Journal of the Midwest Association for Information Systems

Volume 2025 | Issue 2 Article 1

Date: 07-01-2025

On Higher Education, the Workplace of the Future, and Generalized Large Language Models: Research Questions in a Time of Technological Change and Adaptation

Barbara D. Klein
University of Michigan-Dearborn, bdklein@umich.edu

Rassule Hadidi
Metro State University, Rassule.Hadidi@metrostate.edu

Abstract

Resistance to and gradual adoption of technological innovations is not new to educational systems and, in some ways, continues as generalized large language models become widespread and, at times, seem to threaten traditional ways of teaching and learning. As educators and students experiment with these models innovatively, educational institutions will be compelled to evolve in ways that will enhance student preparation for the future workplace. This paper will discuss the potential impacts of generative language models on the workplace of the future, implications for workforce preparation, and implications for faculty teaching with an emphasis on business and information systems management. Research questions are posed to suggest future research opportunities for scholars working in the area.

Keywords: generalized large language models; higher education; workplace of the future; information systems pedagogy

DOI: 10.17705/3jmwa.000093

1. Introduction

Prior to the widespread availability and use of computers and calculators, occupations that required skill in arithmetic calculation such as bookkeeping, engineering, and surveying were widespread and schools devoted significant time and resources to teaching students to quickly and accurately perform these calculations. As the adoption of computers and calculators has become widespread, teaching and learning in primary schools is focused primarily on the development of conceptual understanding of mathematics. A parallel nascent development is now occurring in the domain of language as powerful generative large language models are becoming widely available. These models are trained on vast repositories of text data and, once trained, can generate sophisticated text material in a wide variety of domains and for a wide variety of tasks. These generative language models are likely to transform the workplace and the types of jobs humans will perform in the future. For certain types of routine language generation, it may soon seem as unimaginable to assign the task to humans as it would have seemed for the past several decades to assign routine arithmetic calculation tasks to humans. A parallel transformation to the one seen earlier in the teaching of mathematics is likely in education with a transformation in the way humans are taught about language generation and the language generation tasks for which they will be prepared in schools. And just as teachers were once alarmed by the use of calculators in schools and worried that students would lose their ability to perform routine mathematical calculations, many teachers and university professors are alarmed today by the existence and the use of generative language models. Indeed, for many, the focus is more on detection and prevention of the use of generative large language models with an emphasis on pursuing academic integrity violations rather than on preparing students for the future workplace.

This paper will discuss the potential impacts of generative large language models on the workplace of the future, implications for workforce preparation, and implications for faculty teaching with an emphasis on business and information systems management.

2. From Early Conceptions of Artificial Intelligence to Generative Large Language Models: Implications for a Transformed Human-Machine Relationship

The roots of artificial intelligence lie in the work of path breaking scholars such as Herbert Simon, Allen Newell, and Marvin Minsky who sought to develop ways in which computer programs could mimic activities generally thought to be in the exclusive domain of human intelligence (Haenlein and Kaplan, 2019). Fields such as cognitive science, linguistics, logic, and computer science were applied to these developments. Early debates addressed the issue of whether the best approach would lie in systems designed to mimic general intelligence or those designed to perform specific tasks. For a time, the field of expert systems seemed promising as efforts were made to develop a deep understanding of expertise in specific fields and then apply that understanding to artificial intelligence systems designed to replicate the performance of experts in narrow domains. More recently generalized artificial intelligence such as generalized large language models have used machine learning algorithms along with very large collections of training data with the aim of replicating human understanding and generation of language (Chun and Noveck, 2025; Ooi et al., 2025). These systems have the potential to perform a wide variety of tasks and are currently of great interest to business organizations and educational institutions.

Computer vision and robotics have already transformed some aspects of jobs and the workplace in domains such as the manufacturing of goods (Brynjolfsson and McAfee, 2016). Similarly, generalized large language models hold the potential to radically transform knowledge work and the educational systems designed to prepare students for roles in the knowledge economy. In doing so, these systems hold the potential to reimagine the relationship between humans and machines in ways that Simon, Newell, and Minsky (Simon and Newell, 1964; Minsky, 1988) imagined at the beginning of the age of artificial intelligence.

3. The Workplace of the Future of Generalized Large Language Models

The workplace and ways in which business practices are designed and executed have long been influenced by new technologies and algorithms. While workers once used manual approaches and adding machines to perform straightforward calculations, the adoption of spreadsheets changed the way in which quantitative business processes were executed. Some jobs such as linotype operators (https://en.wikipedia.org/wiki/Linotype_machine) simply ceased to exist with the development of computerized tools for copy layout. More recently, organizational tasks and roles have been transformed by the use of sophisticated computer algorithms that support organizational processes and decision making

(McAfee and Brynjolfsson, 2017).

Potential effects of generalized large language models in the workplace of the future include automation of jobs focused on knowledge work, the shift to increasingly more complex tasks in some jobs, the creation of new types of jobs such as roles focused on the integration of generalized large language models into business processes and decision making, increased focus on collaboration between artificial intelligence agents and humans, and the need for employees to develop skills focused on tasks that cannot be accomplished by artificial intelligence agents. The skills needed for language-related tasks in the workplace are likely to shift significantly as routine writing tasks will increasingly be allocated to generalized large language models. Humans will likely focus on skills such as the detection and correction of errors, synthesis, prompt generation and refinement, creativity, and innovation. As prompt engineering becomes increasingly important in knowledge work, professionals in a wide variety of domains will need to understand the characteristics and limitations of generalized large language models; understand how to craft, test, and refine effective prompts; and understand how to evaluate and appropriately use the outputs of these models.

4. On Adaptation to Technological Change in Education: A Brief History of Calculator Adoption and Use

Resistance to and gradual adoption of technological innovations is not new to elementary and secondary schools, colleges, and universities (Ellington, 2003; Hembree and Dessart, 1986). The case of the electronic calculator provides a parallel to what we are seeing today during the early stages of generalized large language models. Initially, the electronic calculator was fiercely resisted in elementary and secondary schools and teachers and parents feared that calculator use would harm students' calculation skills and mathematical understanding. As calculators became more affordable, they inevitably made their way into classrooms and homes and gradually mathematics curricula and pedagogy adapted to incorporate the calculator as a tool to be used alongside the human mind. As students were freed from the need to focus on calculations, more challenging problems and strategies were introduced into schools (Ruthven, 1998). Eventually the electronic calculator became a routine aspect of mathematics curricula and classrooms and its use even spread to standardized exams. As graphing and scientific calculators became widespread, students and knowledge workers were expected to have the skills needed to formulate problems and interpret outputs generated by electronic calculators. As calculator apps have become available on cell phones, many teachers and employers would be very surprised to encounter students or employees who do not know how to appropriately use them.

A similar evolution is likely to occur with the adoption of generalized large language models in educational settings and the workplace. While some academics express alarm about their use in the classroom and it is not uncommon for academic integrity complaints to be filed against their users, over time we expect the use of generalized large language models to be widespread and for the skills and knowledge students and employees bring to language-based tasks to shift toward those requiring higher order thinking, creativity, and innovation. Routine writing tasks are likely to shift toward execution by generalized large language models over time, and students and employees will be expected to have the prompt engineering skills needed to effectively put these models to work (Phoenix and Taylor, 2024; Tababalan, 2024). Employees and organizations are likely to adapt to this transformed environment by emphasizing employee selection and training with an emphasis on technical skills such as machine learning and chatbot development as well as business process and change management skills (Babashahi et al., 2024).

5. Higher Education and Generalized Large Language Models

As with the initial availability and adoption of calculators, students and faculty have tended to view the availability of generalized large language models from different perspectives. Students embracing them as labor saving tools that can execute assignments focused on lower-order cognitive domains (Kiel and Linkov, 2025; Lubbe et al., 2025) and as tools that can generate study aids such as flashcards and study guides. At least some faculty have been alarmed by the capacity of these tools to execute academic tasks that they have created to be done solely by their students (Wong, 2024).

While we currently see significant emphasis on academic integrity issues in the academic use of generalized large language models (e.g., Eke, 2023; Plata et al., 2023; Sullivan et al., 2023; Wiredu et al., 2024), we anticipate that focus will shift to the design and adaptation of curricula and schools to address the workplace of the future in which humans will perform knowledge work in conjunction with generalized large language models (e.g., Yusuf et al., 2024).

Despite the current focus on academic integrity in the use of generalized large language models, we also see promising efforts to better understand how faculty and students can productively, effectively, and even ethically use these tools (Chiu, 2024; Hadidi and Klein, 2025). In this spirit, we invite contributions to the *Journal of the Midwest Association for Information Systems* building on the literature seeking to enhance our understanding of the ways in which generalized large language models can be incorporated into information systems curricula and classes in order to better prepare students for the workplace of the future.

6. Ethical Consistency and Generalized Large Language Models

University faculty have long been concerned with plagiarism and academic dishonesty, especially in the context of writing assignments that ask students to think critically and generate original analysis and text. As access to written material became more freely available and as word processing software made it easier for students to copy text and represent it as their original work, technical approaches to detecting such academic integrity issues were developed and adopted by universities. These systems tended to develop a mindset among faculty and administrators in which they viewed their roles as encompassing the prevention, detection, and correction of deception in student writing. As generative large language models became available to faculty and students, this mindset has persisted among some faculty and the desire among some to prohibit the use of these tools and to detect forbidden use of generative large language models has developed. In some cases, faculty have been encouraged to adopt and publish course-level policies related to the use of these tools (e.g., Course Policies and Syllabi Statements, 2025). Simultaneously, universities have begun to encourage faculty to think creatively and broadly about the use of generative large language models as productivity tools and as tools for enhancing teaching, learning, and scholarship (e.g., Course and Assignment Re-Design, 2025). For example, a recent eight-week training course offered by the University of Michigan focused on topics such as privacy and security, content creation using generative AI, prompt literacy, and generative AI as a communication tool (The University of Michigan 8-Week AI Challenge, 2025). These competing views of learning and work may, in some cases, encourage faculty to adopt an attitude that use of generative language models is "OK for me (i.e., faculty), but not for you (i.e., students)." A more forward-looking perspective guided by the search for new ways to use generalized large language models as tools for teaching and learning is arguably a healthier approach since the workplace of the future will be one in which the use of generative large language models will be as common as the use of word processing and electronic spreadsheet tools are today.

The principle of ethical consistency offers a way forward during this time of rapid change in educational institutions and the workplace. Ethical consistency is the principle that one should use the same ethical principles to evaluate one's own behavior and actions and the behavior and actions of others (Hopkins et al., 2008). This is in contrast to the notion of ethical hypocrisy which is the notion of evaluating the behavior and actions of others differently (and generally using more demanding standards) compared to one's own behavior and actions (Foad et al., 2022). A pedagogical approach characterized by the principle of ethical consistency recognizes the challenges and opportunities of a transformed workplace both in terms of faculty teaching and scholarship as well as the workplaces students will experience during their own careers.

We invite manuscripts focused on the following research question: Given recent developments related to generalized large language models, how can we approach our scholarly work and teaching and learning in the classroom in a spirit of ethical consistency?

7. Additional Research Questions

In addition to research guided by the principle of ethical consistency, we invite manuscripts for submission to the *Journal of the Midwest Association for Information Systems* focused on the following research questions related to higher education, the workplace of the future, and generalized large language models.

1. What are the prompt engineering skills and abilities likely to be of most use in the future workplace? How can these skills and abilities best be developed in our students?

Increasingly, preparation for the workplace will require faculty to focus on the design and delivery of curricula incorporating prompt engineering (Knoth et al, 2024). Students will need to develop the ability to craft and revise effective prompts to serve as inputs to generalized large language models (Federiakin et al., 2024; Lee and Palmer, 2025). To accomplish this, students will need a deep understanding of language, knowledge of the problems and tasks to be addressed by the generalized large language model, an understanding of the goals of the model output, and the ability to evaluate and refine outputs in an iterative fashion (Cain, 2024). Prompt engineering curricula will need to focus on the fundamentals of artificial intelligence, natural language processing, and the domain knowledge in which students will develop professional expertise. Case studies, project-based learning, community-based experiences, and internships are likely to play a role in the development of the required knowledge and skills.

2. What knowledge and skills should students be taught about the use of generalized large language models in intercultural and global contexts?

Given that students will spend much of their business careers working in global organizations and intercultural contexts and given that generalized large language models may emphasize text written in a specific language, how can students be taught to navigate these language challenges in global business environments (Dai et al., 2025; Wu et al., 2025).

3. What types of jobs are likely to disappear and what types of jobs are likely to be created due to the existence of generalized large language models? What are the implications for pedagogical design, course topics, and job placement for students?

Generalized large language models have the potential to facilitate significant redesign of jobs. In the process some jobs may disappear or become far less numerous, while other jobs may be created that do not currently exist. While scholarly work addressing this topic is in its infancy, some guidance has begun to emerge. Jobs such as language transcription may disappear entirely, while others such as customer service positions may become far less numerous and require more complex skills targeted at tackling unusually complex customer issues. Jobs focused on training generalized large language models and those focused on compliance issues associated with these models may become more numerous over time. Employees with language and cultural skills may be in demand as organizations focus on global deployment of their generalized large language models. Human resource specialists with expertise in artificial intelligence may work in new jobs focused on collaborative work design between humans and generalized large language models. As these shifts in work design and jobs occur, faculty and students will need to adapt to new pedagogical approaches, course topics, and career preparation. Preparation for managerial roles will also need to evolve so managers are prepared to provide guidance and supervision for employees using generalized large language models both informally and formally in organizations (Retkowsky et al., 2024; Shokran et al., 2025).

4. What skills related to innovation will students need for a workplace in which generalized large language models will be deployed? As routine tasks done by knowledge workers in the past are shifted to these models, how do we ensure that our students are able to compete in a transformed labor market?

The workplace of the future is likely to offer jobs that are increasingly complex, more demanding, more focused on innovation and creativity, and more focused on assessment of the outcomes of generalized large language models (Sollner et al., 2025). Pedagogy, course content, and experiential learning will need to adapt to prepare students for this workplace shift.

5. What knowledge and skills will students need to navigate data privacy and data accuracy issues as they use generalized large language models in the workplace and as personal assistants?

In order to effectively use generalized large language data models in the workplace for organizational tasks and as personal assistants, employees will need a strong understanding of data privacy and data accuracy. Students and employees will need to understand how training data, test data, input prompts, and model outputs are collected and stored and how breaches can expose data in ways that may be harmful to individuals and organizations (Das et al., 2025; Mesko and Topol, 2023). Knowledge of data accuracy frameworks (Wang and Strong, 1996) and error detection and correction (Klein et al., 1997; Klein, 2001) will also be needed. Privacy settings and encryption standards are likely to evolve over time and organizations and universities will need to update their training and curricula to help employees maintain up-to-date knowledge of these technical issues.

6 How can student outcomes and the effectiveness of degree programs be assessed and revised over time to evaluate the design and implementation of curricula infused with generative large language models, either by design or by students' informal adoption and use of the tools?

Generalized large language models have been found to perform well on at least some course assessments designed and used before the era of generalized artificial intelligence, and some faculty have expressed concerns about student use of these models on course assignments and assessments (Nikolic et al., 2023). However, as the nature of work is transformed by generalized large language models, learning outcomes and goals will inevitably evolve to focus on more complex skills and knowledge emphasizing higher order thinking skills in order to deliver career preparation needed for this transformed workplace. Course material, assignments, and assessments can then be modified in ways that acknowledge and use these models to achieve student and program level learning outcomes. Employer feedback related to student adaptability in a workplace in which generalized large language models are used can also be adapted as an indirect measure of learning outcomes (Khlaif et al., 2024; Lubbe et al., 2025; Xia et al., 2024).

