Revolutionizing Online Learning with Digital Twins

by
Andi Qu

Summary

In today's digital age, online learning is becoming an increasingly important tool for promoting a culture of lifelong learning. Although there is more information than ever before in human history freely available online, computer-aided learning infrastructures have not significantly improved how we learn yet, with massive open online course (MOOC) completion rates remaining below 5%.

The problem is not a lack of online resources but a surplus – a deluge of information. We need a way to navigate all this information, and digital twins – virtual representations of real-world systems that evolve with time – are promising candidates for such a tool.

Imagine you are a college student studying neuroscience who wants to learn electronic circuit design for a personal project. To find a starting point, you search for some kind of tutorial on Google. Immediately, the results overwhelm you – Google returns dozens of tutorials of unknown quality, each assuming a different level of background experience. Given your limited free time, you reconsider venturing outside your comfort zone, but luckily, technology is on your side.

Thanks to a digital twin of your knowledge space, your computer knows your academic background and learning style. It knows about your knowledge of differential equations, experience modeling neural feedback loops, and freshman-year physics classes (which have been untouched for years). Using this information, the digital twin directs you to the middle of an MIT MOOC about analyzing electronic circuits. It skips the calculus review and goes straight to the part about device physics, allowing you to dive into the content without wasting any time.

After you report back to the digital twin, it refers you to a DIY-soil-moisture-sensor project to help you apply and consolidate the skills you just learned. It does this because it can infer the connection between neural feedback loops and electronic sensors. Before long, you're designing original circuits and sharing project ideas online for others to learn from. Even months later, when you need a quick refresher, you know exactly where to go because the digital twin remembers which resources worked best to help you learn.

This scenario may sound like something ripped out of science fiction, but digital twin technology can realistically reshape how we learn new skills via the internet. How would this kind of technology work, and what are some of its potential pitfalls?

Revolutionizing Online Learning with Digital Twins

Maximizing Personal Growth through Customized Learning Journeys

Introduction

In today's digital age, online learning is becoming an increasingly important tool for promoting a culture of lifelong learning. There is more information than ever before in human history freely available online, and websites like Wikipedia exist to make that information as accessible as possible. Despite these advances, computer-aided learning infrastructures have not significantly improved learning yet – massive open online course (MOOC) completion rates remain below 5% [16], and adaptive tutor platforms have seen, at best, modest success when implemented in schools [1].

The problem is not a lack of online resources but a surplus – a deluge of information. Some resources are good, some are bad, and even the good ones are not appropriate for everyone. We need a tool to help navigate all this information, which I envision as digital twins of people's knowledge spaces that can suggest personalized learning paths. Digital twins are virtual representations of real-world systems that use live data to track the systems' evolution and suggest ways to optimize them [2]. A digital twin of a person's knowledge space would leverage the wealth of information available online to accurately model their learning journey, create a digital archive of their ideas, and streamline the tedious process of finding the most relevant learning resources online.

A Motivating Use Case

Imagine you are a college student studying neuroscience who wants to learn electronic circuit design for a personal project. To find a starting point, you search for some kind of tutorial on Google. Immediately, the results overwhelm you – Google returns dozens of tutorials of unknown quality, each assuming a different level of background experience. Given your limited free time, you reconsider venturing outside your comfort zone, but luckily, technology is on your side.

Thanks to a digital twin of your knowledge space, your computer knows your academic background and learning style. It knows about your knowledge of differential equations, experience modeling neural feedback loops, and freshman-year physics classes (which have been untouched for years). Using this information, the digital twin directs you to the middle of an MIT MOOC about analyzing electronic circuits. It skips the calculus review and goes straight to the part about device physics, allowing you to dive into the content without wasting any time.

After you report back to the digital twin, it refers you to a DIY-soil-moisture-sensor project to help you apply and consolidate the skills you just learned. It does this because it can infer the connection between neural feedback loops and electronic sensors. Before long, you're designing original circuits and sharing project ideas online for others to learn from. Even months later, when you need a quick refresher, you know exactly where to go because the digital twin remembers which resources worked best to help you learn.

This scenario may sound like something ripped out of science fiction, but digital twin technology could realistically reshape how we learn new skills via the internet. How would this kind of technology work?

