Advances in Engineering Education

2025: VOLUME 13 ISSUE 2

Opinion: Preparing Engineers for the Data-Driven World: The Case for Contextualized Data Science Engineering Education

WESLEY F. REINHART
Pennsylvania State University

REED WILLIAMS
Siemens Corporation, Technology

RYAN SOLNOSKY
R. ALLEN KIMEL
REBECCA NAPOLITANO
Pennsylvania State University

ABSTRACT

Data science has become an increasingly popular topic among engineering students and practitioners as high-profile engineering applications of machine learning and artificial intelligence continue to make headlines. Companies in engineering domains are placing a growing emphasis on hiring engineers who can extract insights and create value from the large amounts of data generated, rather than having distinct roles for designers and data scientists working separately. However, the question of who is responsible for teaching these emerging skills remains an outstanding one: Should they be taught in-house by engineering faculty? Should students take generalized courses outside their program, such as those housed in computer science departments? Or should students be responsible for self-teaching outside their formal training? In this commentary, we argue that undergraduate engineering curricula should include contextualized training in programming and data science concepts. We describe our recent efforts to build a "Data Science Corps" in collaboration with dozens of other institutions throughout the United States and provide specific examples of our recent approach to contextualized data science education in two engineering disciplines, materials engineering and architectural engineering. As engineering programs strive to incorporate new curricula, contextualization becomes vital in bridging the gap between abstract data science concepts and their application in specific engineering disciplines. The collaboration between academia and industry will be crucial for developing effective educational programs that equip future engineers with the necessary skills to thrive in a fast-moving, data-driven world. By embracing this collaboration and integrating data science education within engineering disciplines, we can pave the way for a future where engineers are proficient in both domain-specific knowledge and data science competencies, leading to transformative advancements in the field.

Key words: Engineering curriculum, Data science, Contextualized learning

MOTIVATION

It is becoming cliché to say that data science continues to make headlines in both mainstream media and scientific venues, yet it remains true. Generative models especially have dominated news cycles in recent years (Chang and Kidman, 2023). For instance, the release of Stable Diffusion's source code in August 2022 provoked heated debate over copyright, ownership, and fair use of images (Rombach et al. 2022). Later, in November 2022, the public beta of OpenAI's new generative language model ChatGPT (OpenAI 2022) launched, sparking similar ethical controversies around language (Bishop 2023). While at the time of writing, there remains significant uncertainty about the ethical, legal, and practical implications of using these models for personal or professional projects, it appears very clear that the genie is out of the bottle, and these technologies have irreversibly changed the way engineers will work and learn (Georgieva et al. 2022; Aksoy et al. 2024).

While most of the news coverage of data science topics has been over provocative proofs of concept in the image and language domains (such as generating images of an astronaut riding a horse with Stable Diffusion or asking ChatGPT to impersonate a full-stack web developer), scientific and engineering applications of data science are already embedded into how the world learns (e.g., AI-powered research tools (Heidt 2023)) and builds (e.g., generative design tools in Computer Aided Design software (Rawat and Tiwari 2023)). Traditional phenomenological modeling strategies are slowly ceding ground to data-driven surrogate models that provide superior performance. Commercial and academic datasets are growing in volume and complexity as such models provide increasingly more real-world value. Computing hardware grows more powerful each year, providing more detailed and higher throughput calculations.

Next-generation scientists need to be comfortable with these data-driven paradigms to be competitive in the job market (Dumbill et al. 2013). At a minimum, this means understanding how to interact with such models through software interfaces, but in practice, it requires developing basic programming proficiency during their undergraduate training (National Academies of Sciences, Engineering, and Medicine 2018). For instance, scripted data manipulation and visualization have become increasingly important as access to published data and models continues to improve yearly in response to federal mandates (National Science and Technology Council 2022).

In engineering industries, new data science tools are improving not only the flagship use cases such as image classification and document summarization but also invigorating R&D in domain-specific problems that previously had no tractable solutions (e.g., CAD program synthesis, design of hierarchical structures (Li et al. 2023; Li et al. 2024). Engineering businesses increasingly seek engineers who can derive value from the vast streams of data they create rather than separating the designer and data scientist into two separate roles (Liston et al. 2022). Data science is increasingly leveraged to extract insights from historical company data, enhancing current and future products, and enabling exploration of vast design spaces. It enables the generation and evaluation of an immense number of potential design alternatives at a scale that would be intractable through manual human analysis alone (Hille, 2023). More and more, we ask, "Could a virtual advisor (Chanatry 2021) or assistant aid a human expert with this task?" and more and more, we answer, "Yes!"

