How does the copyright environment affect what kind of research & scholarship is conducted, and why does this matter for society? For democracy?

I think that I can share from my perspective, as someone who studies political information and political discourse in the digital space, meaning everything from blogs, to websites, to of course social media, but also mass media that’s produced online and shared online. The rules and regulations around copyright, in many instances, wind up making the kind of core scientific endeavor quite challenging, particularly because under the best practices of the scientific method, you would ideally share the core data, the baseline data, so that others can deeply investigate the work that you’ve done and ensure that you’ve followed the proper methodology and replicate it as appropriate.

In the area that I work in, computational social science, we’re doing a lot of things like building out algorithmic models to detect and understand various concepts or various phenomena. Having that underlying data available is essential. It’s really, really important. And so we’re sometimes hindered by the fact that, for example, if I’m using a great deal of data pulled from mass media outlets, news organizations, websites, I very often can’t share that with other researchers. Those who are on my team can also work with that data, but copyright rules and regulations prevent us from sharing it with other researchers outside of our team, and so the scientific process winds up being hindered. The sort of...
questions that we're investigating are about the spread and impact of disinformation. Understanding the types of political discourse that are happening in a variety of spaces online is really essential to our deeper understanding of both the health and wellness, and on the other hand, the harms that are occurring within our democratic societies.

When you find that you can't share your data with other people in order to have them replicate your results and test them, what are your workarounds?

The most straightforward thing that we do is we share URLs, but as anyone who’s worked with digital data knows all too well, URLs are essentially a fleeting commodity. They change, and specific sites or pages tied to a particular URL disappear. Some of that is helped by, for example, the Internet Archive, which creates cached versions of websites; and as much as possible, we try to point people towards those. But ultimately, what winds up happening is we create a waste of resources. Even if all of those URLs did wind up allowing people to find all of the information over again, they still have to do the same things that we were doing initially in many instances, which is scraping the data, or they have to use incredibly expensive databases like Lexis Nexis and so on, that not everyone has access to. So we have this duplication of effort and resources that happens again and again and again, and we don’t have the kind of necessary, or the ideal, streamlining of scientific inquiry that we would like to have.

How is fair use an important principle for academic research and teaching, including studies on the impact of social media companies on society or the use of media content in analysis?

My first answer touched on some of the things that are at the heart of this question. But what I didn’t talk about there was the importance of fair use per se, and fair use in particular for teaching. This is one of the challenges that I’ve come across because there’s so many materials out there that I would really love to use at the core of my teaching. But they’re often difficult to share if I want to follow basic fair-use principles with my students. And so I wind up having to choose other materials to try to play by the rules, essentially.

Could you give us an example?

There are a number of graphs, figures, and digital and visual materials, and it would be really beneficial if I could have my students dig into them, especially for visual materials. But simply posting those materials for the students, it’s effectively impossible for me to do so. These aren’t things that, save through our library, that the students would normally have access to. I pay individually for my own access to those things. And then it means that I’m not able to use those directly as a teaching resource. So students are assigned to dig through these archives and do something in preparation for class, which we then talk about, but it also limits their ability to use those materials for research that they would do for a class paper or something like that. I often feel constrained, particularly in teaching, by the limits of fair use. You can always find a workaround, but what winds up happening is some of the richest educational materials, the richest examples that we would want to draw on, wind up not being available to us. So I find other examples, I find secondary instances that my students can get something out of, but they won’t
be as rich or as ideal as I had hoped.