8. Conclusion

Generalized large language models have the potential to transform higher education and the workplace. By embracing rather than resisting this and similar technologies, faculty can collaborate with their students and industry partners to ensure that students receive an education focused on the complex, higher-order thinking skills and abilities that will prepare them to thrive in their careers. Research focused on the questions discussed in this article, as well as related questions, holds the potential to improve our understanding of practices and strategies that can best deliver the promise of these tools to our students, organizations, and society.

9. Overview of the Contents of this Issue

This issue of the journal includes two other articles. Vlad Krotov and Pitzel Krotova in their interesting and teaching related article provide a detailed description of how they have used the Scrum methodology to manage group project in a face-to-face graduate Project Management course. They discuss the feedback they have received from their students about the effectiveness of this approach.

Ricardo de Deijn and Rajeev Bukralia in their interesting and timely article discuss the potential uses of prompt-generated synthetic datasets for snow detection. The study explores the advantages as well as disadvantages of using such datasets for this application.

We appreciate and wish to acknowledge the contributions of reviewers for this issue of the journal, including Queen Booker (Metropolitan State University), Mari Buche (Michigan Technological University), Omar El-Gayar (Dakota State University), Bryan Hosack, (Penske Logistic), Alanah Mitchell (Drake University), Kevin Scheibe (Iowa State University) and Troy Strader (Drake University).

10. References

Babashahi, L., Barbosa, C. E., Lima, Y., Lyra, A., Salazar, H., Argolo, M., de Almeida, M. A., and de Souza, J., M. (2024). AI in the workplace: A systematic review of skill transformation in the industry. *Administrative Sciences*. 14(127), doi.org/10.3390/admsci14060127.

Bynjolfsson, E., and McAfee, A. (2016). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York: W.W. Norton & Company.

Cain, W (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *Tech Trends*. 2024(68), 47-57.

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*. 6, doi.org/10.1016/j.caeai.2023.100197.

Chun, S. A., and Noveck, B. S. (2025). Introduction to the special issue on ChapGPT and other generative AI commentaries: AI augmentation in government 4.0. *Digital Government: Research and Practice*. 6(1).

Course and Assignment (Re-)Design. (2025). https://genai.umich.edu/resources/faculty/redesigning-assessments, accessed May 23, 2025.

Course Policies and Syllabi Statements. (2025). https://genai.umich.edu/resources/faculty/course-policies, accessed May 23, 2025.

Dai, D. W., Suzuki, S., and Chen, G. (2025). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review*. 16(2), 763-774.

Das, B. C., Amini, M. H., and Wu, Yanzhao. (2025). Security and privacy challenges of large language models: A Survey. *ACM Computer Surveys*. 57(6), Article 152.

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*. 13, doi.org/10.1016/j.jrt.2023.100060.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*. 34(6), 433-463.

Federiakin, D., Molerov, D., Zlatkin-Troitschanskaia, O., and Maur, A. (2024). Prompt engineering as a new 21st century skill. *Frontiers in Education*. Doi: 10.3389/feduc.2024.1366434.

Foad, C., Haddock, G., and Maio, G. (2022). Hypocrisy in ethical consumption. *Frontiers in Psychology*. 13:880009, doi: 10.3389/fpsyg.2022.880009.

Hadidi, R., and Klein, B. D. (2025). Exploring faculty and students' ethical uses of GAI in teaching and learning focusing on Bloom's Taxonomy and the needed skills development. *Journal of the Midwest Association for Information Systems*. 2025(1), Article 1.

Haenlein, M., and Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*. 61(4), 5-14.

Hembree, R., and Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for Research in Mathematics Education*. 17(2), 83-98.

Hopkins, W. E., Hopkins, S. A., and Mitchell, B. C. (2008). Ethical consistency in managerial decisions. *Ethics & Behavior*. 18(1), doi: 10.1080/10508420701519544.

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M., and Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*. 14, doi.org/10.3390/educsci14101090.

Kiel, J. M., and Linkov, F. (2025). Creative teaching innovations to align with Bloom's taxonomy: Leadership and immersion experiences. *Journal of Health Administration*. 41(1), 141-148.

Klein, B. D. (2001). Detecting errors in data: Clarification of the impact of base rate expectations and incentives. *Omega*. 29(5), 391-404.

Klein, B. D., Goodhue, D. L., and Davis, G. B. (1997). Can humans detect errors in data? Impact of base rates, incentives, and goals. *MIS Quarterly*. 6(1), 169-194.

Knoth, N., Tolzin, A., Janson, A., and Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computer and Education: Artificial Intelligence*. 6(2024), 100225.

Lee, D., and Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*. 22(7), doi.org/10.1186/s41239-025-00503-7.

Lubbe, A., Marais, E., and Kruger, D. (2025). Cultivating independent thinkers: The triad of artificial intelligence, Bloom's taxonomy and critical thinking in assessment pedagogy. *Education and Information Technologies*.

https://doi.org/10.1007/s10639-025-13476-x.

McAfee, A., and Brynjolfsson, E. (2017). *Harnessing our digital future: Machine platform crowd*. New York: W.W. Norton & Company.

Mesko, B., and Topol, E. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npr Digital Medicine*. 6(120), doi.org/10.1038/s41746-023-00873-0.

Minsky, M. (1988). Emotions and the Society of Mind. In: Clynes, M., Panksepp, J. (eds) Emotions and Psychopathology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-1987-1_7

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., and Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*. 48(4), 559-614.

Ooi, K.-B., Tan, G. W., Al-Emran, M., Al-Sharafi, M.A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P. Sarker, P., Sharma, A., Samuel, C. T., Wanba, F., and Wong. L.-W. (2025). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*. 65(1), 76-107.

Phoenix, J., and Taylor, M. (2024). *Prompt engineering for generative AI: Future-proof inputs for reliable AI outputs*. Sebastopol, CA: O'Reilly Media.

Plata, S., De Guzman, M. A., and Quesada, A. (2023). Emerging research and policy themes on academic integrity in the age of ChatGPT and generative AI. *Asian Journal of University Education*. 19(4), doi.org/10.2419/ajue.v19i4.24697.

Retkowsky, J., Hafermalz, E., and Huysman, M. (2024). Managing a ChatGPT-empowered workforce: Understanding its affordances and side effects. *Business Horizons*. 67, 511-523.

Ruthven, K. (1998). The use of mental, written, and calculator strategies of numerical computation by upper primary pupils within a 'calculator aware' number curriculum. *British Educational Research Journal*. 24(1), 21-42.

Shokran, M., Islam, M.S., and Ferdousi, J. (2025). Harnessing AI adoption in the workforce: A pathway to sustainable competitive advantage through intelligent decision-making and skill transformation. *American Journal of Economics and Business Management*. 8(3), 954-976.

Simon, H and Newell, A. (1964). Information processing in computers and man, *American Scientist*, 52:281-300.

Sollner, M, Arnold, T., Benlian, A., Bretschneider, U., Knight, C., Ohly, S., Rudkowski, L, Schreiber, G., and Wendt, D. (2025). ChatGPT and beyond: Exploring the responsible use of generative AI in the workplace. *Business & Information Systems Engineering*. 67(2), 289-303.

Sullivan, M., Kelly, A., and Mclaughian, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*. 6(1), 1-10.

Tababalan, M. (2024). *Prompt engineering using ChatGPT: Crafting effective interactions and building GPT apps*. Herndon, Virginia: Mercury learning and Information.

The University of Michigan 8-Week AI Challenge. (2025). https://its.umich.edu/computing/ai/8-week-challenge, accessed May 23, 2025.

Wang, R. Y., and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*. 12(4), 5-33.

Wirdu, J. K., Abuba, N. S., and Zakaria, H. (2024). Impact of generative AI in academic integrity and learning outcomes: A case study in the Upper East Region. *Asian Journal of Research in Computer Science*. 17(7), 214-232.

Wong, W. K. O. (2024). The sudden disruptive rise of generative artificial intelligence? An evaluation of their impact on higher education and the global workplace. *Journal of Open Innovation: Technology, Market, and Complexity*. 10(024).

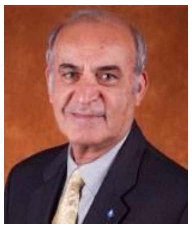
Wu, T. T., Hapsari, I. P., and Huang, Y. M. (2025). Effects of incorporating AI chatbots into think–pair–share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 1–39. https://doi.org/10.1080/09588221.2025.2478271.

Xia, Q., Weng, X., Quyang, F., Lin, T. J., and Chiu, T. K. F. (2024). A scoping review of how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*. 21(40). doi.org/10.1186/s41239-024-00468-z.

Yusuf, A., Pervin, N., and Roman-Gonzalez, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*. 21(21), doi.org/10.1186/s41239-024-00453-6.

Author Biographies

Barbara D. Klein is Professor of Management Information Systems and Information Technology Management at the University of Michigan-Dearborn. She received her Ph.D. in Information and Decision Sciences from the University of Minnesota, her M.B.A. from the State University of New York at Albany, and her B.A. from the University of Iowa. Professor Klein has published in the *Journal of the Midwest Association for Information Systems*, *MIS Quarterly*, *Omega*, *Database*, *Information & Management*, *Information Resources Management Journal*, and other journals. Her research interests include information quality, user error behavior, and information systems pedagogy. Professor Klein has also worked in the information systems field at IBM, Exxon, and AMP.



Rassule Hadidi is Dean of the College of Business and Management, Metro State University, Minneapolis, Minnesota. His current research areas of interest include Gen AI and its implications for teaching and learning, online and blended teaching and learning pedagogy and its comparison with face-to-face teaching; curriculum development and quality assessment; cloud computing and its applications for small and medium-sized enterprises; and quality of online information. He has served as the president as well as the At-Large Director of the Midwest Association for Information Systems and is the founding Managing Editor of the *Journal of the Midwest Association for Information Systems*. He is an AIS Distinguished Member – Cum Laude and is a member of the Board of Directors of the Society for Advancement of Management.

Journal of the Midwest Association for Information Systems

Volume 2025 | Issue 2 Article 2

Date: 07-01-2025

Using Scrum for Managing Group Projects in a Face-to-Face Information Systems Project Management Course

Vlad Krotov
Murray State University, vkrotov@murraystate.edu

Pitzel Krotova
Murray State University, mkrotova@murraystate.edu

Abstract

Scrum has become the most widely used Agile project management methodology, and educational institutions around the globe are integrating Scrum into their curricula. Using Scrum in class can potentially help educators not only equip students with hands-on knowledge of the Scrum framework and Agile values, but also teach students valuable soft skills and foster positive group work dynamics. This article provides a detailed account of how Scrum has been adapted to manage group projects in an on-campus graduate Information Systems course devoted to Project Management. The adaptation is deliberately simple and "low-tech" to keep administrative overhead low and to appeal to a broad audience of educators. An online survey was conducted among the students to assess the usefulness of the Scrum adaptation. While certain issues were raised, the feedback was mostly positive. The survey results indicate that implementing the framework may help students acquire practical knowledge of Scrum, Agile, and project management, and enhance positive group work dynamics.

Keywords: Scrum, Agile, Education, Adaptation, Project Management

DOI: 10.17705/3jmwa.000094

1. Introduction

Scrum is arguably the most widely used Agile Project Management methodology in the Information Technology (IT) field. The 17 "State of Agile Report" published in 2023 by VersionOne reveals that 71% of organizations surveyed practice Agile development methodologies; 63% use Scrum for managing their projects (VersionOne, 2023). The popularity of Scrum as alternative to the traditional, "waterfall" approach to managing projects is not surprising: Scrum is "a framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value." (Sutherland & Schwaber, 2013, p.3). Scrum works exceptionally well for projects with evolving or unclear requirements, high uncertainty, and the need for frequent stakeholder feedback (Highsmith & Cockburn, 2001). It is most effective when used by small, cross-functional, self-organizing teams in environments that support adaptability, rapid delivery of value, and continuous improvement (Moe et al., 2010; Rigby et al., 2016).

Given the popularity of Scrum as a project management framework, practical knowledge of this Agile methodology is a valuable skill for IT and general business professionals in all industries. Because of that, educational institutions around the globe are integrating Scrum into their curricula (Matkovic et al., 2016; Reynolds et al., 2023). The main goal of this article is to assist educators from various fields in integrating Scrum and Agile into their curricula by providing a detailed account of how Scrum has been adapted for managing group projects involving the creation of detailed project management plans in a graduate, on-campus Information Systems course devoted to Project Management. The implementation of Scrum described in this article is deliberately simple and "low-tech" to appeal to educators from various fields, including non-technical ones. This is also in alignment with the spirit of Scrum, since the methodology does not prescribe any specific tools or technologies (although team members are free to use them as needed) (Sutherland & Schwaber, 2013). Additional Scrum tools, techniques, and artifacts can be added to this customization of Scum—depending on the nature and the goals of the class where it is used.

The rest of this article is structured as follows. First, it provides a brief overview of the Scrum methodology and explains how and why Scrum is increasingly used in education. This section highlights the benefits of using Scrum in the educational process and the need to adapt the framework—originally developed for software projects—to the educational context. Next, the article offers a detailed account of how Scrum was implemented to manage group work in a project management course based on the traditional "waterfall" approach. The unique features and benefits of this adaptation, along with the instructor's insights and recommendations, are discussed as part of this implementation. The article concludes with an empirical evaluation of the effectiveness of the approach, presenting both quantitative and qualitative feedback from students, followed by recommendations for improvement and final reflections on the use of Scrum in education.

2. Using Scrum in Class

2.1 The Scrum Framework

In comparison to traditional, "waterfall" project management methodologies, the Scrum framework is deliberately "light" (see Figure 1). The framework is based on simple, generative rules that empower people and give them an opportunity to show their best performance. For example, the framework does not prescribe any technologies or tools; it is up to the team to decide which technologies and tools should be used to deliver a product. As often mentioned by Jeff Sutherland, one of the creators of Scrum, in his public speeches, the Scrum framework is easy to understand, but often takes a lifetime to master. In most cases, effective use of Scrum is a result of extensive experience and tacit knowledge rather than intellectual understanding of this framework (Sutherland & Schwaber, 2013).

Scrum emphasizes empirical process control, which means that project-related decisions are based on observation and feedback rather than rigid planning. This allows for flexibility and adaptability throughout the project. The framework encourages regular reflection and improvement through retrospective meetings held after each Sprint, where the team discusses what went well, what could be improved, and decides on actions for the next iteration.

Overall, Scrum provides a framework that enables teams to deliver high-quality products in a collaborative and efficient manner, with a focus on continuous learning and adaptation.