Mechanisms of the Technology

Digital twins have seen successful applications in industries ranging from urban planning to solar energy [2,3]. Although they have historically only been used to simulate physical phenomena, their underlying techniques can represent any data-driven system, from computer networks to a person's knowledge space [4,10].

When a user first uses the digital twin, they would need to calibrate the model by providing it with various forms of data. Exactly which data will be most effective here must be formulated as this technology develops; some potential examples include their browser history, educational history like their college transcript, and some minimal demographic data like age. They could also rate their proficiency in a few skills to make the model even more accurate. This initial calibration would give users an instant payoff when they start using the digital twin, as the model would be able to provide personalized recommendations from the very beginning. This instant payoff could encourage them to keep using the technology, making it well worth the extra work and data required to implement it.

The digital twin could then crawl through the user's browser history and use natural language processing (NLP) techniques to identify keywords and concepts in each recently-visited website. Using those keywords, it could approximate the user's knowledge base by building a graph of relevant websites where related ones are linked; such a graph would also be an easily-searchable digital archive. The model could then assign each website a value indicating the user's proficiency in the concepts presented there, which would decay over time until the user reviews those concepts. These proficiency values could initially be derived from data like the user's educational history and timestamps in their browsing history, which would unify those data sources.

Using the current model of the user's knowledge space, the digital twin could recommend personalized next steps in their learning journey by querying a centralized database storing how relevant each resource was to other users. It could also act as an editorial filter by favoring content from reputable sources like large universities. When the user accesses these websites, the digital twin could use the same NLP techniques to add them to the graph and update the corresponding proficiency values.

Potential Benefits

The main benefit of digital twin technology is its potential to make online learning more accessible by streamlining the tedious process of finding the right learning resources on the internet. Many online learning innovations like Scratch [5] and connectivist MOOCs (a MOOC where students co-create knowledge through social media) [6] only work for the few people with the technical prowess and free time to navigate through all the loosely-connected content online. This high barrier to entry is one of the reasons why walled-garden platforms like Facebook and Twitter became more popular than blogs on the open web [1]; it's easier to consume and share content aggregated in one place, and doing so often triggers the brain's reward center and encourages further use.

When Scratch and connectivist MOOCs work, they work extraordinarily well – think of all the success stories about Scratch kindling a kid's passion for computer science and game design. But these success stories are few and far between, and for most people, online learning is an inefficient and frustrating experience. With many people turning to online learning as an affordable alternative to traditional school-based learning, it is becoming increasingly important to make it as accessible as possible. Lack of accessibility is one of the driving forces behind MOOCs' low completion rates – although their target audience is people without resources to access school-based learning, the overwhelming majority of people who complete MOOCs are already-educated, affluent learners [1]. Digital twin technology has the potential to create this much-needed accessibility by lowering the barrier to entry for these platforms. It would achieve this by suggesting personalized learning resources to users, which saves them the time and effort they would have otherwise spent searching for them. This automation would give users more time to engage with the content, making online learning more efficient and effective.

In addition to making online learning more accessible, digital twin technology would encourage participation in online learning communities. By creating a more personalized learning experience, users would feel more motivated to engage with the content and share their progress with others. The digital twin could even prompt users directly to share their ideas with others or recommend communities based on the user's interests and demographics. Sharing ideas fosters a sense of community and makes online learning a more social experience, which may further increase motivation and engagement.

Another benefit of digital twin technology is its ability to create a digital archive of a person's knowledge space. In today's fast-paced world, we are bombarded with so much information that it's impossible to retain all of it. With a digital archive ensuring easy access to the information they need, users could revisit and reinforce skills over time. This feature would be especially useful for people in fields requiring constant learning and skill acquisition, such as medicine or engineering.

Although a few knowledge management systems exist [7,8], none of them model skill retention rates. This lack of modeling becomes problematic when skills naturally atrophy over time when not reviewed and reinforced, which is especially noticeable for learning over several months or years. A 2012 study at MIT gave some fourth-year undergraduates an exam similar to the ones they took in their first-year introductory mechanics class; on average, they scored more than 50% lower than their first-year counterparts [9]. The digital twin's ability to model the evolution of the user's knowledge space would be a straightforward solution to this problem. Accurate reporting of the user's proficiencies in different skills would allow the digital twin to provide a more realistic report of the user's learning journey. This report could help users set and meet more realistic goals, as they would no longer over/underestimate their skills or the time it would take to reach their goals.

