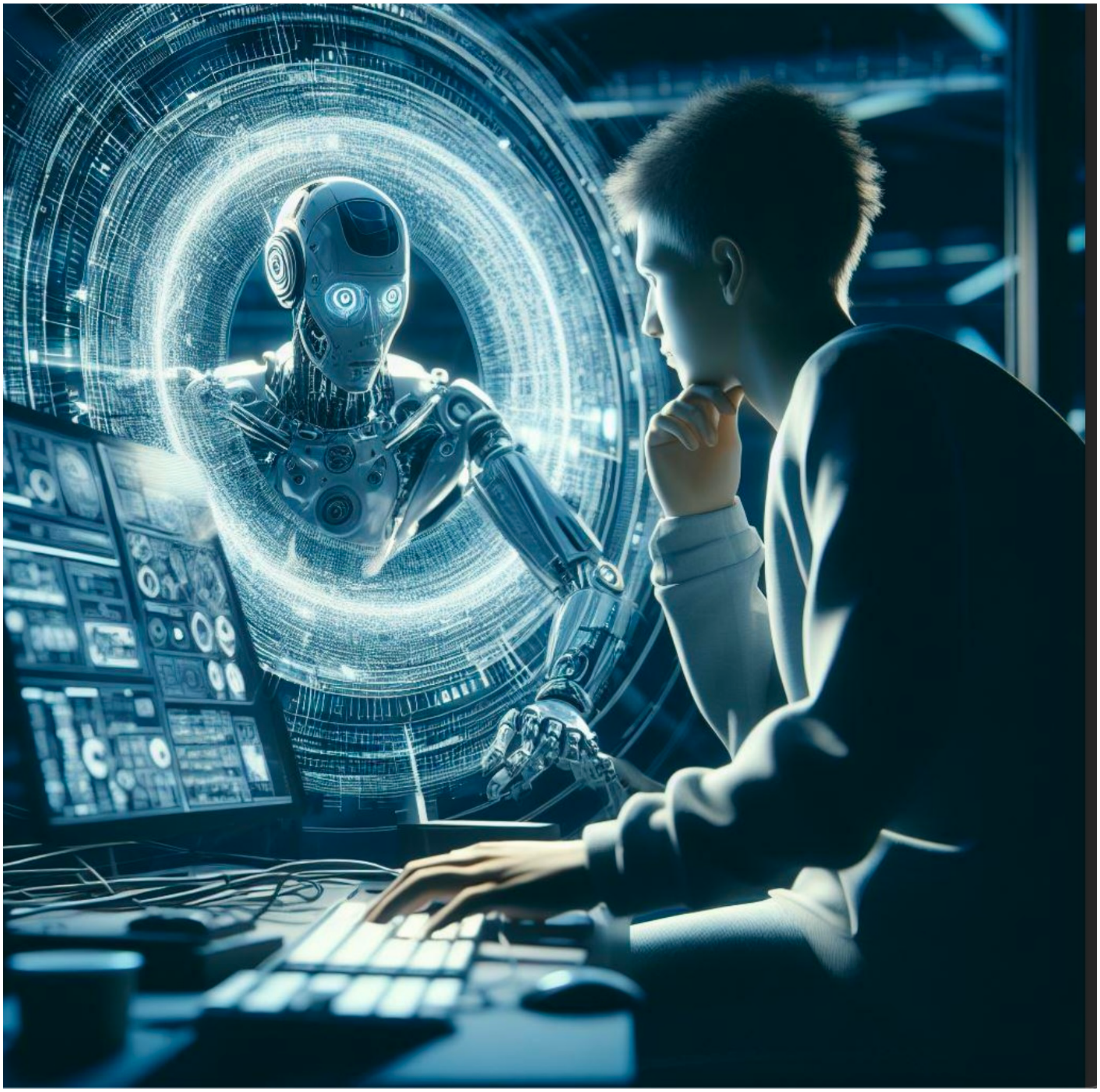
Exploring certain films and texts was an important part of the course

The students looked in depth at two movies and a novel, all of them exploring various aspects of mankind's relationship with technology:

The Matrix (1999): Described by Cooper as "the AI apocalypse movie *par excellence*," it tells the story of human rebels battling against an AI-controlled world, where technology has been allowed to flourish unchecked. "I like *The Matrix* because it challenges the assumption that technology is automatically good for society and benefits everyone. I want students to think about the greater forces that enable AI, the techno elite who benefit, and who might not benefit."