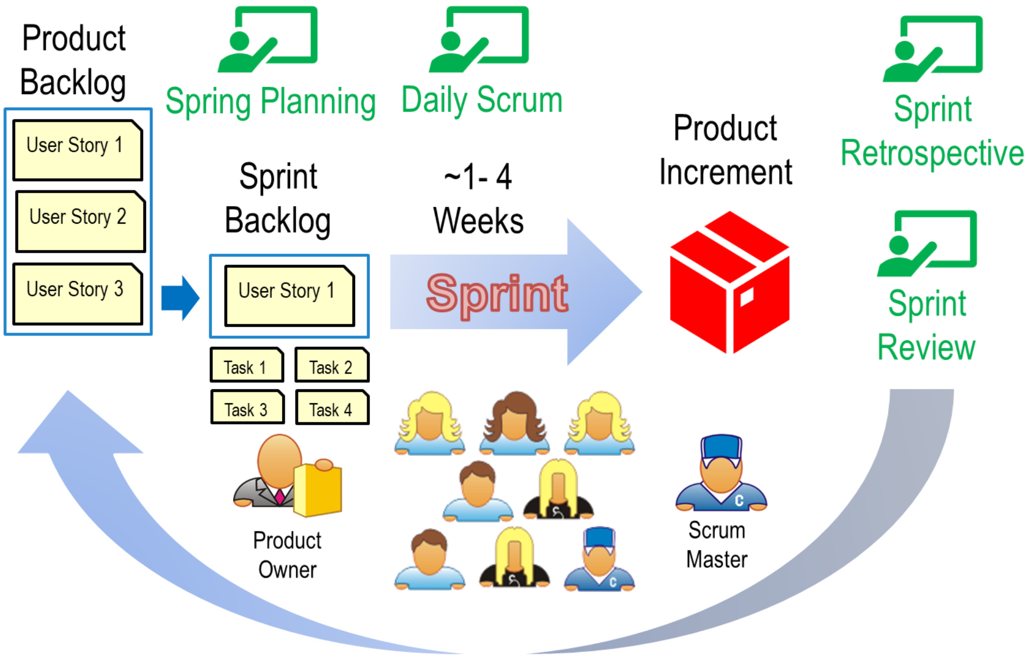


Figure 1. The Scrum Framework (Adapted from Mountain Goat Software, 2005)

Scrum begins by assembling a Scrum Team consisting of six to nine members (ScrumEductor.org, 2023). This team includes a Scrum Master, responsible for facilitating teamwork, removing obstacles, and ensuring adherence to Scrum values and processes; a Product Owner, who represents the customer and has final authority over product requirements; and a Development Team, composed of individuals with the technical and non-technical skills needed to deliver the product.

Once the team is formed, they create a Product Backlog, a list of tasks or User Stories that describe product features from the customer's perspective (e.g., "As a traveler, I want to browse different travel insurance options available on the website"). The Product Backlog captures the full scope of the product's functionality and must be prioritized. The backlog remains flexible, allowing for modifications and reprioritization between Sprints.

After establishing the Product Backlog, the team selects a set of prioritized items to work on during a Sprint—a time-boxed development cycle lasting approximately 2-4 weeks. The Scrum Team commits not only to implementing these selected features but also to delivering a shippable product increment, meaning a functional and demonstrable portion of the product.

Throughout each Sprint, the team holds Daily Scrums, which are short, 15-minute stand-up meetings. Each team member answers three key questions:

1. What did you do yesterday?
2. What will you do today?
3. Are there any obstacles in your way?

These meetings are not meant for problem-solving but rather for team members to stay aligned and commit to their tasks in front of their peers. Problem-solving discussions take place outside the Daily Scrum, helping eliminate unnecessary meetings.

At the end of each Sprint, the Scrum Team conducts a Sprint Review, where they present their completed work and demonstrate new product features. These meetings involve the entire team and should be open to anyone in the organization who wants to attend.

Additionally, the team holds Sprint Retrospective meetings between Sprints. The goal of these sessions is to continuously improve the team's workflow by reflecting on what worked well, what didn't, and what can be improved. While these retrospectives can be conducted after every Sprint, they are especially valuable when the team identifies areas needing refinement.

2.2 The Benefits of Scrum in Higher Education

While Scrum is often associated with computer science education, the framework is useful in a variety of educational contexts, such as nursing and education (Zahorodko, 2023). Research indicates that Scrum can enhance the learning experiences of students in relation to teamwork and diverse subjects by providing structured frameworks for collaboration and communication involving students and their instructors (Reynolds et al., 2023).

One of the key advantages for using Scrum in higher education is to foster student collaboration and engagement. With the help of Scrum, students can engage not only with the instructor and other team members, but also—in the case of client-sponsored projects—with external stakeholders. Students can take on roles that mirror real-world responsibilities, thus enhancing their learning experience through service learning and preparing themselves for the industry (Dong, 2023). Scrum also allows students to get practical insights into project management and teamwork. Similarly, the use of Scrum in project-based courses has been shown to improve student performance and satisfaction, as it encourages team self-regulation and constant reflection on the learning progress and the quality of the deliverable produced as a result of this project (Vogelzang et al., 2021; Fernandes et al., 2021).

The use of Scrum for managing student group projects also addresses challenges related to team dynamics and group project management in higher education. The role of the Product Owner, which can be assumed by the instructor teaching the course, can allow the instructor to guide student projects while also giving students the freedom and responsibility to self-organize their work (Baham, 2020). By establishing clear roles and responsibilities via formal roles, Scrum helps to streamline communication and coordination among team members, which is essential for successful project outcomes (Wolff, 2024). Additionally, the structured nature of Scrum ceremonies, such as Daily Scrums and Sprint Reviews, promotes student accountability and continuous feedback from the instructor, which are important for student learning (Pope-Ruark, 2012; Morales-Trujillo et al., 2021).

Furthermore, the use of Scrum in remote learning environments has gained traction as well (Wolff, 2024). Research indicates that Scrum can enhance communication and collaboration among geographically distributed teams, leading to improved project management and learning outcomes (Wolff, 2024). This adaptability is crucial in today's educational landscape, where remote learning is becoming increasingly prevalent.

2.3 The Main Uses of Scrum

There are many ways in which Scrum is used in the educational context (Sharp et al, 2020). Some instructors use the so-called "Scrum games" as fun, team building activities in their classes that also equip students with hands-on understanding of Scrum and Agile principles (e.g., May et al., 2016). Scrum is also used to deliver educational content in various courses in a student-driven, self-organized, and self-paced manner (Rush & Connolly, 2020). One of the most popular uses of Scrum in class is for managing student group projects (e.g., Wagh, 2012; Manamendra et al., 2013; Zorzo, & Lucredio, 2013; Nejmeh & Weaver, 2014; Mahalakshmi & Sundararajan, 2015).

2.4 The Need for Scrum Adaptation

One of the challenges of using Scum in education is the fact that the framework was originally created for managing software development teams within real organizations. Some of the processes prescribed by Scrum cannot be implemented in class. For example, an instructor cannot have Daily Scrums with students, as most classes meet only 1-2 times per week. Thus, in order to make the framework useful in education, Scrum needs to be adapted to the educational context.

Several adaptations of Agile and Scrum to the educational context have been published in academic and practitioner outlets. Below, we discuss some of these implementations and highlight how the adaptation of Scrum described in this paper differs from these educational adaptations and adds unique value to learning.

First, there are conceptual papers that provide the foundation for "agile teaching." For example, Krehbiel et al. (2017) present an "Agile Manifesto for Teaching and Learning," which reimagines the Agile Manifesto's principles within the educational context. The authors argue for learner-centered, adaptive, and feedback-driven approaches to pedagogy, paralleling the values of Agile software development. While not offering specific Scrum adaptation instructions, this conceptual paper offers a foundational rationale for applying Agile principles in the classroom and includes some guidance on how Scrum can be adopted in class.

Second, there are studies detailing various approaches to using Scrum in the classroom. For example, Rush & Connolly (2020) provide a fairly detailed description of how Scrum was used to deliver educational content related to project management in an Agile fashion. Baham (2020) examines how to improve Business Product Owner (BPO) engagement in student Scrum projects, emphasizing such practices as co-creation workshops and formalized role assignments to enhance stakeholder involvement and project success. Babik (2022) offers a detailed account of a hands-on "Scrum Boot Camp" that builds student confidence in Scrum principles and tools through active learning in both face-to-face and online courses. Thouin and Hefley (2024) explain how they use an experiential simulation based on SimAgile to teach Scrum Product Owner skills, finding that students gain authentic insights into product management and decision-making processes from the simulation exercise. Finally, Mutchler et al. (2024) demonstrate how Scrum can be adapted for concept-heavy courses—specifically Information Security Management—by allowing students to tailor deliverables to their interests, resulting in higher engagement and perceived skill development.

There are also proprietary Scrum-based educational methodologies developed by organizations such as eduScrum (eduScrum, 2020) or ScrumEducator (ScrumEducator.org, 2023), as well as literature reviews of various Scrum implementations by Salza et al. (2019), López-Alcarria et al. (2019), and Reynolds et al. (2023). Collectively, these rarticles communicate general trends in Scrum adaptation and underscore the adaptability of Scrum across diverse course types, student populations, and learning objectives.

We believe that the Scrum adaptation described in this paper provides a unique, valuable, and practical contribution to the literature on the use of Scrum in education. The adaptation includes some unique and valuable features that were not found in previously published adaptations of Scrum. Some of these unique features and value propositions are discussed below.

First, the article provides a detailed and rich description of how Scrum can be adopted in the classroom. These rich descriptions are supplemented with the instructor's own experiences and insights. Many of the reviewed studies lack specific, practical details on implementing Scrum in class.

Second, although various technological tools can be added based on the needs and goals of the course, the Scrum adaptation described here is deliberately simple and "low-tech." It emphasizes tangible tools like physical Scrum boards and burndown charts, making implementation accessible to instructors who do not have access to technical resources or who prefer not to invest time and money into learning new tools.

Third, the emphasis of this adaptation is on the translation and implementation of Scrum in an Information Systems course devoted to the "waterfall" approach to Project Management. The project managed via Scrum involves the creation of traditional project management artifacts, such as a project charter, business case, and quality assurance plan. This integration of Agile practices with "waterfall" material allows the instructor to "kill two birds with one stone": (1) familiarize students with the "waterfall" approach to project management via Team-Based Learning (TBL), and (2) equip students with practical, hands-on skills related to Scrum and Agile.

Fourth, this adaptation focuses on quality face-to-face interactions among students and the instructor. We believe that face-to-face experiences provide students with rich opportunities to learn from their peers and the instructor. The adaptation emphasizes the development of teamwork and other soft skills through these interactions, which is grounded in the principles of Team-Based Learning (TBL).

Fifth, this paper proposes integrating Scrum not as a short workshop, activity, or assignment, but as a full-scale, long-term, semester-long exercise. While the Scrum framework is relatively simple and easy to learn, it often takes a long time to master (Sutherland & Schwaber, 2013). This is largely due to the fact that Scrum is largely about doing (tacit knowledge) rather than intellectual understanding Scrum roles, processes, and artifacts (explicit knowledge). Thus, providing students with ample hands-on opportunities to gain working knowledge of Scrum and Agile processes is essential to learning these important topics.

Finally, the adaptation includes both quantitative and qualitative evaluation of its effectiveness as perceived by students. This feedback allows readers to gain additional insights into how the implementation of Scrum can be improved further or tailored to their unique educational goals or contexts.

3. Scrum Adaptation

The Scrum framework described in Sutherland & Schwaber (2013) has been adapted to manage student group projects in a required, on campus, graduate IT Project Management course offered at a North American university. The group project involves the creation of a detailed IT project management plan for an organization. The Scrum framework has been implemented in six phases (see Figure 2). The main elements (i.e., artifacts, processes, roles, and rules) of this implementation are outlined in Table 1. The names of these elements are capitalized and italicized throughout the text. Details of the implementation processes (e.g., the nature of the graduate course, group project description, detailed description of the phases, necessary supplies, artifacts, processes, roles, rules, some instructor reflections, etc.) are provided in the sections below.

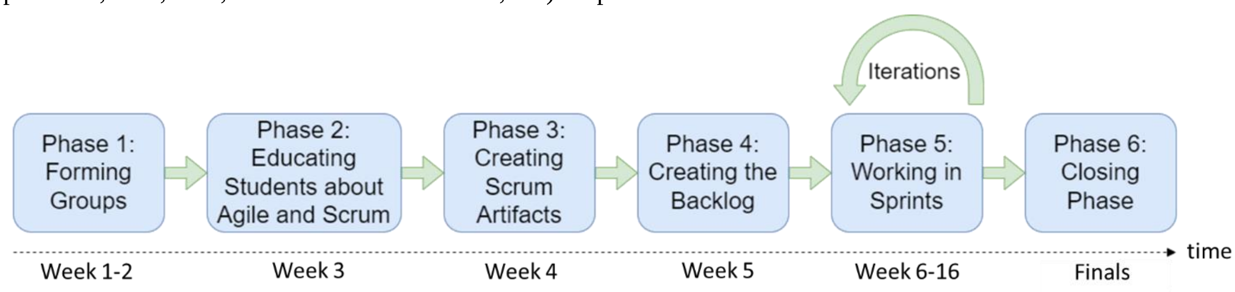


Figure 2. Scrum Implementation Phases (ScrumEductor.org, 2023)

3.1 Information Technology Project Management Course

The Scrum framework has been implemented in a required, graduate, on-campus IT Project Management course. The course is offered by the College of Business at a North American regional university during the spring and fall semesters that span fifteen weeks each. The enrollment in each section is usually close around 20-30 students. This course uses the textbook by Kathy Schwalbe titled "Information Technology Project Management" (6-8th edition). While the textbook touches on the Agile approach to project management and contains brief descriptions of some of the most popular Agile methodologies (e.g., Scrum and Extreme Programming), the textbook adopts a rather traditional, "waterfall" view on project management.

The IT Project Management course is a required course for graduate students majoring in Computer Information Systems (CIS) and Cybersecurity Management (CM). The CIS program is offered by the College of Business. The CM program is offered jointly by the College of Business and the College of Engineering. Other business majors can take this course as an elective, but they rarely do that due to class capacity issues and a false belief that this course may involve programming.

3.2 Group Project Assignment: IT Project Management Plan

The Scrum framework has been used to manage group work involving the creation of a detailed IT project management plan. This group project requires students to create a detailed project management plan involving implementation of an information system in an organization of their choice. Students majoring in CM usually chose to work on a plan involving implementing various networking elements, while CIS majors usually work on project plans involving software development. The project management plan is comprised of 17 Deliverables (see Table 2). The order in which the Deliverables are arranged reflects the dependencies among them.

The Story Points attached to each of the Deliverables are estimated by the instructor. These points reflect the relative workload and grading weight associated with each of the Deliverables. The Final Report involves compiling and summarizing Deliverables 1-15 (see Table 2). The Final Presentation involves presenting the Final Report to the entire class. Many of the Deliverables overlap (i.e., have shared parts). Thus, by the nature of the project, students are forced to work on the plan in an iterative fashion.