The ability to design things better, faster, and more confidently using data-driven modeling is already changing from a differentiator in the market to a fundamental requirement for an engineering business. As companies try to staff positions that require these new skills while still expecting strong domain knowledge from traditional engineering disciplines, the lack of context for these methods within the existing academic infrastructure becomes clear. Companies face equal challenges in upskilling experienced engineers with data science skills and providing new computer science graduates with the domain knowledge and engineering expertise required to effectively apply these tools to real-world problems (Orit and Mike 2020). Even the jargon of data science (e.g., "latent spaces") is not properly contextualized for learners intending to use these tools in industries with well-established equivalents (e.g., "configuration spaces"). Addressing these gaps will require changes to classroom education in engineering undergraduate programs. This brings us to the key questions: Who will teach these emerging skills to future engineers, and how will they do it?

CHALLENGES

Challenges associated with courses housed in computer science departments: Almost every secondary education program has an opportunity for students to engage in an introductory computer science course. However, while computer science faculty may be better trained in data science topics, they cannot always provide examples that connect general topics to each student's discipline (Nagashima 2018). Contextualization challenges are well documented in education research and are reflected in the popular Bloom's Taxonomy, one of the most common hierarchical models used to discuss student skill levels and learning objectives (Krathwohl 2002). Bloom's Taxonomy shows that "Apply" and "Analyze" demand deeper familiarity with a subject than "Remember" and "Understand." However, it has been argued in the computer science education literature that context can motivate student learning more powerfully than abstract concepts (Cooper and Cunningham 2010). Additionally, programming concepts are learned in computer science courses but are not regularly reinforced and applied within the engineering curriculum (El-Zein et al. 2009; Dunne et al. 2005). In these cases, students often struggle to retain and utilize pertinent skills effectively in their home disciplines which can lead to negative perceptions of the topic (Cetin and Ogden 2015). If we are preparing our students to critically evaluate the problems of the 21st century workforce, which will certainly include synthesizing data science with foundational engineering knowledge, we need to reconsider if this type of approach is pedagogically defensible.

Challenges associated with self-study: As another avenue to acquiring in-demand data science skills, students sometimes engage in self-directed learning outside their formal coursework, which presents unique challenges for data science education. Thanks to the hard work of many academics in statistics, mathematics, computer science, and related fields, there is already an abundance of open courseware and other educational materials for data science topics available for free on the Internet (Adhikari and Jordan 2021). However, without the structured guidance of a formal curriculum, students must navigate a vast array of resources, potentially becoming overwhelmed or confused (Pears et al. 2007; Glantz et al. 2023). In addition, it can be difficult for students to comprehend how these concepts can be applied to their discipline without contextualization (Craig et al. 2017). For instance, one homework problem in the excellent "Data8" course curriculum made publicly available by UC Berkeley tasks students with evaluating professional basketball players' salaries using statistical modeling (Adhikari et al. 2022). While it is true that these tabular data could be easily exchanged for relevant data from an engineering domain, such as the mechanical properties of metal alloys or the structural health of infrastructure, making the leap to an engineering context is an extra challenge for students seeing the technical content for the first time. Furthermore, the challenges associated with self-learning can contribute to widening the gap between students with more resources and those with less, potentially exacerbating existing inequities in engineering education (Valero 2022; Rajan 2015).

Challenges associated with embedding data science into engineering courses: Engineering curricula are already packed with discipline-specific courses, leaving limited room for additional programming- and data-science-focused classes (Gerrard et al. 2017; Law 2022). Introducing new courses may require restructuring existing programs or sacrificing other vital topics. Additionally, programming and data science fields are constantly evolving, with new tools, libraries, and best practices emerging regularly. Keeping course content up-to-date and relevant can be a daunting task for engineering departments, especially if they lack dedicated resources for curriculum development (Aguilar-Santelises et al. 2014). Furthermore, engineering departments must balance teaching theoretical concepts and providing hands-on, practical experience. Finding the best mix of these can be challenging, as it requires a combination of classroom instruction, lab work, and real-world projects. This could burden engineering faculty who may be familiar with these topics but are not formally trained; the National Academies of Sciences, Engineering, and Medicine have described recruiting and retaining faculty to create and teach integrative introductory courses as a significant hurdle to contextualized data science education (National Academies of Sciences, Engineering, and Medicine, 2017; National Academies of Sciences, Engineering, and Medicine, 2018).