Risks and Mitigations

Since the digital twin would store so much personal information about the users, how could we make the system less invasive and protect the users' privacy? For a start, we should encrypt all stored data. Although simple encryption seems rudimentary, it is surprisingly effective at protecting data – over seven million data records get compromised daily because they are unencrypted [13]. To further decrease the risk of a breach, we should also minimize the amount of data stored at any given time; for example, the system could aggregate or delete old data that is no longer relevant. Aggregation is a good strategy for data protection, as it decreases the volume of data and hides personally-identifiable information while not compromising too much on the data's descriptiveness. The 2020 US census implemented differential privacy, an augmented form of aggregation, for this exact reason [14]. Finally, we should try to collect only the bare minimum demographic data necessary to make good-enough recommendations when the user first uses the digital twin. This approach would make the digital twin feel less invasive to the users, encouraging more widespread adoption. However, this approach would also be at odds with our incentives to properly calibrate the model, as collecting more data improves predictive models. As such, we must strike a healthy balance between these two goals, although we do not yet know what this balance should be.

As with most content recommendation systems, the issues of censorship and misinformation also arise. Who would build this technology, and what if they have a vested interest in excluding some sources of information from being recommended? What if a third party pays to have information (or even propaganda) made by them actively recommended? Social media platforms like Facebook have acted maliciously like that before, so it is a real risk. One way to mitigate it is to increase transparency by open-sourcing the technology's code. Doing so would hold the people building the technology more accountable for changes they make to the system, as any changes would automatically be available for public scrutiny.

On the other hand, some mild form of preferential recommendations may be necessary to curb the spread of misinformation through the digital twin. Given the vastness of the internet, it's unsurprising that it contains a lot of misinformation, spread intentionally (like fake news) and unintentionally (like a misstated fact or figure). Consequently, a content recommendation system could spread misinformation without malicious code powering it. One way to mitigate this risk is to preferentially recommend content from "trustworthy" sources like MOOCs developed by well-known universities like MIT, peer-reviewed journals, and official tutorials/manuals for technologies when they are available. These websites are far more likely than other arbitrary websites to contain the truth and teach effectively, so the digital twin would act as an editorial filter by implementing this approach.

Despite this approach's effectiveness, it still has a few limitations. Trustworthiness is a nebulous metric, so excluding some websites from that list for no apparent reason may feel unfair. We could implement something akin to Twitter's old "blue checkmark" system [11], where a website would become a trustworthy source after receiving sufficient traffic from the digital twin's users. However, this feature may be exploitable by bots to promote spam websites. This approach could also limit access to hidden gems that explain concepts better, as too much preference for large, established sources of information would give them a monopoly and hinder innovation. The most straightforward way to address this limitation would be to make the preference for trustworthy sources very slight. Despite these two limitations, the benefits of having an editorial filter in the system far outweigh the downsides.

Skeptics may point to the numerous past failures in educational technology and claim that this technology would never see widespread adoption. This skepticism is not unfounded – indeed, many adaptive tutor platforms like Knewton [15] and Khan Academy initially promised to revolutionize education through personalized learning, yet now are mainly used as modest supplements to traditional classroom instruction [1]. Nevertheless, I believe the second mouse gets the cheese, as we could learn from these platforms' mistakes instead of letting them discourage us.

Firstly, this technology does not aim to replace existing educational infrastructure; instead, it would augment them. It would be designed to fit naturally into its users' workflows, just like how a calculator fits naturally inside a math classroom. Aside from making the digital twin feel less out of place, it would also not antagonize educators as adaptive tutor platforms had done before. Additionally, there is little setup overhead when using the technology. Most processes would be automated from the start, so users would not have to worry about using the digital twin correctly. As a result, the technology would have low floors and high ceilings. It would be easy to get started and get a few suggestions to learn something quickly, but it would be possible to use the digital twin for more long-term, larger-scope learning too.