Table 1. Main Elements of the Scrum Implementation (ScrumEducator.org, 2023)

Element	Description
Scrum Board	A physical board containing: (1) Group Name; (2) Class Code; (3) *To Do* (also known as the *Backlog*), *Doing*, and *Done* columns. The board is populated with "sticky notes" containing information about *Deliverables*.
Deliverable	A *Deliverable* is a document comprising the group project. A template is provided to students for each *Deliverable*. Students are free to modify these templates. Each *Deliverable* is described using a "sticky note". At the minimum, each note should contain: (1) the name of the *Deliverable*; (2) the value of the *Deliverable* in *Story Points*; and (3) the name of the person responsible for the *Deliverable*. Small "sticky notes" can be used to add additional information about the *Deliverables*.
Story Points	*Story Points* are scale-free estimates of the relative workload required to complete a particular *Deliverabl*e. The story points are also used to "weigh" grades associated with individual *Deliverables*.
Burndown Chart	A *Burndown Chart* shows team work velocity and progress over time. The line on the chart starts at 60 points - the total value of all *Deliverables* comprising the project. When a *Deliverable* is completed (see the *Definition of Done*), the line "goes down". The slope of the line gives information about the "velocity" of the group work, meaning how productive the team is in completing the work. A "flat line" may suggest that not much work is being done and students may miss the final deadlines. A "steep line" with a negative slope may suggest good progress.
Backlog	This is the *To Do* column on the *Scrum Board*. Contains all the group project *Deliverables* that need to be deliverable and the overall plan for completing the project created by each *Group*.
Scrum Master	A person conducting *Weekly Scrums* and making sure each group member has all the tools and information to be productive. For the first few *Sprints*, the role can be performed by the instructor, but will eventually be transferred to a leader of each of the groups.
Product Owner	This is the instructor teaching the course who explains to students what needs to be done for each of the *Deliverables*, provides feedback on the submitted *Deliverables*, and assigns grades.
Group	A permanent team comprised of 4-6 students working together on a group project for the duration of the entire semester.
Sprint	A period of time (usually 1-2 weeks) during which students work on a particular portion of the group project. By default, each *Sprint* is one week long, although it can be longer if a weekly session is cancelled due to holidays or exams.
Weekly Scrum	An in-class group work session comprised of two meetings. The first meeting is among the *Group* members. During this meeting, the *Group* members discuss what has been done and what needs to be done for the next *Sprint*. Based on this discussion, the *Scrum Board* and the *Burndown Chart* are updated. Other issues can be discussed during the *Group* meeting. The second meeting is a stand-up meeting between the group and the instructor. During these meetings students report to the instructor on what has (or has not) been done, the issues they are facing, and the items they commit to work on for the next Sprint.
Definition of Done	A *Deliverable* is done when it is submitted via the Learning Management System (LMS) for grading; only then the *Scrum Board* and the *Burndown Chart* can be updated.
Peer Evaluation	An online survey asking students to provide quantitative and qualitative evaluation of their own contribution to group work together with the contribution of each of their group members. This feedback is used to assign the peer-evaluation grade component for each of the students.

The instructor does not set any deadlines for Deliverables 1-15. But students are informed that (1) they will make their Final Presentations on the last day of the class; (2) the Final Report will be due on the last day of the semester as per the official Academic Calendar of the university; (3) the last day to submit their anonymous online Peer Evaluations (another major grading component for the group project) is the date of the final exam. Thus, delaying the completion of Deliverables 1-15 may delay the final two Deliverables (16-17).

Table 2. Group Project Deliverables (ScrumEducator.org, 2023)

Deliverable	Story Points
1. Team Contract	2
2. Project Organization Chart	1
3. Project Charter	3
4. Business Case	4
5. Statement of Work	4
6. Scope Statement	2
7. Work Breakdown Structure	3
8. Gantt Chart	3
9. Network Diagram	4
10. Cost Estimate	3
11. Stakeholder Registry	1
12. Stakeholder Management Strategy	2
13. List of Risks	1
14. Communication Plan	3
15. Quality Assurance Plan	4
16. Final Presentation	10
17. Final Report	10
Total:	60

3.3 Scrum Implementation Phases

The framework has been implemented for managing group work on the IT Project Management Plan in several phases: (1) Forming Groups; (2) Educating Students about Agile and Scrum; (3) Creating Scrum Artifacts; (4) Creating the Backlog; (5) Working in Sprints (an iterative phase); and (6) Closing Phase. Each of these phases is described in detail below (ScrumEducator.org, 2023).

Phase 1: Forming Groups

The first week of the course is used by the instructor to introduce himself and the course and to cover Chapter 1 from the textbook. During Week 1, students are also briefly informed about the nature of the group project and asked to form groups of 4-6 people. To speed up the group formation, the instructor requires students to do an in-class group case analysis during the second week of the course. The deadline for forming groups is the end of Week 3 of the course. Once students finalized their group membership, group information is entered by the instructor in Canvas – the Learning Management System (LMS) used in this course. This allows the instructor to create collaboration areas for students in Canvas and to assign group project assignments to groups of students.

Phase 2: Educating Students about Agile and Scrum

By the end of Week 3, the instructor covers the first three chapters of the textbook. These chapters give students a fairly good introduction into the field of Project Management. Moreover, Chapter 3 contains an overview of Agile methodologies used in Project Management, including Scrum. This gives the instructor an opportunity to educate students about Scrum and explain how Scrum will be used for managing their group project work. First, the instructor directs students' attention to the "The Scrum Guide" – a 16-page document outlining the essence of the framework (Sutherland & Schwaber, 2013). The instructor tells the students to read the guide by the end of Week 3 as this material will be needed for the group project and will be covered by the midterm exam. In addition, there is a YouTube video tutorial explaining the essence of Scrum and how it is used in software development in under 10 minutes (Axosoft, 2014). This video is shown in class. The content of this video seems to be sufficient to help students understand the most fundamental aspects of the Scrum framework (even if students did not read "The Scrum

Guide"). After showing the video, the instructor explains how Scrum will be adapted and used in class. Basically, Phases 3-6 described further in this section are presented to students. The instructor deliberately does not use any formal manual containing the rules and processes of Scrum implementation in this particular class to allow for room for discussion and innovation.

Phase 3: Creating Scrum Artifacts

During Week 4, the instructor asks students to create all the necessary Scrum artifacts: Scrum Board, Burndown Chart, and "sticky notes" for the Deliverables. This requires acquiring certain supplies prior to this exercise (see Table 3). While the instructor has experimented with several electronic tools and artifacts for implementing Scrum, such as Trello Boards (see Figure 3), the instructor gives preference to physical artifacts (e.g., physical Scrum Board and physical Burndown Chart made from poster boards) to create opportunities for quality, face-to-face interactions in class.

Table 3. Supplies Needed

Item	Quantity	Additional Notes
Poster Boards	1-2 per team	These poster boards will be used for creating *Scrum Boards* – one per team. If an instructor wishes to have a separate board for a *Burndown Chart*, then additional poster boards (one per team) should be acquired. Preferably, the board used for the *Burndown Chart* should contain a square grid. This will make plotting easier.
Markers	1 pack per team	At the minimum, the instructor should have one marker per team. Ideally, each team should have access to a set of markers of various colors. Students usually like to use several colors for creating their *Scrum Boards* and *Burndown Charts*.
Large "Sticky Notes"	1 stack per team	"Sticky notes" (e.g. Post-it® Notes) are to be used by students to capture the information about the *Deliverables* (or "user stories") comprising their *Backlog*. At the minimum, these notes should contain: (1) the name of the *Deliverable*; (2) the value in terms of *Story Points* (can be indicated by putting thick dots using a marker; and (3) the person in charge.
Small "Sticky Notes"	1 stack per team	Small "sticky notes" can be used by students to add additional information to their "user stories" (captured with the help of large "sticky notes").
Mobile Tripods	1 per team	Mobile tripods are used for mounting *Scrum Boards* and *Burndown Charts* during "stand-up" meetings between team members and the instructor.
Rulers	1 per team	Long rulers (depending on the size of the poster boards used) to assist students with drawing lines on their *Scrum Boards* and *Burndown Charts*.
Pencils and Erasers	1 per team	Pencils are used by students to sketch layouts of their *Scrum Boards* and *Burndown Charts*.
Storage Box	1	A box for storing rulers, "sticky notes", markers, pencils, etc. The instructor can give out these supplies during *Weekly Scrum* sessions and put them away once group work is over.
Folders	1 per team	These folders are used for storing physical copies of templates for each of the *Deliverables*. This allows students to have something in front of them when they discuss the details of each of the Deliverables.

To encourage experimentation and innovation, the instructor deliberately gives students only rough guidelines on how these artifacts should look like (e.g., by showing some examples from previous courses). Over the years, students have created several versions of Scum Boards and Burndown Charts of the course of four semesters. One representative example of a Scrum Board and Burndown Chart is provided in Figure 4.

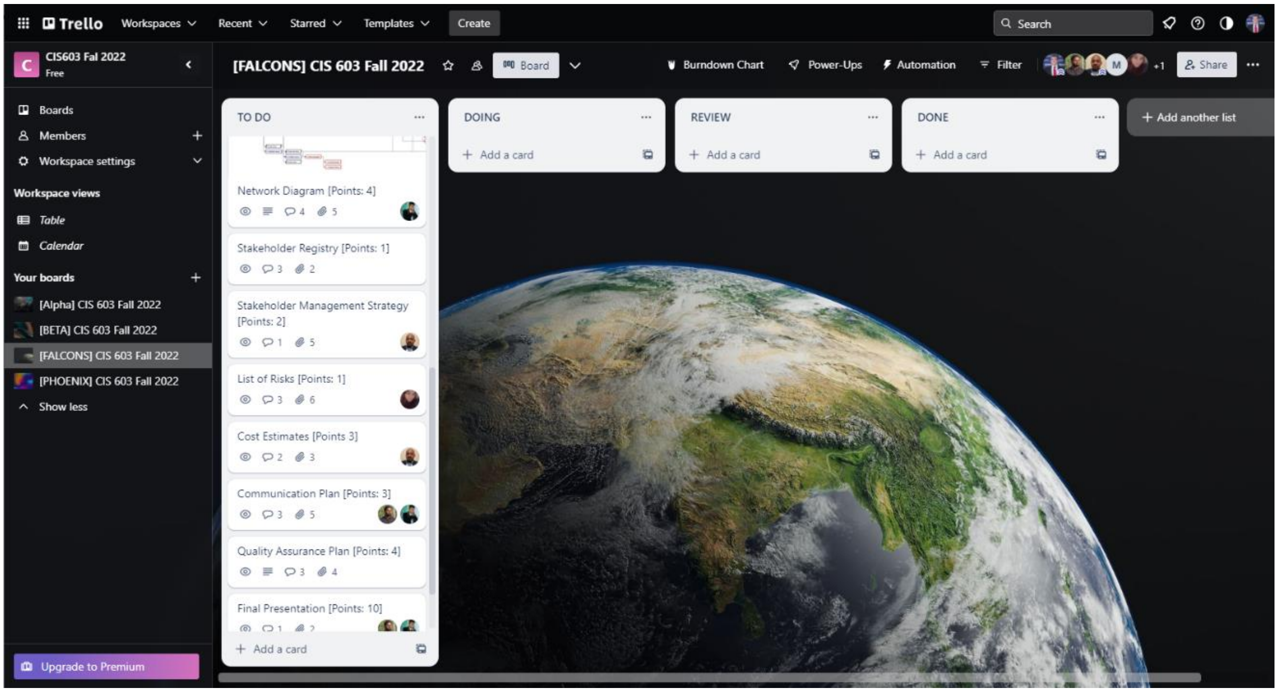


Figure 3. Trello Scrum Board (ScrumEducator.org, 2023)

Eventually, the instructor and students converged on the following version of a combined Scrum Board and Burndown Chart board (see Figure 4).

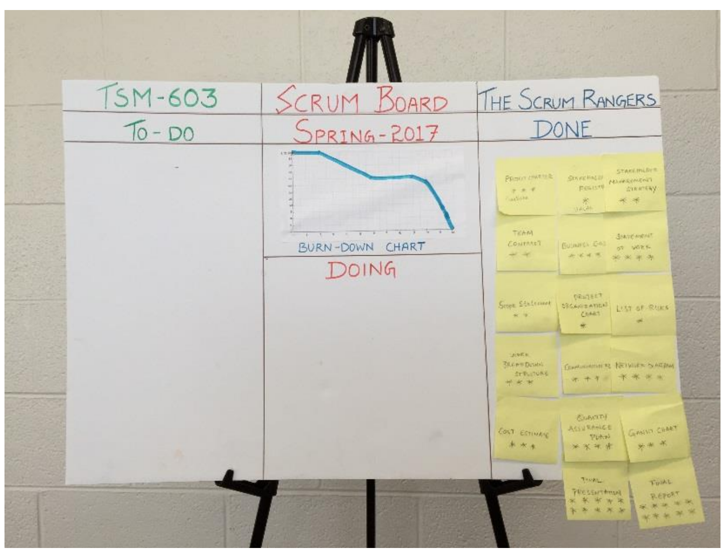


Figure 4. Scrum Board and Burndown Chart Combined (ScrumEducator.org, 2023)

Note that the Scrum Board in Figure 4 is simpler (it has three columns) and also includes a small Burndown Chart in the middle (made from a sheet of graphing paper). This makes it cheaper and simplifies the logistics associated with Weekly Scrums (e.g., only one tripod is needed to mount this board). The Burndown Chart, just like the previous version thereof, contains 60 points reflecting the total weight of all the Deliverables (see Table 2).

Phase 4: Creating the Backlog

To start working on the group project, students need to familiarize themselves with the Deliverables comprising the group project and to create a Backlog. Templates for each of the Deliverables are posted to Canvas at the start of the class. For Week 5 meeting, the instructor should print the templates for each of the team and place them in a folder. This is needed so that each team always has physical copies of the required Deliverables in front of them. Usually, students bring smartphones or laptops to class, so they can also access these templates online via Canvas.

Creating a Backlog requires using a large "sticky note" to capture information (or "user story", as it is called in Scrum) for each of the Deliverables comprising the group project. At the minimum, each of these large notes should contain: (1) the name of the Deliverable; (2) the value in terms of Story Points (can be indicated by putting thick dots using a marker) (3) the person in charge of the Deliverable. The instructor explains to students that several people can collaborate on a particular Deliverable. In fact, this is often necessary as individual Deliverables overlap or rely on several other Deliverables. However, there should be one person in charge of each Deliverable. This person will be responsible for finalizing the Deliverable and submitting it to the instructor via Canvas for grading. Small "sticky notes" can be used to subdivide large Deliverables into sub-parts or to simply add additional notes to each of the Deliverables.

Once a note (or notes) is created for each Deliverable, students need to place them into the "To Do" column of their Scrum Boards in the order in which these Deliverables should be completed. This requires students to develop a shared understanding of what it will take to complete the entire project. The instructor acts as a consultant to make sure students arrange the Deliverables in an order that reflects the actual dependencies among them. For example, students cannot work on their Gantt Chart or Network Diagram unless their Work Breakdown Structure is finalized (see Table 2). Ordering the notes in the To Do column completes the creation of the Backlog.

Once each Group creates a Backlog, the instructor asks students to select a few items to work on for the next week (or Sprint). The "sticky notes" devoted to these items are placed into the "Doing" column of the Scrum Board. Some Groups make very ambitious plans for the first week selecting several labor-intensive items. The instructor can advise them to start with a Project Charter and a Scope Statement. Spending more time on these brief yet important documents can put the entire project on the right track.

Phase 5: Working in Sprints

Starting Week 6, Groups should be ready to start working in Sprints. It is communicated to students that, by default, each Sprint will be one week long. Groups are asked to commit to certain amount of work of each Sprint (week). Sometimes Sprints are longer than 1 week due to class cancelations (e.g., due to holidays or exams). Students seem to prefer longer Sprints, as this gives them more time to work on their Deliverable. Yet the instructor has found that one-week Sprints allow to intervene earlier in case a student is "stuck" or simply fails to do any work.