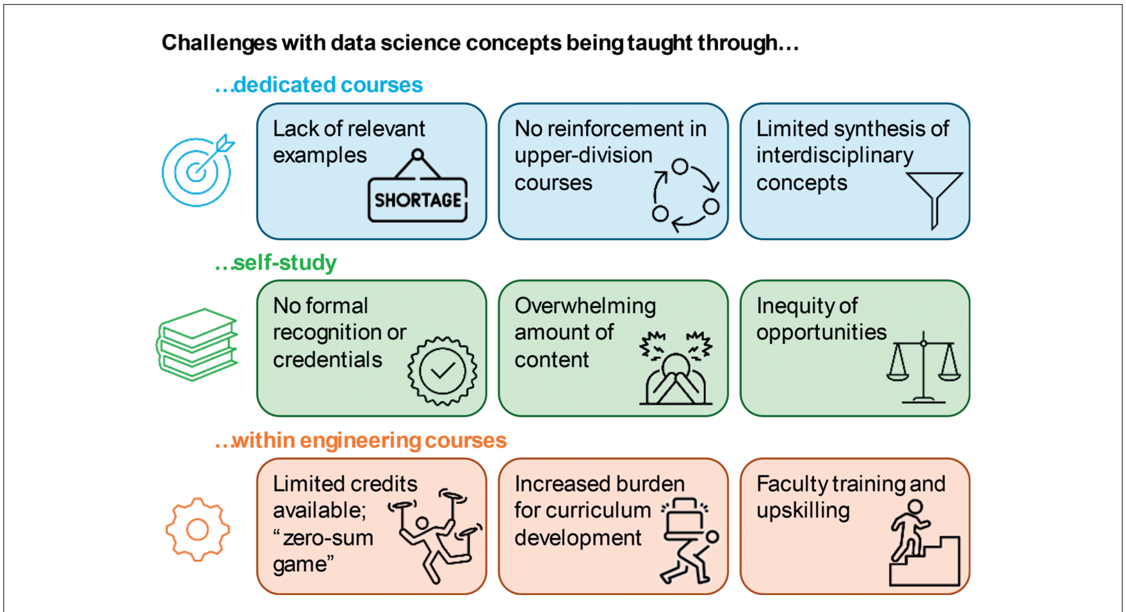


Figure 1. Schematic illustration of the challenges with data science concepts being taught in dedicated courses, through self-study, and in contextualized engineering courses.

Finally, some good news: research on contextualization in computing education has found that while there is a significant initial overhead to the development of contextualized curricula, the resulting products can often be shared and reused by other faculty teaching similar courses (such as in other departments or at other institutions) (Guzdial 2009). Considering this, we define a pragmatic solution: as data science is a rapidly evolving field, and we anticipate it will remain so for the foreseeable future, we advocate for an equally dynamic approach to its integration into engineering curricula. Fundamentals (such as programming, statistical analysis, and regression) should be established in a contextualized introductory course tailored to a specific engineering discipline; new tools and methods should be incorporated piecemeal throughout the existing engineering curriculum, allowing for rapid adaptation to the changing landscape. This approach provides students with a solid understanding of data science principles within the context of their chosen engineering field. It also prepares them for the lifelong learning process they will undoubtedly face in the workforce (Shen et al. 2020). By encountering new tools and techniques throughout their engineering education, students develop the ability to adapt and learn on the job, a skill that will prove invaluable in an industry where the pace of change is only accelerating (Flemming and Panizzon 2010). In the following section, we will give some examples of how this strategy was implemented in two engineering majors at the undergraduate level. Subsequently, we will share our outlook, lessons learned, and suggestions for future implementations.

IMPLEMENTATION

The challenges and our chosen solution outlined above highlight the importance of an ongoing National Science Foundation initiative, the Data Science Corps (DSC), which seeks to equip next-generation scientists and engineers with data management and analytics competency. This training program is part of the "Harnessing the Data Revolution" umbrella, one of the NSF 10 Big Ideas announced in 2016, which focuses on fundamental data science research, education, and cyberinfrastructure. Our DSC project is focused on building modular, transferrable curricula for engineering programs to create longstanding institutional collaborations within this dynamic educational landscape.

The core of our educational initiative comes through traditional classroom experiences in two specific disciplines: Architectural Engineering (AE) and Materials Science and Engineering (MATSE). "AE 240: Introduction to Programming and Data Science for Architectural Engineers" introduces students to essential tools for manipulating and modeling architectural engineering data, including introductory programming concepts, statistical analysis, regression, and data visualization. By teaching these concepts in-house, special attention can be given to peculiarities of architectural engineering data, including digital representations for building information modeling. Likewise, "MATSE 219: Introduction to Materials Informatics" introduces data science concepts through a "materials informatics" lens for second-year students. Through instruction in these concepts, including