Next Steps and Conclusion

Many of the techniques I have described (NLP, digital twins, and learning theory) are still budding fields that would need more attention and research to blossom; as such, we should invest more in advancing those fields. While researchers create the technology, we should concurrently revise student privacy laws like FERPA and COPPA – laws established decades before our current digital age – to prepare us better for privacy-related issues arising from digital twin technology and other data-driven educational technologies [17].

Digital twin technology shows great promise for enhancing learning using the internet. It would do this by making existing online learning platforms more accessible and accurately tracking a user's progress to help them stay on track. It could thus promote lifelong learning by providing its users with tools to create a streamlined and personalized learning experience. Despite these promising benefits, censorship/misinformation, privacy, and the historical underperformance of educational technology are some potential pitfalls we must address. Although they are complex issues, they are still tractable, and I have described several possible approaches for addressing them.

Nevertheless, digital twin technology has the power to revolutionize online learning in ways we never thought possible. This cross-disciplinary project would bring together the brightest minds from across the world of computing, exemplifying the tenets of this technology. Even in the unlikely event that it fails, the ideas behind it could power exciting new digital twin applications and reshape how we think about privacy and misinformation. So what are we waiting for? Let us collaborate and leverage the power of digital twins to shape a brighter future for education!

Acknowledgments

Thank you to [REDACTED 1] for providing valuable feedback as I worked on this essay, [REDACTED 2] for introducing me to the concept of digital twins, and [REDACTED 3], [REDACTED 4], and [REDACTED 5] for letting me bounce initial ideas off of them.

References

[1] Reich, Justin. *Failure to Disrupt.* Harvard University Press, 2020.

[2] IBM. *What is a digital twin?.* https://www.ibm.com/topics/what-is-a-digital-twin. 2021.

[3] Unity Game Engine. *Digital Twin Applications and Use Cases.* https://unity.com/solutions/digital-twin-applications-and-use-cases. 2021.

[4] Almasan, Paul, et al. *Digital twin network: Opportunities and challenges*. arXiv preprint arXiv:2201.01144. 2022.

[5] Scratch. https://scratch.mit.edu/. 2023.

[6] Downes, Stephen. *Connectivism and Connective Knowledge*. https://cck11.mooc.ca/. 2011.

[7] Bricman, Paul. *Conceptarium*. https://paulbricman.com/thoughtware/conceptarium. 2022.

[8] Girit, Uzay. *Archivy*. https://archivy.github.io/. 2023.

[9] Pawl, Andrew, et al. *What do seniors remember from freshman physics?*. Physical Review Special Topics-Physics Education Research 8.2. 2012.

[10] Buitrago, Mauricio, and Andrés Chiappe. *Representation of knowledge in digital educational environments: A systematic review of literature*. Australasian Journal of Educational Technology 35.4. 2019.

[11] Kanalley, Craig. *Why Twitter Verifies Users: The History Behind the Blue Checkmark*. Huffington Post. https://www.huffpost.com/entry/twitter-verified-accounts_b_2863282. 2013.

[12] Gupta, Shashank. *Sentiment Analysis: Concept, Analysis, and Applications*. Towards Data Science. https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17. 2018.

[13] Sobers, Rob. *The World in Data Breaches*. Varonis. https://www.varonis.com/blog/the-world-in-data-breaches. 2020.

[14] United States Census. *Understanding Differential Privacy*. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html. 2020.

[15] Warner, John. *Knewton Is Gone. The Larger Threat Remains*. Inside Higher Ed. https://www.insidehighered.com/blogs/just-visiting/knewton-gone-larger-threat-remains. 2019.

[16] Lederman, Doug. *Why MOOCs Didn't Work, in 3 Data Points*. Inside Higher Ed. https://www.insidehighered.com/digital-learning/article/2019/01/16/study-offers-data-show-moocs-didnt-achieve-their-goals. 2019.

[17] Creel, Kathleen, and Tara Dixit. *Privacy and Paternalism: The Ethics of Student Data Collection*. MIT Case Studies in Social and Ethical Responsibilities of Computing. 2022. https://doi.org/10.21428/2c646de5.b725319a.