At the end of each Sprint, the instructor asks each team to have a Weekly Scrum. The Weekly Scrum is comprised of two meetings. First, students are asked to convene with their Group members to discuss what has (or has not) been done the week prior to that and what needs to be done for next week. Before asking students to break into Groups, the instructor gives students overall feedback on the submissions. During the Group meeting, students need to develop a shared understanding of what has been done and what needs to be done. After this, Groups need to update their Scrum Board and Burndown Chart (if some items were actually completed). The second meeting happens closer to the end of the Weekly Scrum. It is a "stand-up" meeting between the Group and the instructor. This requires the entire Group to come forward to the instructor and place their Scrum Board and Burndown Chart in front of the instructor. During the first Sprints, the instructor acts as a Scrum Master and Product Owner. The instructor asks the team the following three questions:

1. What have you done during this Sprint?
2. What issues are you facing in your work?
3. What are you planning to do for the next Sprint?

Going through these questions is usually enough to generate a brief yet useful discussion among the Group members and the instructor. After a few weeks, the instructor can usually identify a "natural leader" within each Group and transfer the responsibilities of a Scrum Master to him or her. This can be done by telling the leader that he or she needs to go through these three questions with the Group in front of the instructor. The instructor can help the student Scrum Master "interrogate" the Group members who do not actively participate in these "strand-up" meetings. It is especially important for the instructor to politely ask those Group members who have not accomplished anything during the Sprint or whether they need any help. This is usually enough to put "peer pressure" on the lagging team members without the Group leader having to do any disciplining in front of everyone. The instructor remains the Product Owner for the rest of the semester, acting as the ultimate authority on what needs to be done for the group project.

In order for a student to say that something has been done during a Sprint, a specific Deliverable needs to be submitted via Canvas. Submission to Canvas constitutes the so-called Definition of Done for this Scrum adaptation. This rule is reiterated and strictly enforced by the instructor throughout the semester. Sometimes students forget the rule and report things as being done without actually submitting them to Canvas. This probably happens due to the fact that these students are embarrassed to admit in front of everyone that they have not done anything during the Sprint.

Sometimes students bring up some grading issues during Weekly Scrums. For example, they may remind the instructor that a particular important Deliverable has not been graded or ask questions about the grade assigned. They may also ask the instructor to clarify some of the feedback that was given to them within the document (the instructor grades each document in Canvas by inserting in-text notes and highlights for each of the documents submitted). The instructor may open these Deliverables on his class computer and discuss those issues with students. The instructor makes sure that all submissions for the week are graded at least one day prior to each class. This gives students some time to look at the feedback and raise meaningful questions during Weekly Scrum meetings.

If students wish to resubmit a particular Deliverable and have it regraded, they need to obtain instructor permission. The instructor grants this permission only if it is important to redo a particular Deliverable to increase the quality of the entire project. Ideally, the instructor wants to give unlimited submission attempts for each of the Deliverables. This will help the students to proceed with their group project in a truly iterative fashion. However, this is simply not possible when a class is large. But students can "resubmit" all of their Deliverables when they submit their Final Report, which is a compilation of all Deliverables submitted during the semester.

At the end of each stand-up meeting with each Group, the instructor gives the team a concluding message to motivate them. For example, if the instructor feels that the team is making good progress, then the instructor will praise the entire Group. If the instructor thinks that the Group is lagging and has noticed certain quality issues with the submitted work (e.g., poor writing style or lack of understanding of some fundamental concepts in project management), the instructor will bring it up and emphasize the need to take some corrective actions from the team. If the class is too big or the instructor simply cannot remember the details of the feedback given to each of the teams, then notes should be taken prior to the class and used as memory aids for formulating this "weekly message" to the team.

Some corrective actions may be required on the instructor side as well. For example, poor writing quality has been an issue with several Groups due to the fact that the majority of students enrolled in the class are international students. The instructor addresses this problem by conducting additional in-class workshops on effective technical writing and proper referencing using APA style.

Phase 6: Closing Phase

After going through around 10 Sprints, teams should be done with the project. As it was mentioned previously, the last week of the semester is devoted to Final Presentations. The Final Report is due at the end of this week as well. The instructor also asks students to complete a simple online survey (created using Google Forms) where they provide quantitative and qualitative Peer Evaluations of their own contribution to the group project and the contributions of their peers. While each student sees his or her overall grade for the peer evaluation component of the group project, the student will not see individual scores assigned by students and the qualitative feedback.

This feedback is used by the instructor to calculate the peer evaluation grade component for each of the students. Typically, the instructor averages out the quantitative scores given to a particular student by each of the Group members. Qualitative feedback is used by the instructor to validate the numeric scores given. For example, if somebody gives a very low score to another student, then it will be explained that the student being evaluated has not contributed much to the group project at all due to poor attendance. The instructor informs the students that he reserves the right to delete "outliers"—scores that are either too high or too low relative to others and are not backed by convincing justifications. So far, the instructor has not observed too many outliers. Students tend to assign fair and consistent grades to their Group members.

Before the final exam, the instructor asks students to update their Burndown Charts to "zero points left" and place all of the notes into the Done column of their Scrum Boards to officially close the project. As a joke, the instructor recommends the students to go and celebrate the completion of the project and the course as a group after the final exam. Some teams follow this advice.

4. Student Feedback

This section discusses the feedback obtained from students in relation to the Scrum implementation in class. First, this section provides an overview of the survey used to capture student perceptions. The survey is provided in its entirety in Appendix A. Second, the section provides an overview of the demographics of survey participants. Third, the qualitative and quantitative feedback obtained from the survey participants is discussed. Detailed quantitative results of the survey are provided in Appendix B.

4.1 Student Survey

An online survey was designed using Google Forms and administered to students to obtain their feedback on the use of Scrum for managing group project work in the course (see Appendix A). The first part of the survey captures student demographics. The second part of the survey asks students to provide qualitative, open-ended feedback on the things they like and do not like about the Scrum implementation in the course. The third part of the survey also asks students to indicate the degree to which they agree or disagree with statements outlining potential benefits associated with using Scrum in class. A total of 32 questions were asked, with each of the questions falling into one of the following four value dimensions: (1) practical knowledge of Scrum, Agile philosophy, and project management in general; (2) group work dynamics; (3) team-based learning; and (4) overall group performance. Five-point Likert scale (from 1 – "Strongly Disagree" to 5 – "Strongly Agree") is used to capture student responses to each of the questions.

4.2 Student Demographics

While the survey was sent to all students who took this course over the period of two years, most of the students who responded were enrolled in the latest course. Responses were obtained from a total of 28 students. Majority of the respondents are males in their 20s majoring in CIS and CS and having some IT experience (see Table 4).

4.3 Qualitative Feedback

Some examples of positive and negative qualitative feedback obtained from students with the help of the survey are provided in Table 5. All meaningful quotes were included. Some quotes were edited to improve grammar, punctuation, and clarity.

4.4 Quantitative Feedback

An aggregated summary statistics for the four value dimensions of using Scrum in the classroom is provided in Table 6. As shown in the table, an overwhelming majority of students agree that using Scrum for managing the class group project helps students to (1) acquire practical knowledge of Scrum, Agile philosophy, and project management in general; (2) enhance positive group work dynamics; (3) improve team-based learning; and (4) improve the overall performance of groups on the group project assigned to them in class. Detailed survey results are provided in Appendix B.

Table 4. Student Demographics

Demographic Variable	Demographic Group	(n)	(%)
Gender	Female	7	25%
	Male	21	75%
Age	21-25	18	64%
	26-30	10	36%
Major	Computer Information Systems (CIS)	14	50%
	General Business	1	4%
	Telecommunications Systems Management (TSM)	13	46%
Years of Professional Experience	None	12	43%
	6 months to less than 1 year	1	4%
	1 year to less than 3 years	7	25%
	3 years to less than 5 years	4	14%
	5 years or more	4	14%
Years of Professional Experience in IT Sector	None	13	46%
	6 months to less than 1 year	2	7%
	1 year to less than 3 years	10	36%
	3 years to less than 5 years	3	11%
	5 years or more	0	0%

Table 5. Qualitative Feedback from Students: Likes and Dislikes

Likes	Dislikes
• Knowing the progress of the project in real time and right from the beginning of the course • Helps in enhancing the quality of the group project • Helped complete the project on time • Shows the flow of work done in specific time • Helps organize the project activities easily • Helps to see the bigger project through smaller tasks • Weekly meetings and incremental approach to group project work was quite effective • It makes accountability and control within the team easier • Good way to visualize work • Makes it easier to make changes to how the project is done • Good way to organize teamwork • Stand-up meetings were effective because everyone in the team knew what to do in the following week	• Sprints should be 2 weeks or more; one-week sprints make things hectic sometimes • Sometimes we forget to update the Scrum Board • The Scrum framework forces you to rely on the entire team. If a team member is sick, then his work has to be reassigned to another team member to keep up with Sprint commitments • It would be more helpful if Scrum Board was online • Giving updates on task might distract from actual work • Too much paper work to do • Project quality is hard to manage unless the team members do some quality assurance at the end of each sprint • Becoming an effective Scrum Master takes time while the project is due fairly soon

4.5 Possible Improvements

While student response to the implementation of Scrum in the course is mostly positive, some improvements may be necessary based on the student feedback and the instructor's own reflections. First, additional online tools (e.g., Trello for creating and managing online Scrum Boards) can be used to improve convenience and accessibility of some of the main Scrum artifacts while also preserving the benefits associated with face-to-face interaction. Perhaps an online board can be broadcasted to the projector screen during Weekly Scrums. Second, grading should be somehow incorporated into the methodology. For example, the Burndown Chart can be modified to include not only "points completed" but also "points earned" towards the grade. Perhaps a modified Earned Value Management (EVM) approach can be used (Schwalbe, 2007). Third, ways of formalizing and automating assessment of individual member's contribution to the group project and "soft skills" necessary to be an effective team member should be researched. All these improvements should be done in a way so that the framework remains "light", consistent with the vision of the creators of Scrum (Sutherland, 2014), and does not increase the "overhead" associated with the implementation of the framework substantially. "Keeping it simple" has been the most important goal of the instructor since the very first time the framework was introduced in his course. This makes the implementation simple and general enough to be used in any Information Systems or general business course involving a group project. Still, additional Scrum tools, techniques, and artifacts can be added to this adaptation of Scrum based on the nature and the goals of the course where Scrum is used.

Table 6. Quantitative Feedback from Students: Value Dimensions of Scrum

Dimension Using Scrum for managing the group project allowed me…	**Mean**	**Standard Deviation**	**"Agree" or "Strongly Agree" (%)**
…acquire practical knowledge of Scrum, Agile philosophy, and project management in general	4.54	0.57	96%
… enhance positive group work dynamics	4.49	0.72	91%
… improve team-based learning	4.53	0.62	93%
… improve the overall performance of groups on the group project	4.45	0.63	93%

5. Conclusion

This paper has presented a practical and accessible model for integrating Scrum into a face-to-face Information Systems Project Management course, offering a low-tech yet effective approach to managing student group work. By applying Scrum to the development of traditional project management artifacts—rather than software deliverables—this adaptation bridges the gap between Agile principles and "waterfall-based" learning objectives in a typical Project Management class. The Scrum adaptation model consists of six clearly defined phases, with each phase containing detailed implementation instructions and the instructor's recommendations based on his experience implementing Scrum in the course. The implementation is supported by simple tools, such as physical or electronic Scrum boards, and processes, such as peer-evaluations, making it applicable across a wide range of educational settings. Empirical feedback from students suggests that this approach not only enhances engagement and accountability but also supports the development of both Agile competencies and teamwork skills. For instructors, the framework offers a straightforward method for monitoring progress, promoting team collaboration, and aligning classroom practices with real-world project dynamics. Future research may explore the adaptation of this model to online or hybrid settings and investigate long-term impacts on student performance and soft skill development. As the demand for Agile-ready graduates continues to grow, this classroom-tested implementation of Scrum provides a meaningful and effective pathway for educators to prepare students for today's collaborative, fast-paced work environments across diverse organizations and industries.

6. References

Axosoft. (2014). *Scrum in under 10 minutes* [Video]. YouTube. https://www.youtube.com/watch?v=XU0llRltyFM

Baham, C. (2020). Improving business product owner commitment in student scrum projects. *Journal of Information Technology Education: Research, 19*, 243–258. https://doi.org/10.28945/4549

Babik, D. (2022). Teaching Tip: Scrum boot camp: Introducing students to Agile system development. *Journal of Information Systems Education*, *33*(3), 195-208.

eduScrum. (2020). *The eduScrum guide: The rules of the game*. https://www.scribd.com/document/551896600/The-eduScrum-guide-English-2-0-update-21-12-2019

Dong, L. (2023). Strengthening sustainability in Agile Education: Using client-sponsored projects to cultivate Agile talents. *Sustainability*, *15*(11), 8598.

Fernandes, S., Dinis-Carvalho, J., & Ferreira-Oliveira, A. T. (2021). Improving the performance of student teams in project-based learning with scrum. *Education Sciences, 11*(8), 444. https://doi.org/10.3390/educsci11080444

Highsmith, J., & Cockburn, A. (2001). Agile software development: The business of innovation. *Computer, 34*(9), 120–127. https://doi.org/10.1109/2.947100

Krehbiel, T. C., Salzarulo, P. A., Cosmah, M. L., Forren, J., Gannod, G., Havelka, D., & Merhout, J. (2017). Agile manifesto for teaching and learning. *Journal of Effective Teaching, 17*(2), 90–111.

López-Alcarria, A., Olivares-Vicente, A., & Poza-Vilches, F. (2019). A systematic review of the use of agile methodologies in education to foster sustainability competencies. *Sustainability, 11*(10), 2915. https://doi.org/10.3390/su11102915

Mahalakshmi, M., & Sundararajan, M. (2015, February). Tracking the student's performance in web-based education using Scrum methodology. In *Computing and Communications Technologies (ICCCT), 2015 International Conference on* (pp. 379–382). IEEE. https://doi.org/10.1109/ICCCT2.2015.7292761

Manamendra, M. A. S. C., Manathunga, K. N., Perera, K. H. D., & Kodagoda, N. (2013, April). Improvements for agile manifesto and make agile applicable for undergraduate research projects. In *Computer Science & Education (ICCSE), 2013 8th International Conference on* (pp. 539–544). IEEE. https://doi.org/10.1109/ICCSE.2013.6554019

Matkovic, P., Tumbas, P., & Vukovic, V. (2016). Agile development processes and software architecture in study programmes of higher education institutions. In *Proceedings of ICERI2016 Conference* (pp. 3092–3100).

May, J., York, J., & Lending, D. (2016). Play ball: Bringing Scrum into the classroom. *Journal of Information Systems Education*, *27*(2), 85–98.

Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Scrum project. *Information and Software Technology, 52*(5), 480–491. https://doi.org/10.1016/j.infsof.2009.11.004

Morales-Trujillo, M. E., Galster, M., Gilson, F., & Mathews, M. (2021). A three-year study on peer evaluation in a software engineering project course. *IEEE Transactions on Education, 65*(3), 409–418. https://doi.org/10.1109/TE.2021.3117032

Mutchler, L. A., Connolly, A. J., & Rush, D. E. (2024). Teaching tip: Leveraging scrum to increase student engagement and skill building in information security management. *Journal of Information Systems Education*, *35*(4), 422-437.