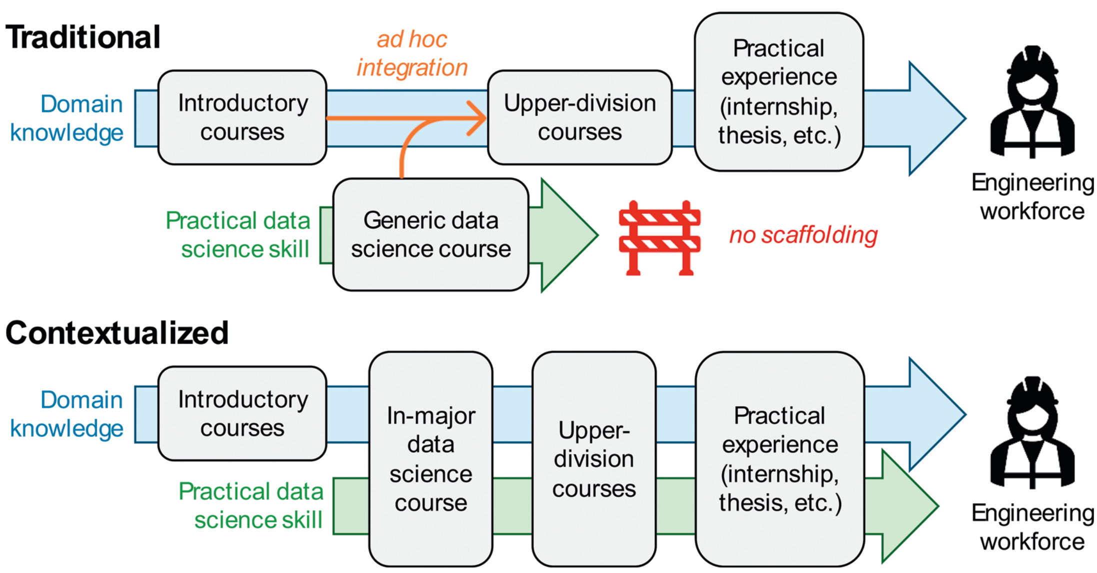


Figure 2. Schematic illustration of the differences between traditional and contextualized learning paths for data science in engineering.

digital representations of the periodic table and crystal structures, special attention can be given to unique features of materials data. Teaching these introductory courses within the engineering disciplines allows students to more easily apply the principles to their other coursework (and future career) than if they were taught through a computer science lens.

Planning for contextualized data science content within the AE curriculum began in 2020. A committee was formed at the department level to identify potential ways these topics could be integrated into the curriculum. Based on discussions with students, alumni, and faculty, the committee advocated for the creation of a new standalone course. However, with the existing curriculum for the AE major already at 160 credits, changes had to be identified to create room for such a course to be added. A course on electrical circuits was removed as a requirement, as it was deemed to be sufficiently covered in the rest of the curriculum, and AE 240 took its place in the curriculum. Following this, the course ran for a provisional year, and a syllabus was submitted to the Faculty Senate for formal review. Once approved, AE 240 became required for all AE students in their second year.

A similar strategy was adopted for the MATSE installation of this course. However, there was already a required computer programming course in the curriculum, so a simple replacement could be performed. Unfortunately, despite widespread departmental buy-in, formal course approval did not go as smoothly as in the case of AE. Initially offered as "MATSE 297: Introduction to Data Science for Materials Scientists," the MATSE equivalent of AE 240 received excellent student feedback but was not approved by the Faculty Senate, whose subcommittee cited an overabundance of similar course offerings focused on contextualized data science being proposed in recent years. Subsequently, we revised the course title to MATSE 219: Materials Informatics and updated about 15% of the proposed topics to be more specific to materials science. The course was subsequently approved to begin running in Fall 2023. We expect this resistance to perceived duplication or overlap of data science courses may be a common counterpoint to our proposed strategy that will need to be addressed within the context of each institution.

An important practical question for both course offerings was which programming language to use to teach these topics. While Matlab has been a common choice for engineers in the past and R for mathematicians, we consider Python the de facto language for data science. For instance, the popular deep learning frameworks TensorFlow, PyTorch, JAX, and Keras are all Python-based. Python is also free and open-source, has a huge open-source development community, and even has commercial-friendly licensing, which makes it a viable tool for students planning to join the private sector after graduation. We teach the course using interactive Jupyter notebooks hosted on the Google Colaboratory platform, which hosts a free, pre-configured scientific Python environment on the Google Cloud. This means students can launch a web browser and start running Python code within minutes of the start of the course without needing any troubleshooting from the instructors to get a working installation. It also makes it easy to ensure that everyone (students and instructors) works in the same environment. While Python and Jupyter notebooks are not the only sensible choices, we advocate for similarly pragmatic solutions that reduce implementation barriers for instructors wherever possible, with the explicit understanding that the landscape will continue to evolve and further updates will likely be needed every year or two.