Her (2013): This tale of a man who falls in love with an operating system explores the idea of artificial companionship and what it is that might prompt someone to seek friendship outside of living, human circles. "The film asks some important questions for the class to discuss," says Cooper: "Is technology making us more lonely? Does artificial intelligence have the potential to help humans achieve spiritual and emotional fulfilment? What does it mean to be human?"

Klara and the Sun (2021): This novel by Kashuo Ishiguro, set in a dystopian future, explores the relationship between a sickly child and her "artificial friend," a robot named Klara, from whose point of view the story is told. "Like the movie *Her*, the novel looks at how AI can provide artificial companionship and explores the implications of the idea that the human brain might be just another piece of technology," says Cooper.



The class studies how AI impacts that news that is fed to us on social media

Module Three, The Social Sphere

In this part of the course, Professor of Government Michael Franz looked at the regulation of AI as well as the effects the technology can have on citizens.

On the political front, says Franz, there has been much more progress in Europe when it comes to regulation. Recently, the European Parliament passed the Artificial Intelligence (AI) Act, the first-ever legal framework for dealing with AI, providing for EU-wide rules concerning data quality, transparency, and accountability.

In the US, however, it's a different picture, explains Franz: "We're a fairly polarized country that doesn't do any legislating, so there hasn't been much policymaking on this at a federal level." There has been some action at the state level, he adds, where a number of states, including

California and Connecticut, have enacted legislation designed to protect people from the negative effects of AI systems. The lack of progress at the national level prompted President Biden to issue an executive order in October 2023 to try to ensure that AI systems are safe and secure.

Franz's class also looks at how AI impacts citizens, particularly with regard to the "news" that is fed to us on social media. "With more people, especially the younger population, relying on social media channels to get their news, they are increasingly exposed to what is being fed them by engagement algorithms," he explains. "We look at how manipulation of social media feeds changes the way we are exposed to certain pieces of information and ask how this might be regulated."