Nejmeh, B., & Weaver, D. S. (2014, October). Leveraging Scrum principles in collaborative, interdisciplinary service-learning project courses. In *2014 IEEE Frontiers in Education Conference* (pp. 1–6). IEEE. https://doi.org/10.1109/FIE.2014.7044087

Pope-Ruark, R. (2012). We Scrum every day: Using Scrum project management framework for group projects. *College Teaching, 60*(4), 164–169. https://doi.org/10.1080/87567555.2012.660761

Reynolds, S., Caldwell, A., Procko, T., & Ochoa, O. (2023, October). Scrum in the classroom: An implementation guide. In *2023 IEEE Frontiers in Education Conference* (pp. 01–08). IEEE. https://doi.org/10.1109/FIE58773.2023.10339309

Rigby, D. K., Sutherland, J., & Takeuchi, H. (2016). Embracing Agile. *Harvard Business Review, 94*(5), 40–50.

Rush, D. E., & Connolly, A. J. (2020). An agile framework for teaching with Scrum in the IT project management classroom. *Journal of Information Systems Education, 31*(3), 176–187.

Salza, P., Musmarra, P., & Ferrucci, F. (2019). Agile methodologies in education: A review. In D. Parsons, & K. MacCallum (Eds.), *Agile and lean concepts for teaching and learning: Bringing methodologies from industry to the classroom* (1st ed., pp. 25-45). Springer.

ScrumEducator.org. (2023). *Using Scrum in Education: A Definitive Guide*. https://scrumeducator.org/wp-content/uploads/2023/07/Scrum_in_Education_V01_07122023-.pdf

Schwalbe, K. (2007). *Information technology project management* (6th ed.). Boston, MA: Cengage Learning.

Sharp, J. H., Mitchell, A., & Lang, G. (2020). Agile teaching and learning in information systems education: An analysis and categorization of literature. *Journal of Information Systems Education, 31*(4), 269–281.

Sutherland, J. (2014). *Scrum: The art of doing twice the work in half the time*. New York, NY: Random House Business Books.

Sutherland, J., & Schwaber, K. (2013). *The Scrum Guide*. Scrum.org. https://scrumguides.org

Thouin, M. F., & Hefley, W. E. (2024). Teaching tip: Teaching scrum product owner competencies using an experiential learning simulation. *Journal of Information Systems Education*, *35*(1), 37-47.

VersionOne. (2023). *17th annual state of agile report*. https://stateofagile.com/

Vogelzang, J., Admiraal, W. F., & van Driel, J. H. (2021). Scrum methodology in context-based secondary chemistry classes: Effects on students' achievement and perceptions. *Instructional Science, 49*(5), 719–746. https://doi.org/10.1007/s11251-021-09553-1

Wagh, R. (2012). Using Scrum for software engineering class projects. In *AGILE India 2012* (pp. 68–71). IEEE. https://doi.org/10.1109/AGILEIndia.2012.13

Wolff, E., & van Ryneveld, L. (2024). The effective management of remote higher education curriculum design teams using Scrum. *Revista de Gestão Social e Ambiental, 18*(9), Article e03291. https://doi.org/10.24857/rgsa.v18n9.2024.3291

Zahorodko, P. V. (2023). Overview of Agile frameworks in computer science education. *Educational Dimension, 9*, 206–214. https://doi.org/10.31812/educdim.v9i3.937

Zorzo, S. D., de Ponte, L., & Lucredio, D. (2013, October). Using Scrum to teach software engineering: A case study. In *2013 IEEE Frontiers in Education Conference* (pp. 455–461). IEEE. https://doi.org/10.1109/FIE.2013.6684899

Appendix A: Online Student Survey

Scrum Framework Survey

Dear Participant,

This study explores the effectiveness of using Scum Framework for managing class group work. This research can potentially benefit various educational institutions by equipping them with a theoretical and practical understanding on how to use Scrum for improving group work and team-based learning in the classroom.

Completing this survey should take approximately 15 minutes of your time.

Your participation in this survey is absolutely voluntary. You are free to choose not to participate in the survey or discontinue the survey at any time. Your responses will be anonymous. The information that you provide through this survey will not be linked to your identity. Because this survey is anonymous, the researcher reserve the right for a variety of ethical uses of this data at their own discretion. The use may involve sharing this data with other researchers or the public.

Please feel free to contact [deleted] if you have any questions or comments about this survey. Thank you!

Personal Information

What is your gender?

How old are you?

Which of the following describes you best?

Are you an online student?

What is your major?

How many years of professional experience do you have?

How many years of professional experience in the IT sector do you have?

General Comments

In general, what did you like about using the Scrum Framework for managing your group project? How did the framework help you with your group project?

In general, what are the things that you DID NOT like about using the Scrum Framework for managing your group project work? Are there any specific elements of this framework that you found to be of little value or distracting?

What suggestions do you have for the instructor on how the Scrum Framework should be used in class? What are some of the specific ways in which the instructor can improve the effectiveness of the framework for managing class group projects?

Project Management Knowledge Dimension (D1)

Using the Scrum Framework for managing our group project work helped me(us)...

D1Q1 [...Learn the fundamental principles of Agile project management philosophy]

D1Q2 [...Learn about the Scrum methodology]

D1Q3 [...Gain a practical understanding of the Scrum methodology]

D1Q4 [...Learn the main roles under the Scrum framework]

D1Q5 [...Learn about the Scrum artifacts and how to use them]

D1Q6 [...Learn Scrum processes and practices]

D1Q7 [...Learn about the concepts, ideas, practices, documents, and deliverables used in project management]

D1Q8 [...Gain a practical understanding on how to manage projects]

D1Q9 [...Gain practical experience working on projects]

Group Work Dimension (D2)

Using the Scrum Framework for managing our group project work helped me(us)...

D2Q1 [...Learn how to lead teams]

D2Q2 [...Learn how to work as a part of a group]

D2Q3 [...Learn how to facilitate group work]

D2Q4 [...Improve communication among the group members]

D2Q5 [...Improve communication between the instructor and the team]

D2Q6 [...Reduce conflicts among the team members]

D2Q7 [...Reach consensus among the team members on various issues related to the group project]

D2Q8 [...Improve collaboration among the group members]

D2Q9 [...Increase involvement of team members into the group work]

D2Q10 [...Solve problems related to the group project]

D2Q11 [...Become more organized with our group work]

Team-Based Learning Dimension (D3)

Using the Scrum Framework for managing our group project work helped me(us)...

D3Q1 [...Learn from other team members]

D3Q2 [...Get timely feedback on my understanding of various project management concepts, documents, and practices from my team members]

D3Q3 [...Get timely feedback on my understanding of various project management concepts, documents, and practices from the instructor]

D3Q4 [...To understand various project management concepts, documents, deliverable, and practices]

D3Q5 [...Reach a mutually shared understanding of various project management concepts, documents, deliverables, and practices]

Group Project Performance Dimension (D4)

Using the Scrum Framework for managing our group project work helped me(us)...

D4Q1 [...Complete group project tasks in shorter time]

D4Q2 [...Avoid mistakes in relation to the group project]

D4Q3 [...Improve the quality of group project deliverables]

D4Q4 [...Increase our grade for the group project]

D4Q5 [...Enjoy working in the group]

D4Q6 [...Enjoy the class]

D4Q7 [...Benefit from the class]

Appendix B: Detailed Survey Results

Dimension 1 (D1)	**Mean**	**Median**	**SD**	**% of 5**	**% of 4**
D1Q1	5	5	0.638285	57%	36%
D1Q2	5	5	0.57275	61%	36%
D1Q3	5	5	0.57275	61%	36%
D1Q4	5	5	0.507875	54%	46%
D1Q5	4	5	0.685257	50%	39%
D1Q6	4	5	0.576204	50%	46%
D1Q7	5	5	0.417855	79%	21%
D1Q8	5	5	0.503953	57%	43%
D1Q9	4	4	0.57275	46%	50%
Dimension 1 Overall	**5**	**5**	**0.566947**	**57%**	**39%**
Dimension 2 (D2)	**Mean**	**Median**	**SD**	**% of 5**	**% of 4**
D2Q1	4	5	0.772374	50%	32%
D2Q2	5	5	0.693889	61%	29%
D2Q3	5	5	0.692935	64%	25%
D2Q4	5	5	0.566947	64%	32%
D2Q5	5	5	0.475595	68%	32%
D2Q6	4	5	0.854493	50%	32%
D2Q7	4	5	0.780042	50%	39%
D2Q8	5	5	0.57275	61%	36%
D2Q9	4	5	0.999338	68%	21%
D2Q10	4	5	0.685257	50%	39%
D2Q11	5	5	0.611832	75%	18%
Dimension 2 Overall	**4**	**5**	**0.715056**	**60%**	**31%**
Dimension 3 (D3)	**Mean**	**Median**	**SD**	**% of 5**	**% of 4**
D3Q1	4	5	0.780042	54%	29%
D3Q2	4	5	0.576204	50%	46%
D3Q3	5	5	0.48795	64%	36%
D3Q4	5	5	0.637248	61%	32%
D3Q5	5	5	0.558721	68%	29%
Dimension 3 Overall	**5**	**5**	**0.616825**	**59%**	**34%**
Dimension 4 (D4)	**Mean**	**Median**	**SD**	**% of 5**	**% of 4**
D4Q1	4	4	0.645497	**36%**	**54%**
D4Q2	4	4	0.662687	25%	57%
D4Q3	4	4	0.558721	39%	57%
D4Q4	5	5	0.566947	64%	32%
D4Q5	5	5	0.693889	61%	29%
D4Q6	5	5	0.547964	71%	25%
D4Q7	5	5	0.460044	71%	29%
Dimension 4 Overall	**4**	**5**	**0.6267**	**53%**	**40%**

Author Biographies

Dr. Vlad Krotov received his PhD in Management Information Systems from the Department of Decision and Information Sciences, University of Houston (USA). Currently, he is a Professor of Information Systems at Murray State University and a consultant at Accreditation.Biz - an international accreditation consulting company for business schools. He is also a founder of ScrumEducator.org – a free resource for training students, educators, and professionals in essential soft skills. His research and teaching interests are at the intersection of business and technology and fall under such topics as business analytics, strategic IT management, project management, innovation, and ethical aspects of AI.

Dr. Pitzel Krotova is the Director of Institutional Assessment at Murray State University (Kentucky, USA). She earned her EdD in P-20 and Community Leadership with a specialization in Postsecondary Leadership from Murray State University. She has fifteen years of international experience in higher education, specifically in academic administration and institutional effectiveness. Prior to joining Murray State University, she was the College of Business's Academic Quality Assurance Coordinator at Abu Dhabi University in the UAE. Her research interests are related to assurance of learning, assessment, educational leadership, program review, accreditation, and curriculum alignment.

This page intentionally left blank.

Journal of the Midwest Association for Information Systems

Volume 2025 | Issue 2 Article 3

Date: 07-01-2025

Leveraging Synthetic Data from Generative Models for Snow Detection in Data-Scarce Environments

Ricardo de Deijn
Minnesota State University, Mankato, ricardodedeijn@gmail.com

Rajeev Bukralia
Minnesota State University, Mankato, rajeev.bukralia@mnsu.edu

Abstract

Data scarcity poses a significant challenge for training robust machine learning models in safety-critical applications like snow detection, where real-world data collection is often limited and seasonal. This study explores the potential of synthetic data sets generated by prompt-based image synthesis models to enhance machine learning applications in such data-scarce environments. Using OpenAI's DALL·E 3 and xAI's Aurora, synthetic images of snowy and clear sidewalks were compared against a real-world data set for training image-classification models. The findings reveal that an Aurora-based model achieved the highest F2 scores, excelling in snow detection because of its high photorealism and contextual relevance. However, the real-world data set demonstrated greater accuracy in detecting clear sidewalks, resulting in fewer overall classification errors. These results highlight the potential of synthetic data to supplement real-world data sets, particularly in data-scarce domains, while also emphasizing that real-world data remains crucial for balanced classification. This research underscores the necessity for advancements in generative models to more effectively capture complex environmental conditions and improve the generalizability of AI-generated data sets for scalable and practical machine learning applications.

Keywords: prompt-based image synthesis; synthetic images; snow classification; real-world data

DOI: 10.17705/3jmwa.000095

1. Introduction

Since the increased popularity of generative artificial intelligence (Gen AI) started with the introduction of OpenAI's ChatGPT-3.5 in late November 2022 and following models like ChatGPT-4o, xAI's Grok, Meta's Llama, and Google's Bard / Gemini, numerous fields have been largely impacted in their daily processes. These large Gen AI models have been shown to be powerful in creating text similar to how a human would write these (Bandi et al., 2023; Chen et al., 2023). This phenomenon resulted in a large increase in work productivity, as they are now widely used as productivity enhancers (Haan, 2023; Shaji George et al., 2023; Valeriya et al., 2024) helping employees optimize and automate parts of (repetitive) processes or improving decision-making (Dwivedi et al., 2023; Shaji George et al., 2023; Valeriya et al., 2024). The increase in accuracy that led to the rise of these generative artificial intelligence models can be attributed to numerous factors, such as the invention of how to train using parallelization with large numbers of graphics processing units (GPUs) (Atallah et al., 2023; Bandi et al., 2023; Vaswani et al., 2017), the increase in computing power that a GPU can output within a certain timeframe (Bandi et al., 2023). The increased computing efficiency of the models (Bandi et al., 2023), and the increased high-quality internet data sets that could be used for training (Atallah et al., 2023; Bandi et al., 2023). But similar to these text generators, a surge of high-resolution image Gen AI models was observed in the same time frame, with models such as OpenAI's DALL·E, xAI's Aurora, and Google's Imagen (Meng et al., 2023; Zhou et al., 2023). Showcasing an increase in image synthesis capabilities, trying to close the gap between generated images and real-world pictures (Bandi et al., 2023; Meng et al., 2023; Zhou et al., 2023).

Although applications for text-based Gen AIs are largely being developed by businesses and academia, potential applications for image synthesis models are opening as well. This study explores using synthetic data generated by these models to overcome data scarcity, which is a common issue in many domain-specific applications. If successful, this approach could be applied to other scenarios where data is hard to collect. For example, detecting rare events (like certain weather phenomena or accidents) could similarly benefit from generative data (Chang et al., 2020; Elfeki et al., 2017). When an image synthesis model can accurately mimic the data required to train artificial intelligence (AI) models in these data-scarce sectors, the challenges of imbalanced or small data sets will shrink and increase the potential of model convergence, improving its results (D.-C. Li et al., 2017). Current state-of-the-art (SOTA) models like DALL·E and Aurora show a possibility to accurately generate images using textual descriptions that the generator tries to simulate with very minimal costs. It shows that if these models create images indistinguishable from real-world pictures, training data sets can be expanded significantly for costs far cheaper than manual data collection and annotation (Arif & Mahalanobis, 2021; Barbosa et al., 2018).