Taking a step back from choices of interpreters and development environments, we must also consider the incredible range of software interfaces that ship with today's tools. For instance, Large Language Models typically have no-code interfaces (such as web applications) that allow for simple interaction patterns such as interactive chat. However, these interaction patterns are insufficient for scientific use cases such as entity extraction from documents or summarization of scientific literature, which may require thousands of queries. Furthermore, even for simple tasks like few-shot problem-solving, recent work has shown that "prompt programming" is required to achieve the best possible performance, with some seemingly random, automatically discovered prompts out-performing handcrafted ones (Korzynski et al. 2023). This underscores the need for familiarity with programming languages rather than relying exclusively on graphical software interfaces in order for students to keep their skills up to date (Kuhail et al. 2024). Beyond the concepts required for scripting queries to pre-trained models, students must understand data modality, domains, and performance metrics to perform sanity checks when applying models to their data (Holz et al. 2006). Ethical use of today's sophisticated generative models – or even off-the-shelf regression models, in some cases – requires an accurate understanding of their working principles, which is challenging to acquire without practical experience.

Having established these introductory courses in each domain, a problem remained of how to reinforce this knowledge as students progressed throughout the curriculum. Initial efforts have been made in "AE 430: Indeterminate Structural Analysis" and "MATSE 419: Computational Materials Science and Engineering." The intention was that these classes would continue to reinforce programming concepts taught in the introductory courses. However, due to variable teaching loads and faculty availability, the AE instructor has not been able to offer Python-based AE 430 again. The MATSE instructor has recently offered a Python-based MATSE 419, which was historically taught in Matlab based on the previous computer science prerequisite. Incidentally, even though some students in MATSE 419 were grandfathered in from the Matlab course and given the option to complete the assignments in Matlab, every student opted to complete the assignments in Python, mainly due to the simpler interface in Google Colab compared to using the proprietary Matlab software on the university's remote desktop service. Furthermore, "MATSE 413: Solid State Materials" now contains a Python-based computation module using Google Colab, the first instance in our program of computational modules being used by an instructor not already familiar with Python.

Our difficulty in maintaining consistent course offerings highlights two critical issues. The first is that course offerings in engineering departments are frequently subject to changes in staffing. Specialized faculty who know the material are not consistently staffed in upper-level courses taught to the whole major, leading to gaps in reinforcing the skills from introductory courses at later stages in the curriculum. This leads to a second issue, which is the investment required to upskill faculty unfamiliar with data science to redevelop existing curricula. In AE, additional work is planned to assess if plug-and-play modules can be inserted throughout the AE curriculum. This leaves open the question of how time and resources will be allocated to this objective and how the outcomes will be valued by institutional leadership. Once more, these are non-technical challenges that must be resolved within the landscape of each institution.

Despite these challenges, we have successfully implemented contextualized data science education in two engineering curricula. We have launched introductory data science courses tailored to each discipline, namely AE 240 and MATSE 219, both required courses in their respective majors. These courses have received positive student feedback (i.e., both 2022 offerings had median scores of 7 on the 7-point Student Rating of Teaching Effectiveness), indicating their effectiveness and relevance for the students. A sample of open-ended student feedback from each course is provided below:

> "Connecting real life examples related to AE with coding really helped with understanding how to implement and solve the problems at a much faster pace." – AE Spring 2022 student

> "I think this class let us know how [we can] apply data science into material science and triggered my interest in learning this really helpful tool." – MATSE Fall 2022 student

The outcomes of these courses underscore the success of our chosen approach to practically integrate data science into our engineering curricula. However, they also expose the numerous challenges we have faced in the past, and some that still lie ahead. This emphasizes the importance of providing faculty training and institutional support to realize contextualized data science education within engineering programs. Facilitating faculty training is a primary objective of our Data Science Corps project, which hosts sample lessons, entire courses, and instructor resources online at https://sites.psu.edu/datasciencecorps.

OUTLOOK

Looking ahead, it is evident that data science education will quickly continue to gain importance across all engineering disciplines. Machine learning and artificial intelligence advancements have already transformed various industries, such as finance and healthcare, and a revolution in engineering is also underway. As engineering professionals increasingly rely on data-driven decision-making processes, the need for engineers who are well-versed in data science concepts and tools will only grow. However, incorporating new curricula to keep up with the rapid pace of innovation in data science poses a challenge for engineering programs. Thus, we return to the key question of who will teach these emerging skills to future engineers, and how they will do it. The existing course load for engineering students is often at its maximum capacity (Gerrard et al. 2017), and accreditation standards tend to change slowly (Desha et al. 2009). Considering this reality, it is crucial to find ways to integrate important data concepts and skills into domain science courses without increasing the overall credit requirements.

Our experience implementing contextualized data science education in engineering curricula has led to a few key observations. First, we have seen that standalone introductory data science courses contextualized within specific engineering disciplines are essential to provide students with a solid foundation to build on throughout the curriculum. This approach enhances student engagement and understanding and prepares them for what they will face in the rapidly evolving technological landscape. Embedding data science education within specific engineering disciplines, such as materials informatics, can break the “zero-sum game” of required course credits. This approach ensures that students receive contextualized training, allowing them to understand how data science can be applied to their field of study. However, we have also seen that maintaining consistent course offerings and reinforcing data science concepts throughout the curriculum can be challenging due to changes in staffing and faculty availability. This highlights the need for a more sustainable approach to curriculum development and the importance of investing in faculty upskilling to ensure the long-term success of data science integration in engineering education.