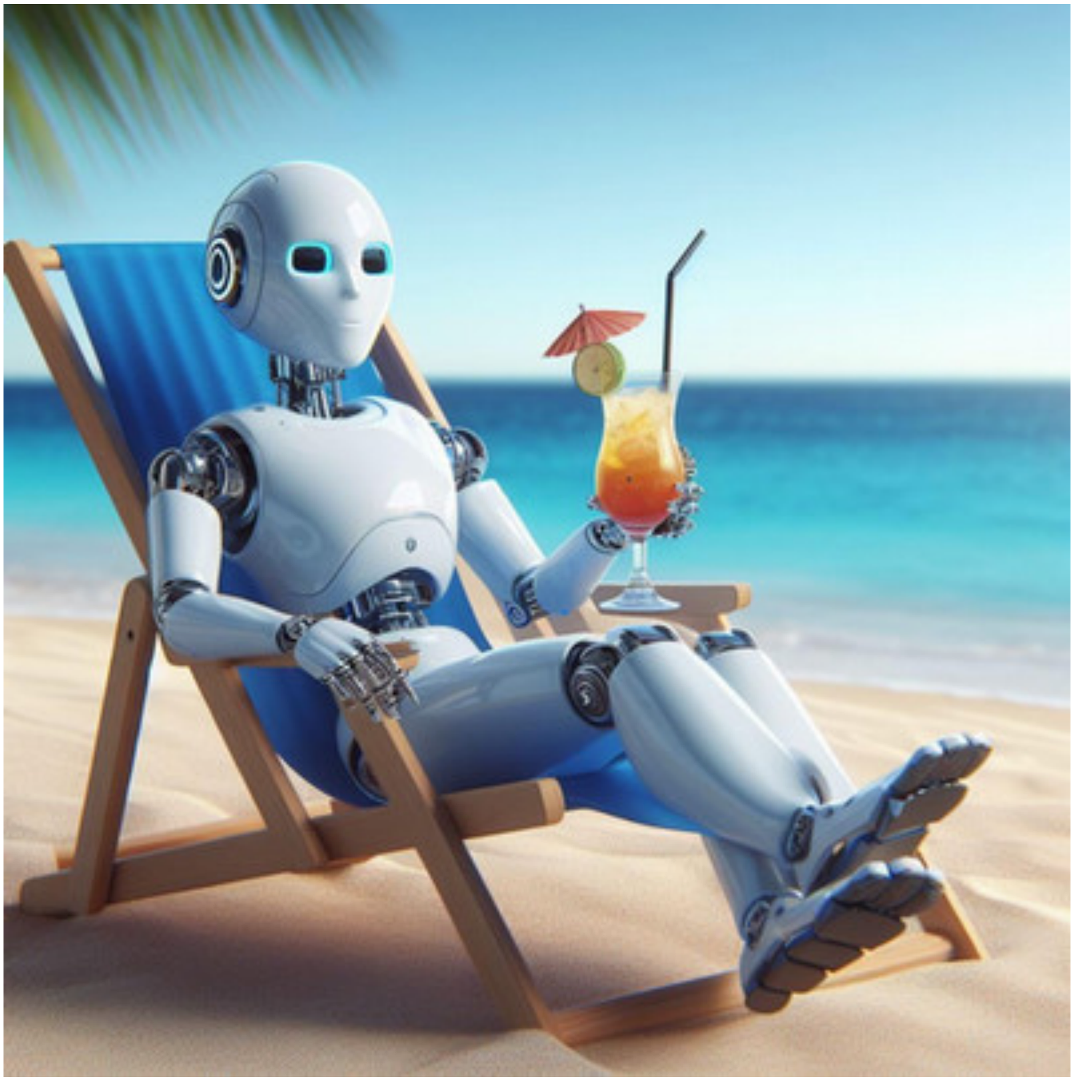
Module Four, The Developers' Sphere

Something that developers of AI systems have to be aware of is that they are building something that possesses a level of autonomy and will, in a way, go on to have a life of its own, explains Assistant Professor of Digital and Computational Studies Fernando Nascimento, who has a background in philosophy as well as computer science.

This can lead to what he calls a "misalignment problem," something that formed a central part of class discussion in this part of the course. "AI develops its own models so has more agency than other systems. Therefore, their societal and ethical implications are much broader than previous digital artifacts," explains Nascimento, who along with Chown coauthored *Meaningful Technologies: How Digital Metaphors Change the Way We Think and Live* (University of Michigan Press, 2023).

"For example, in 2018, Facebook put together a new AI algorithm to select the posts we see in our feeds. The company said the goal was to maximize meaningful relationships, so you would see information that matters to you, from your family, friends, and colleagues." The problem, he adds, is that the AI algorithm modeled "meaningful" according to the reactions and comments of the post. So "meaningful" was unintentionally translated to "emotional," and what stronger emotions are there than fear and hate? "So, instead of promoting harmonious relationships among friends and loved ones, the AI algorithm created polarization because hateful or provocative posts are more likely to get a reaction, get reposted, and promote more traffic. And to make things even more complicated, on top of possible technical misalignments, one has also to consider the alignment of big technology incentives with broader societal goals."

This, says Nascimento, is just one example of why the role of the software developer has added meaning and importance in the age of AI. "The technical side of AI is just one piece of the puzzle. The problem also has to be tackled from the liberal arts perspective," he explains, "involving several layers of society and considering the impact this technology will have on everyone from a diverse and humanistic perspective."



How could AI be used to make our lives better?

Module Five, The Users' Sphere

The concluding segment of the course is designed to make students think about the impact that AI has on their day-to-day lives and what can be done to mitigate the negative effects of this impact. This includes a class activity in which students are invited to design their own social network.

"If you ran Facebook or TikTok and your goal was to get people to stay on your platform all of the time, what kinds of experiments would you run to figure out how to manipulate their attention?" asks Chown. "Well, it turns out that if you design those experiments, which we did, and you go to Facebook's website, that's exactly what they're doing. Our data is under constant scrutiny: how long we look at things, what we click on, and so forth. Essentially, the aim is to hijack our attention, then sell it to advertisers, and they're using AI to do it. So, that's the bad stuff."

However, Chown continues, he doesn't want students to come away from the class with a sense of "existential dread." It's more about encouraging them to act as humans, take control of their digital lives, and be mindful about their use of technology. “Be metacognitive, think about what you're doing: Do I need to go on TikTok right now? Think about interventions in your own life.”

Chown says he and his faculty colleagues thought it important to finish the course on a positive note regarding the future of AI and give students the sense that “there are a lot of smart, like-minded people out there who are working on it. Indeed, some of our students from last year were interested in applying for jobs in this area.”

There is nothing predetermined about the future of AI, conclude Chown and his colleagues. Whether it becomes a force for good or for ill in the future, they say, depends on what we, as humans, do to approach the subject responsibly, and this is where an appreciation of the humanities becomes crucial.

Published July 03, 2024