This study investigates the effectiveness of two prominent prompt-based synthetic image generators, DALL·E 3 and Aurora, in creating training data for a snowy sidewalk image classification task, benchmarking their performance against real-world images. Employing an image classification architecture from de Deijn (2024), the research evaluates three data sets: 500 real-world images of snowy and clear sidewalks, and two synthetic data sets of 500 images each, generated by OpenAI's DALL·E 3 and xAI's Aurora. The choice of DALL·E 3 and Aurora over other popular models like Midjourney or Stable Diffusion was deliberate: DALL·E 3 was selected for its advanced prompt fidelity and ability to generate highly realistic and consistent snowy and clear sidewalk scenes, while Aurora was chosen for its capacity to create diverse, contextually relevant synthetic images, aligning with the study's focus on viable real-world data substitutes. The primary aim is to assess how well models trained on synthetic data can differentiate snow and clear sidewalks in real-world conditions compared to a model trained on real-world images, with all models evaluated on unseen data. We hypothesize that synthetic image generators, such as DALL·E 3 and Aurora, do not yet produce data sets sufficient to match the performance of real-world data in this task, while exploring the alternative possibility that these synthetic data sets are on par with real-world data in terms of model training objectives. This investigation examines the potential of synthetic data sets as practical alternatives to real-world data for training machine learning models, particularly in cases like snow detection, where gathering extensive real-world data sets is often impractical. Performance is assessed using multiple metrics, including precision, recall, accuracy, Area Under the Receiver Operating Characteristic Curve (ROC AUC), and the F2-score, which prioritizes recall over precision to emphasize the importance of detecting snow for safety-critical applications. These comprehensive metrics provide a robust basis for comparing the models' performance in this context and testing the proposed hypothesis.

The remainder of this paper is organized as follows. Section 2 reviews related work, tracing the historical progress of generative AI from generative adversarial networks (GANs) to diffusion models and discussing their applications in synthetic image generation. Section 3 details the methods, including the data set creation process, model training, and evaluation metrics used to compare synthetic and real-world data. Section 4 presents the experimental results, analyzing the performance of models trained on DALL·E 3, Aurora, and real-world data sets, with statistical comparisons to test the hypotheses. Section 5 concludes with key findings, highlighting the potential and limitations of synthetic data for snow detection. Section 6 discusses challenges encountered, such as prompt dependency and data set diversity, and suggests directions for future research. Appendices provide additional details, including confusion matrices for model performance.

2. Related Work

Generative artificial intelligence (Gen AI) is a concept referring to deep-learning models that can generate new-like, meaningful content. This content can include multiple modalities, such as text, images, audio, and video (Bandi et al., 2023; Feuerriegel et al., 2024). These models are designed to mimic human creativity by learning patterns and structures from large data sets, enabling them to create outputs that resemble real-world data or entirely imaginative constructs (Creswell et al., 2018; Donahue et al., 2016; Dumoulin et al., 2016; Feuerriegel et al., 2024; Goodfellow et al., 2014; Mescheder et al., 2017; Mirza & Osindero, 2014; Nichol & Dhariwal, 2021; Ramesh et al., 2021; Sohl-Dickstein et al., 2015; Zhou et al., 2023). Gen AI has seen widespread adoption in various domains, such as automated content creation, image synthesis, natural language processing, music composition, and video generation (Bandi et al., 2023; Zhang et al., 2023).

Understanding the historical progress of Gen AI is essential to appreciating its current capabilities and challenges. The development of key frameworks, from foundational generative adversarial networks (GANs) to more recent diffusion models, reveals the iterative process through which researchers have enhanced the quality, stability, and applicability of generative models. This historical perspective not only highlights the technical milestones achieved but also underscores the limitations that continue to drive innovation in this space and where we are right now.

2.1 Historical Progress of Generative AI

Generative AI has been in development for multiple decades, but its significance surged with the introduction of generative adversarial networks (GANs) (Bandi et al., 2023; de Deijn et al., 2024; K. Wang et al., 2017), as originally proposed by Goodfellow et al. (2014). GANs create new synthetic data using an example training data set built of real-world or authentic data and minimax two-player game concepts using fully connected neural networks. This concept assumes that there are two components: component A, a generator block, and component B, a discriminator block. The generator block's mission is to generate new data that mimics the distribution of the original real-world data, while the discriminator block tries to distinguish whether the data shown is from the original data set or from the generator. As each component wants its score to be as good as possible, it tries to outperform the other component during its training. This adversarial process should *theoretically* lead to a point in which neither component can improve its score, as the data achieves a high level of realism that the discriminator can't identify synthetic data from authentic data (Creswell et al., 2018; Goodfellow et al., 2014; K. Wang et al., 2017).

Early implementations of GANs performed well on simple data sets such as MNIST and CIFAR-10 but struggled with more complex scenes due to limited understanding of image semantics and spatial relationships (Creswell et al., 2018; de Deijn et al., 2024; S. Wang et al., 2023). To address this, researchers incorporated latent space inference, enabling the generator to capture abstract representations of data. Notable examples include Adversarially Learned Inference (ALI) and Bidirectional GANs (BiGANs), which added encoders to transform input images into latent feature vectors, improving the generator's understanding of image structure and context (Donahue et al., 2016; Dumoulin et al., 2016). This helped guide generation and discrimination more meaningfully than relying on noise vectors alone.

Despite these improvements, GANs continued to exhibit issues such as training instability and mode collapse, especially with high-dimensional images. To counter these limitations, researchers explored probabilistic modeling approaches like variational auto-encoders (VAEs), which aim to encode input data into a latent distribution rather than a fixed vector. VAEs optimize a loss function combining a reconstruction term (which measures how well the model can reconstruct the input) and a Kullback-Leibler (KL) divergence term (which ensures that the learned latent space remains close to a prior distribution, typically Gaussian) (Kingma & Welling, 2013). While VAEs improve training stability, they often yield blurry outputs due to their emphasis on reconstruction fidelity rather than adversarial realism.

To bridge the strengths of both models, Adversarial Variational Bayes (AVB) was introduced as a hybrid framework that integrates the probabilistic inference of VAEs with the generative sharpness of GANs. AVB replaces the traditional VAE encoder with one trained adversarially to match a posterior distribution, thereby preserving sample quality while improving stability Mescheder et al. (2017). This hybrid approach reduces the tension between quality and training convergence and marks a key moment in Gen AI's trajectory. This showed that integrating probabilistic reasoning into adversarial frameworks can lead to more robust generative models.

However, even these advanced models faced limitations in capturing fine-grained visual complexity and maintaining training stability at scale. This led to the rise of diffusion models, which offered a fundamentally different approach. Rather than pitting two networks against each other, diffusion models learn to gradually corrupt input data with noise and then reconstruct it step-by-step in reverse. The process is optimized using reconstruction loss and KL divergence terms similar to VAEs, but avoids adversarial training entirely. This method forms the basis for many recent high-fidelity image generators and sets the stage for the diffusion-based models explored in the next section (Ho et al., 2020; Nichol & Dhariwal, 2021).

2.2 Diffusion Models

A diffusion model often uses UNet UBlocks (Ho et al., 2020; Saharia et al., 2022), ResNet blocks (Nichol & Dhariwal, 2021), as well as down sample and attention blocks such as self-attention (Ho et al., 2020; Saharia et al., 2022), multi-head attention (Nichol & Dhariwal, 2021), or cross attention (Saharia et al., 2022) to forward diffuse images with noise. To reverse diffuse, a model uses a combination of up sample blocks, skip connections to retain information across layers, and more attention blocks. By applying attention to multiple depths, such as 64x64, 32x32, 16x16, and 8x8, the model can better understand spatial relationships in early layers while recognizing global structures on deeper layers (Ho et al., 2020; Nichol & Dhariwal, 2021; Saharia et al., 2022).

Shortly after the introduction of ChatGPT by OpenAI in 2022, large language model (LLM) applications became popular and found their way into diffusion models (Saharia et al., 2022; Zhang et al., 2023). LLMs showed that models can mimic real-life conversations, with near-accurate responses to questions and comments from users. As a result of the transformative research on LLMs such as GPT (Radford et al., 2018), BERT (Devlin et al., 2018), and T5 (Raffel et al., 2020), large text-to-image synthesis models were developed that utilized LLMs to create images (Ramesh et al., 2021; Saharia et al., 2022; Zhang et al., 2023). These diffusion models take a textual prompt and transform it into vector embeddings through pre-trained embedders such as Contrastive Language-Image Pre-Training (CLIP). Adding these text embedders allows the diffusion model to create images for many classes, as the text informs the model during both the forward and reverse diffusion sections on what it tries to create (Radford et al., 2021; Ramesh et al., 2022; Saharia et al., 2022). Using large data sets of web-scraped data with both images and image descriptions, large diffusion models, such as Imagen (Saharia et al., 2022), Stable Diffusion (Rombach et al., 2021), and DALL·E (Ramesh et al., 2021) were created with the ability to handle both high-end input prompts and to create high-quality photorealistic images within seconds (Ramesh et al., 2021, 2022; Saharia et al., 2022).

However, current research shows that a lot of current generations are more likely to look like artwork instead of real-world images (Hao et al., 2023; Hossain et al., 2024; Zhang et al., 2023). This is related to the limited availability of realistic annotated data sets and the constraints of model sizes. Additionally, image quality is further influenced by prompt engineering, a process of crafting input text to optimize the output of generative AI models, as well as by model architecture, fine-tuning processes, and prompt formulation (Zhang et al., 2023). Because of the way textual embeddings are integrated into the models' architecture, text-to-image models are directly dependent on the quality of the input prompt to generate better results. With well-engineered prompts, a model can emulate real-world situations, possibly even functioning as an emulator for real-world training data in scarce domains. But the opposite is true as well; if a prompt is not well-defined and of low quality, results can be nothing like the desired training set (Hao et al., 2023; Zhou et al., 2023). Next to that, there is also a risk that the model would like the prompt to be received in a certain way and performs worse when it sees a word that it has not been trained on. This creates different challenges, as it is often hard to know which words a model prefers to see and which words receive preferred results for the user (Hao et al., 2023; Zhang et al., 2023). For this reason, approaches such as Diffusion Inversion (Zhou et al., 2023) or prompt adaptation frameworks (Hao et al., 2023) are proposed to overcome these challenges. These considerations emphasize the importance of addressing prompt dependency and data set limitations to advance the generation of more realistic and contextually accurate AI outputs, setting the stage for further exploration in this study.

Although diffusion models have become the dominant architecture for high-resolution and prompt-driven image generation due to their stability and effective noise-based reconstruction, they are not the only approach shaping the field of generative modeling. Another important category, autoregressive models, is more upcoming and have shown strong performance by generating images in a sequential manner, predicting one pixel or image patch at a time based on prior context. This approach differs fundamentally from the denoising process used in diffusion models, emphasizing local coherence and detail through ordered prediction. While often more computationally intensive, autoregressive models excel in structured generation tasks, particularly when combined with language modeling techniques. The next section introduces these models and examines their role in advancing the realism and utility of synthetic image generation.

2.3 Autoregressive Models

Autoregressive (AR) models represent distinctive approaches to generative modeling, diverging fundamentally from adversarial and diffusion-based methods. Unlike diffusion models, which iteratively refine noisy data, or generative adversarial networks (GANs) that pit a generator against a discriminator to learn image distributions, AR models generate images sequentially. They model the conditional probability of each pixel or image patch based on all previously generated content. This sequential process ensures that each step is informed by prior steps, fostering strong local coherence and enabling highly structured outputs (Gu et al., 2025). The autoregressive framework excels at capturing intricate dependencies within an image, making it particularly effective for generating detailed and contextually consistent visuals. However, this approach has historically faced challenges, including slow generation speeds and difficulties in modeling both local textures and global structures simultaneously, which have limited its scalability for high-resolution image synthesis (Gu et al., 2025).

Recent advancements have significantly mitigated these limitations, pushing AR models toward greater efficiency and quality. One notable innovation is the Multi-scale Vector-Autoregressive (M-VAR) framework, which introduces scale-separated modeling to decompose image generation into intra-scale and inter-scale processes (Ren et al., 2024). Intra-scale modeling focuses on capturing fine-grained details within a given resolution, while inter-scale modeling addresses long-range dependencies across different resolutions. This decomposition allows M-VAR to efficiently model complex global structures without sacrificing high-fidelity textures, resulting in substantial performance improvements on large-scale benchmarks like ImageNet-256. By enabling faster and more robust generation, M-VAR demonstrates the potential of AR models to compete with diffusion-based methods in high-resolution image synthesis, particularly for applications requiring both local precision and global coherence (Ren et al., 2024).

In addition to improvements in generation speed and flexibility, AR models have evolved to support continuous data representations, moving beyond the limitations of vector-quantized image tokens. Traditional token-based approaches often introduce quantization artifacts, which can degrade image quality, especially for high-resolution outputs. The Multimodal Autoregressive (MAR) model addresses this by operating in continuous latent spaces and applying autoregressive loss functions informed by diffusion-based likelihoods (T. Li et al., 2024). This integration reduces tokenization overhead while preserving the sequential modeling benefits of AR frameworks, resulting in sharper and more natural images. By leveraging continuous representations, MAR achieves a balance between computational efficiency and visual fidelity, making it a compelling choice for applications where artifact-free generation is critical.

Moreover, advances in multimodal alignment have enhanced the reasoning capabilities of AR models, enabling them to generate images that align closely with complex input prompts. Techniques such as chain-of-thought prompting and preference alignment allow models to perform stepwise reasoning during generation, improving both the interpretability and accuracy of outputs (Guo et al., 2025). For instance, when conditioned on detailed textual descriptions, these models can produce images that reflect nuanced semantic relationships, such as specific object placements or environmental conditions. This capability is particularly advantageous for tasks requiring precise control over generated content, as it allows AR models to translate abstract instructions into visually coherent results. Such advancements underscore the potential of AR models to bridge the gap between generative modeling and structured reasoning (Guo et al., 2025).

In the context of synthetic image generation for data-scarce domains like snow detection, AR models offer unique advantages. Their ability to produce highly structured, prompt-conditioned images makes them well-suited for creating synthetic data sets that capture realistic spatial relationships and environmental features, such as snow accumulation patterns or terrain variations. Unlike diffusion models, which often prioritize raw image fidelity, AR models perform well in controllability and semantic precision, enabling them to generate images tailored to specific requirements. While diffusion models remain dominant in applications demanding photorealistic outputs, the structured visual understanding offered by AR models provides a powerful alternative.

3. Methods

Annually, tens of thousands of people end up at the emergency department as a result of winter-related fall and slip injuries, mostly among seniors and the visually impaired. These fall and slip injuries are weather-related, such as snow and ice on sidewalks, which are challenging to detect and are very slippery (Kakara et al., 2021; Mills et al., 2020). Solutions using artificial intelligence, such as image classification networks, might help these demographics to promptly recognize and avoid dangerous spots with timely alerts. These networks can be trained on sidewalk images with snow and ice, and be deployed to mobile phones and/or other wearables (de Deijn, 2024; de Deijn & Bukralia, 2024) but require large quantities of data to achieve high accuracy (de Deijn, 2024; Lecun et al., 2015). This might be one of the bigger challenges for sidewalk snow and ice detection, as these images are not as commonly available on the internet and are season-dependent to collect (de Deijn, 2024; de Deijn & Bukralia, 2024). Next to that, snow and ice have very unique optical characteristics, meaning that light reflection depends on a large number of factors such as lighting type, snow depth, environmental conditions (de Deijn, 2024; Warren, 2019), snow grain size (Dang et al., 2016; de Deijn, 2024; Zhuravleva & Kokhanovsky, 2011), light brightness, and solar zenith angle (Dang et al., 2016; de Deijn, 2024) to name a few. Synthetic data could offer a solution, as it enables researchers to generate new training data cheaply throughout the year. It has the potential to simulate all types of different environments under different reflection conditions. For synthetic data to be used as training data for a snow detection network in the real-world, photorealistic images are necessary. For this reason, a model is required to simulate the real-world, rather than creating cartoon-like images (Hossain et al., 2024).