Finally, we observe that the successful integration of data science education in engineering curricula extends beyond technical challenges. It requires institutional support, resource allocation, and a clear vision for how these efforts will be valued and prioritized within the broader educational framework. Addressing these non-technical challenges is crucial for ensuring the long-term sustainability and impact of data science education in engineering. We anticipate that collaboration with industry partners will be vital to align curricula with real-world needs and provide students with relevant, applied learning experiences. As we continue to refine and expand our approach, we remain committed to sharing our experiences and encourage others to do the same to promote effective, contextualized data science education for the next generation of engineers.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant IIS-2123343. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Adhikari, Ani, John DeNero, and David Wagner. 2022. "Computational and Inferential Thinking: The Foundations of Data Science." Inferentialthinking.com. 2022. https://inferentialthinking.com.

Adhikari, Ani, and Michael I. Jordan. 2021. "Interleaving Computational and Inferential Thinking in an Undergraduate Data Science Curriculum." *Harvard Data Science Review* 3 (2). https://doi.org/10.1162/99608f92.cb0fa8d2.

Aguilar-Santelises, Leonor, María Teresa Corona Ortega, Margarita Cruz-Millán, Antonia Guillermina Rojas-Fernández, Miguel Aguilar-Santelises, and Araceli García del Valle. 2014. "Keeping Teachers Up-To-Date Is Essential to Reach and Maintain a High Quality University Education." *Multidisciplinary Journal for Education, Social and Technological Sciences* 2 (1): 1. https://doi.org/10.4995/muse.2015.2204.

Aksoy, Bekir, Osamah Khaled, Özge Ekrem, Nergiz Aydin, and Mehmet Yücel. 2024. "Artificial Intelligence-Based Engineering Applications a Comprehensive Review of Application Areas, Impacts and Challenges." In *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*, 32–47. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-56310-2_3.

Bishop, Lea. 2023. "A Computer Wrote This Paper: What ChatGPT Means for Education, Research, and Writing." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4338981.

Cetin, Ibrahim, and M. Yasar Ozden. 2015. "Development of Computer Programming Attitude Scale for University Students." *Computer Applications in Engineering Education* 23 (5): 667–72. https://doi.org/10.1002/cae.21639.

Chanatry, William. 2021. "NX Assemblies: Updates to Save Time and Increase Performance - NX Design." Siemens Digital Industries Software Blogs. Siemens Digital Industries Software. July 9, 2021. https://blogs.sw.siemens.com/nx-design/nx-assemblies-updates-to-save-time-and-increase-performance.

Chang, Chew-Hung, and Gillian Kidman. 2023. "The Rise of Generative Artificial Intelligence (AI) Language Models - Challenges and Opportunities for Geographical and Environmental Education." *International Research in Geographical and Environmental Education* 32 (2): 1–5. https://doi.org/10.1080/10382046.2023.2194036.

Cooper, Steve, and Steve Cunningham. 2010. "Teaching Computer Science in Context." *ACM Inroads* 1 (1): 5. https://doi.org/10.1145/1721933.1721934.

Craig, Michelle, Jacqueline Smith, and Andrew Petersen. 2017. "Familiar Contexts and the Difficulty of Programming Problems." In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*, 123–27. https://doi.org/10.1145/3141880.3141898.

Desha, Cheryl J., Karlson Hargroves, and Michael H. Smith. 2009. "Addressing the Time Lag Dilemma in Curriculum Renewal towards Engineering Education for Sustainable Development." *International Journal of Sustainability in Higher Education* 10 (2): 184–99. https://doi.org/10.1108/14676370910949356.

Dumbill, Edd, Elizabeth D. Liddy, Jeffrey Stanton, Kate Mueller, and Shelly Farnham. 2013. "Educating the next Generation of Data Scientists." *Big Data* 1 (1): 21–27. https://doi.org/10.1089/big.2013.1510.

Dunne, Bruce, A.J Blauch, and Andrew Sterian. 2020. "The Case for Computer Programming Instruction for All Engineering Disciplines." In *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition*. https://doi.org/10.18260/1-2--15520.

El-Zein, Abbas, Tim Langrish, and Nigel Balaam. 2009. "Blended Teaching and Learning of Computer Programming Skills in Engineering Curricula." *Advances in Engineering Education* 1 (3).

Fleming, Keith, and Debra Panizzon. 2010. "Facilitating Students' Ownership of Learning in Science by Developing Lifelong Learning Skills." *Teaching Science: The Journal of the Australian Science Teachers Association* 56 (3): 27–32.

Georgieva, Ilina, Claudio Lazo, Tjerk Timan, and Anne Fleur van Veenstra. 2022. "From AI Ethics Principles to Data Science Practice: A Reflection and a Gap Analysis Based on Recent Frameworks and Practical Experience." *AI and Ethics* 2 (4): 697–711. https://doi.org/10.1007/s43681-021-00127-3.

Gerrard, Darlee, Kirstin Newfield, Narges Balouchestani Asli, and Chirag Variawa. 2017. "Are Students Overworked? Understanding the Workload Expectations and Realities of First-Year Engineering." *2017 ASEE Annual Conference & Exposition Proceedings*, June. https://doi.org/10.18260/1-2--27612.

Glantz, Mary, Jennifer Johnson, Marilyn Macy, Juan Nunez, Rachel Saidi, and Camilo Velez Ramirez. 2023. "Students' Experience and Perspective of a Data Science Program in a Two-Year College." *Journal of Statistics and Data Science Education* 31 (3): 248–57. https://doi.org/10.1080/26939169.2023.2208185.

Guzdial, Mark. 2009. "Teaching Computing to Everyone." *Communications of the ACM* 52 (5): 31–33. https://doi.org/10.1145/1506409.1506420.

Hazzan, Orit, and Koby Mike. 2020. "Ten Challenges of Data Science Education – Communications of the ACM." Communications of the ACM. Association for Computing Machinery. July 16, 2020. https://cacm.acm.org/blogcacm/ten-challenges-of-data-science-education/.

Heidt, Amanda. 2023. "Artificial-Intelligence Search Engines Wrangle Academic Literature." *Nature* 620 (7973): 456–57. https://doi.org/10.1038/d41586-023-01907-z.

Hille, Karl. 2023. "NASA Turns to AI to Design Mission Hardware." NASA.gov. National Aeronautics and Space Administration. February 9, 2023. https://www.nasa.gov/feature/goddard/2023/nasa-turns-to-ai-to-design-mission-hardware.

Holz, Hilary J., Anne Applin, Bruria Haberman, Donald Joyce, Helen Purchase, and Catherine Reed. 2006. "Research Methods in Computing: What Are They, and How Should We Teach Them?" *ACM SIGCSE Bulletin* 38 (4): 96. https://doi.org/10.1145/1189136.1189180.

Korzyński, Paweł, Grzegorz Mazurek, Pamela Krzypkowska, and Artur Kurasiński. 2023. "Artificial Intelligence Prompt Engineering as a New Digital Competence: Analysis of Generative AI Technologies such as ChatGPT." *Entrepreneurial Business and Economics Review* 11 (3): 25–37. https://doi.org/10.15678/eber.2023.110302.

Krathwohl, David R. 2002. "A Revision of Bloom's Taxonomy: An Overview." *Theory into Practice* 41 (4): 212–18. https://doi.org/10.1207/s15430421tip4104_2.

Kuhail, Mohammad Amin, Sujith Samuel Mathew, Ashraf Khalil, Jose Berengueres, and Syed Jawad Hussain Shah. 2024. "'Will I Be Replaced?' Assessing ChatGPT's Effect on Software Development and Programmer Perceptions of AI Tools." *Science of Computer Programming* 235 (March): 103111. https://doi.org/10.1016/j.scico.2024.103111.

Law, Mei Yuan. 2022. "A Review of Curriculum Change and Innovation for Higher Education." *Journal of Education and Training Studies* 10 (2): 16. https://doi.org/10.11114/jets.v10i2.5448.

Li, Beichen, Bolei Deng, Wan Shou, Tae-Hyun Oh, Yuanming Hu, Yiyue Luo, Liang Shi, and Wojciech Matusik. 2024. "Computational Discovery of Microstructured Composites with Optimal Stiffness-Toughness Trade-Offs." *Science Advances* 10 (5). https://doi.org/10.1126/sciadv.adk4284.

Li, Pu, Jianwei Guo, Xiaopeng Zhang, and Dong-Ming Yan. 2023. "SECAD-Net: Self-Supervised CAD Reconstruction by Learning Sketch-Extrude Operations." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr52729.2023.01613.

Liston, Maeve, Anne M. Morrin, Trevor Furlong, and Leona Griffin. 2022. "Integrating Data Science and the Internet of Things into Science, Technology, Engineering, Arts, and Mathematics Education through the Use of New and Emerging Technologies." *Frontiers in Education* 7 (August). https://doi.org/10.3389/feduc.2022.757866.

Nagashima, Tomohiro. 2018. "Contextualized Instruction in Data Science and Its Effect on Transfer of Learning." In *EC-TEL (Doctoral Consortium)*. https://ceur-ws.org/Vol-2294/DCECTEL2018_paper_25.pdf.

National Academies of Sciences, Engineering, and Medicine. 2017. "Envisioning the Data Science Discipline." *National Academies Press EBooks*. National Academies Press. https://doi.org/10.17226/24886.

National Academies of Sciences, Engineering, and Medicine. 2018. "Data Science for Undergraduates: Opportunities and Options." *National Academies Press EBooks*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/25104.

National Science and Technology Council. 2022. "Desirable Characteristics of Data Repositories for Federally Funded Research." *The White House*. Executive Office of the President of the United States. https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf.

OpenAI. 2022. "Introducing ChatGPT." Openai.com. November 30, 2022. https://openai.com/index/chatgpt/.

Pears, Arnold, Stephen Seidman, Lauri Malmi, Linda Mannila, Elizabeth Adams, Jens Bennedsen, Marie Devlin, and James Paterson. 2007. "A Survey of Literature on the Teaching of Introductory Programming." *ACM SIGCSE Bulletin* 39 (4): 204–23. https://doi.org/10.1145/1345375.1345441.

Rajan, Rajes K.M. 2015. "Exploring the Impacts of Self-Directed Learning Approach among Polytechnic Mechanical Engineering Students." *International Journal of Academic Research in Education and Review* 3 (7): 204–14. https://www.semanticscholar.org/paper/Exploring-the-Impacts-of-Self-Directed-Learning-Rajan/6c9e15886a141e4fedab27925619e2337a8d1514.

Rawat, Amit Singh, and Gaurav Tiwari. 2023. "Modern Generative Design Tools: Siemens NX's Algorithmic Feature and Rhinoceros 3D's Grasshopper." *Lecture Notes in Mechanical Engineering*, January, 275–84. https://doi.org/10.1007/978-981-99-3033-3_24.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. "High-Resolution Image Synthesis with Latent Diffusion Models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–95. The Computer Vision Foundation. https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.

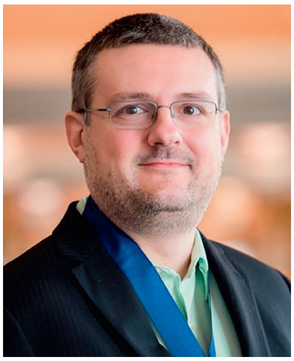
Shen, Ruiqi, Joseph Chiou, and Michael J. Lee. 2020. "Becoming Lifelong Learners: CS Learners' Autonomy." *Journal of Computing Sciences in Colleges* 35 (8). https://doi.org/10.5555/3417639.3417672.

Valero, Miguel. 2022. "Challenges, Difficulties and Barriers for Engineering Higher Education." *Journal of Technology and Science Education* 12 (3): 551. https://doi.org/10.3926/jotse.1696.

AUTHORS

Wesley Reinhart is an Assistant Professor of Materials Science and Engineering and Institute for Computational and Data Sciences Faculty Co-hire at the Pennsylvania State University. He received his B.S. in Chemical Engineering from the University of Minnesota Twin Cities and his Ph.D. in Chemical and Biological Engineering from Princeton University. His research group focuses on developing data-driven solutions to materials design challenges, especially those related to small, sparse, and noisy data sets within large design spaces.

Reed Williams is a Research Scientist at Siemens Foundational Technology, driving research in advanced manufacturing, geometric modeling, product design, and computer-aided engineering. His work is increasingly influenced by AI, from teaching Foundation Models to understand geometry to investigations into ways that AI can improve manufacturing.



Ryan Solnosky is a Teaching Professor in the Department of Architectural Engineering at the Pennsylvania State University at University Park and a registered Professional Engineer. He has taught courses in Architectural Engineering, Civil Engineering, and Pre-Major Freshman Engineering. His research centers on technology-enhanced active learning in the classroom to deepen course knowledge application and engagement, capstone delivery and scope enhancement to make future leaders, and lastly, team dynamics on multi-disciplinary teams.

R. Allen Kimel is an Associate Teaching Professor of Materials Science and Engineering at the Pennsylvania State University. He received a B.S. in Materials Science and Engineering at North Carolina State University, an M.S. degree in Chemistry from the University of North Carolina Greensboro, and a Ph.D. in Materials Science and Engineering at the Pennsylvania State University. At Penn State, he has created new lecture and laboratory courses on materials chemistry, microstructural analysis, quantification of materials properties, and materials sustainability.

Rebecca Napolitano is an Assistant Professor of Architectural Engineering at the Pennsylvania State University. She received her B.A. in Physics and Classical Languages from Connecticut College and her Ph.D. in Civil and Environmental Engineering from Princeton University. Her research interests include the synthesis of quantitative and qualitative data about the built environment for community advocacy.