This paper compares two prompt-based synthetic image generators: OpenAI's DALL·E 3, which is a diffusion-based image generator, and xAI's Aurora, which is an autoregression-based image generator. For the comparison, each model was used to generate 250 images of snowy sidewalks using the following prompt:

> *Generate a realistic winter scene from a first-person pedestrian perspective looking down at a cracked concrete sidewalk. The sidewalk is partially covered with icy, slushy snow, concentrated along the edge where snow transitions into ice. Footprints are visible in the icy patches. The left side of the image shows accumulated snow with small patches of brown grass peeking through. The lighting is natural, with soft daylight, capturing the texture of the snow, ice, and sidewalk in a cold, wintry outdoor setting.*

Similarly, another 250 images of clear sidewalks were generated by each model using this prompt:

> *Generate a realistic outdoor scene from a pedestrian's perspective looking down at a clean, cracked concrete sidewalk on a clear day. The sidewalk is bordered by a patch of dry grass and a landscaped area with mulch, dormant shrubs, and a large rock. Pink stakes are visible marking parts of the landscaped area. The lighting is natural and sunny, casting soft shadows and highlighting the textures of the sidewalk, grass, and surroundings. The scene captures a calm, real-world outdoor setting without including any visible objects like a smartphone.*

These prompts were designed by OpenAI's ChatGPT-4o after analyzing some example real-world sidewalk images, as shown in Figure 1, from a data set created by de Deijn & Bukralia (2024).

Figure 1. Examples of Real-World Snowy Sidewalk Data Set By de Deijn & Bukralia (2024), Used By ChatGPT-4o to Generate Image Generation Prompts

To evaluate the generated images, as shown in Figure 2, we trained models using the snowy sidewalk image classification architecture proposed by de Deijn (2024). This convolutional neural network (CNN) incorporates a spatial attention mechanism with combined average and max pooling. We configured the model with a learning rate of 0.001, a batch size of 4, and trained it for 25 epochs due to the limited data set size. The Adam optimizer and cross-entropy loss function were employed. Images were resized to 224 × 224 pixels, with data augmentation including random rotations of up to 90 degrees and random horizontal flips. By training a model on each synthetic data set, we assessed their performance against real-world data. The evaluation was conducted using a separate, unseen real-world test data set to ensure unbiased comparison.

(a) **(b)**

Figure 2. Examples of the Synthetic Images Created Using (a) DALL·E 3 and (b) Aurora

Additionally, another model was trained on different real-world data of snowy and clear sidewalks, comprising 250 images per category as well. All models were evaluated using a test data set that contains 59 images per category. Despite its small size, this data set provides an effective means of evaluating the performance of the synthetic data. Model performance is compared using the F2 score, which prioritizes recall over precision, highlighting accurate snow detection over clear sidewalk detection, as snow has a higher hazard level for pedestrian safety when not correctly identified as compared to clear sidewalks.

4. Experiment Results

Three similar networks, built on de Deijn's (2024) architecture, were trained using three distinct data sets: synthetic data sets from DALL·E 3 and Aurora, and a real-world training data set created for this study. Each model underwent twelve prediction rounds, with the real-world test data set randomly split into batches to mitigate variability from its limited size. Performance was assessed on an unseen real-world test data set using key metrics: recall (proportion of true snow instances detected), precision (accuracy of snow predictions), F2 score (emphasizing recall over precision to prioritize snow detection while considering clear sidewalk accuracy), accuracy (overall correct predictions), and ROC AUC (ability to distinguish snow from clear sidewalks). To measure statistical significance and the magnitude of performance differences, p-values (with $p < 0.05$ indicating significance) and Cohen's d (standardized effect size, where positive values favor the synthetic model and negative values favor the real-world model) were calculated relative to the real-world model's test results. The results of these tests are presented in Table 1.

Metric	Model	Mean (± Std)	P-value (vs. Real-World)	Cohen's *d* (vs. Real-World)	Interpretation
F2 score	DALL·E 3	82.49% (±0.80)	5.8871×10^{-11}	-7.4000	Significantly Worse
	Aurora	**94.40% (±0.23)**	6.6502×10^{-10}	5.9085	**Significantly Better**
	Real-World	89.53% (±0.91)	-	-	-
Precision	DALL·E 3	62.73% (±1.21)	1.8152×10^{-15}	-19.1604	Significantly Worse
	Aurora	77.13% (±0.77)	1.3991×10^{-14}	-15.9049	Significantly Worse
	Real-World	**94.28% (±0.59)**	-	-	**-**
Recall	DALL·E 3	89.55% (±0.94)	0.0128	0.8944	Significantly Better
	Aurora	**100.00% (±0.00)**	2.3771×10^{-12}	9.9440	**Significantly Better**
	Real-World	88.42% (±1.16)	-	-	-
Accuracy	DALL·E 3	68.15% (±1.40)	1.5834×10^{-14}	-15.7263	Significantly Worse
	Aurora	85.17% (±0.65)	2.7259×10^{-12}	-9.8198	Significantly Worse
	Real-World	**91.53% (±0.51)**	-	-	-
ROC AUC	DALL·E 3	67.52% (±0.68)	9.7188×10^{-21}	-57.8231	Significantly Worse
	Aurora	95.54% (±0.59)	2.0715×10^{-9}	-5.3125	Significantly Worse
	Real-World	**98.09% (±0.22)**	-	-	-

Table 1. Statistical Comparison of Models Trained on DALL·E 3, Aurora, and Real-World Data Across Twelve Test Runs. Means and standard deviations (± Std) are reported, with p-values and Cohen's *d* calculated relative to the Real-World model. A p-value < 0.05 indicates statistical significance, and Cohen's *d* reflects effect size (positive = better than Real-World, negative = worse).

Table 1 shows that the Aurora-trained model achieved the highest F2 score (94.40%) and perfect recall (100%), with t-test p-values of $p = 6.65 \times 10^{-10}$ and $p = 2.38 \times 10^{-12}$, respectively, confirming its mean scores significantly surpass the real-world model (F2: 89.53%, recall: 88.42%). To complement statistical significance, we also report Cohen's *d*, a standardized effect size that quantifies the magnitude of differences across twelve independent runs, showing very large positive values (F2: $d = 5.91$; recall: $d = 9.94$), which underscores Aurora's pronounced advantage in snow detection. This performance likely stems from Aurora's ability to generate highly photorealistic images that closely mimic real-world variability. Conversely, the real-world model led in precision (94.28%), accuracy (91.53%), and ROC AUC (98.09%), with synthetic-data models showing significantly lower means and large negative effect sizes (e.g., Aurora precision: $p = 1.40 \times 10^{-14}$, $d = -15.90$; DALL·E 3 precision: $p = 1.82 \times 10^{-15}$, $d = -19.16$). DALL·E 3 trailed across most metrics (e.g., F2: 82.49%; $p = 5.89 \times 10^{-11}$, $d = -7.40$), likely due to less realistic image detail as verified by visual data set analysis.

Surprisingly, the real-world model did not outperform synthetic models in all areas, an observation reflected not only in p-values but also in moderate to large Cohen's *d* values that highlight substantial performance gaps. Confusion matrices (see Appendix) reveal Aurora's model perfectly identifies snow (recall: $p = 2.38 \times 10^{-12}$, $d = 9.94$) but occasionally misclassifies clear sidewalks, while the real-world model balances both classes with fewer errors. The F2 score's recall bias enhances Aurora's edge (F2: $p = 6.65 \times 10^{-10}$, $d = 5.91$), with effect sizes confirming its superior snow detection. DALL·E 3's consistent underperformance (for example, ROC AUC: $p = 9.72 \times 10^{-21}$, $d = -57.82$) suggests it struggles with photorealistic detail critical for this task.

The test data set's small size (59 images per category) and limited variability (e.g., lighting, snow texture) may have impacted results. A larger, more diverse test set could improve generalizability. Aurora's stable performance, with significant mean differences ($p < 0.05$), highlights the value of high-quality synthetic data for applications like pedestrian safety, while DALL·E 3's significant gaps indicate areas for refinement.

5. Conclusion

This study highlights the transformative potential and current limitations of generative AI in creating synthetic data sets for snow detection. By comparing models trained on real-world data with those trained on synthetic data sets from DALL·E 3 and Aurora, the results reveal significant performance differences, as confirmed by t-test p-values ($p < 0.05$), showing that mean metric scores for synthetic models diverge markedly from the real-world model. Aurora's data set achieved the highest F2 score (94.40%, $p = 6.65 \times 10^{-10}$) and perfect recall (100%, $p = 2.38 \times 10^{-12}$) versus the real-world model (F2: 89.53%; recall: 88.42%), driven by its ability to catch every snow patch, as F2 weights recall more heavily. However, Aurora's lower precision (77.13%, $p = 1.3991 \times 10^{-14}$) due to higher false positives indicates it is less effective at distinguishing snow from clear sidewalks compared to the real-world model. In contrast, DALL·E 3's lower F2 score (82.49%, $p = 5.89 \times 10^{-11}$), ROC AUC (67.52%, $p = 9.7188 \times 10^{-21}$), and accuracy (68.15%, $p = 1.5834 \times 10^{-14}$) highlight its struggles with capturing real-world variability in snow appearance, making it unreliable as a standalone data source for this task.

The real-world model, despite a lower F2 score and recall, outperforms or matches both synthetic-data models across all metrics (F2, precision, recall, accuracy ROC AUC) when considering both false positives and false negatives, achieving higher precision (94.28%) and accuracy (91.53%). This balanced performance suggests that evaluation metrics and data set diversity are critical for assessing generalization. The test data set's small size (59 images per category) may exaggerate synthetic data's apparent superiority, emphasizing the need for larger, more varied data sets to validate these findings.

While generative AI has made remarkable strides, it is not yet a full substitute for real-world data. Aurora's success suggests that synthetic data sets can effectively complement traditional data, particularly for tasks like snow detection, where real-world collection is challenging. However, DALL·E 3's significant performance gap indicates that further refinements are needed. Future research should prioritize enhancing generative models to capture environmental complexities, potentially through physics-based lighting models, increased scenario variability, or advanced prompt engineering, to produce synthetic data sets that meet the demands of robust machine learning applications.

Ultimately, this study underscores that generative AI is narrowing the gap with real-world data. The significant differences in performance, backed by t-test results, point to a future where synthetic data, as exemplified by Aurora, can augment real-world data sets, enabling scalable, effective solutions for machine learning tasks like pedestrian safety.

6. Discussion

This study faced several challenges that influenced its results. One of the primary limitations lies in the dependency of synthetic image generation models on the quality and precision of the prompts provided by the researcher or user. Poorly constructed prompts can negatively impact the quality and photorealism of the generated images. This issue was highly present in this study, as pictures produced by OpenAI's DALL·E were not as photorealistic as those generated by xAI's Aurora, despite both data sets being created using identical prompts. The two prompts were generated by ChatGPT based on some example images, with the thought that this would return the most optimally engineered prompt for the model, but this did not optimally deliver for DALL·E. This underscores the importance of the model architecture, the underlying data used for training, and the quality of prompts in determining the quality and relevance of generated images.

Additionally, while synthetic data sets addressed data scarcity, they still could face challenges in generalizing real-world scenarios. The constrained nature of prompts and the uniformity of generated data sets can lead to overfitting, where models perform well on test sets with similar characteristics but struggle with diverse or unforeseen real-world conditions. This highlights a key limitation of synthetic data: its ability to replicate specific scenarios but not the variability and complexity of real-world environments. For instance, while Aurora generated more realistic images, the synthetic data sets lacked diversity in surface textures, lighting conditions, and environmental contexts, such as the appearance of other white objects that you might see outside, such as chalk, paint, or dandelion puffs, which pose critical factors for robust snow detection models. A solution would be to expand the data set with more variable data or to apply domain adaptation by training a model on synthetic images and fine-tuning it on real images. This would reduce the need for large data sets of real-world data.

Google's Imagen3 model was also evaluated using the prompts. While it successfully generated images for the snow-covered sidewalk prompt, it failed to produce images for the clear pavement prompt, citing policy restrictions. Consequently, this limitation prevented Imagen3 from being included in the comparative analysis. Visually, the snow images generated by Imagen3 appeared more realistic than those from DALL·E but less so than Aurora. However, because of its inability to generate a complete data set under identical conditions, a formal evaluation of Imagen3's performance could not be conducted in this study.

Future research should address these limitations by focusing on improving both the diversity and quality of synthetic data sets. Refining prompt engineering techniques and developing frameworks that automate prompt optimization could reduce the reliance on user expertise, making synthetic data generation more accessible and reliable. Furthermore, as AI-generated images become increasingly indistinguishable from real-world data, it is important to think about mechanisms such as permanent watermarks on AI-produced images. These identifiers would ensure transparency of synthetic data usage, mitigating risks associated with misuse or misrepresentation. Such watermarks must be robust and impervious to tampering, striking a balance between maintaining the utility of the data for machine learning applications and addressing ethical concerns around authenticity.

Acknowledgments

This experiment was conducted without any external funding or grants.

7. References

Arif, M., & Mahalanobis, A. (2021). Infrared Target Recognition Using Realistic Training Images Generated by Modifying Latent Features of an Encoder–Decoder Network. *IEEE Transactions on Aerospace and Electronic Systems*, *57*(6), 4448–4456. https://doi.org/10.1109/TAES.2021.3090921

Atallah, S. B., Banda, N. R., Banda, A., & Roeck, N. A. (2023). How large language models including generative pre-trained transformer (GPT) 3 and 4 will impact medicine and surgery. *Techniques in Coloproctology*, *27*(8), 609–614. https://doi.org/10.1007/s10151-023-02837-8

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, *15*(8), 260. https://doi.org/10.3390/fi15080260

Barbosa, I. B., Cristani, M., Caputo, B., Rognhaugen, A., & Theoharis, T. (2018). Looking beyond appearances: Synthetic training data for deep CNNs in re-identification. *Computer Vision and Image Understanding*, *167*, 50–62. https://doi.org/10.1016/j.cviu.2017.12.002

Chang, D.-L., Yang, S.-H., Hsieh, S.-L., Wang, H.-J., & Yeh, K.-C. (2020). Artificial Intelligence Methodologies Applied to Prompt Pluvial Flood Estimation and Prediction. *Water*, *12*(12), 3552. https://doi.org/10.3390/w12123552

Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, *21*(4), 471–496. https://doi.org/10.1080/14765284.2023.2245279

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

Dang, C., Fu, Q., & Warren, S. G. (2016). Effect of Snow Grain Shape on Snow Albedo. *Journal of the Atmospheric Sciences*, *73*(9), 3573–3583. https://doi.org/10.1175/JAS-D-15-0276.1

de Deijn, R. (2024). *Developing a Snow Detection Algorithm Using Spatial Attention for Pedestrian Safety* [Master's thesis, Minnesota State University, Mankato]. https://www.proquest.com/openview/e98fa5068183de2880a438d6fd5ad9c9/1

de Deijn, R., Batra, A., Koch, B., Mansoor, N., & Makkena, H. (2024). Reviewing FID and SID Metrics on Generative Adversarial Networks. *AI, Machine Learning and Applications*, 111–124. https://doi.org/10.5121/csit.2024.140208

de Deijn, R., & Bukralia, R. (2024, June 30). Image Classification for Snow Detection to Improve Pedestrian Safety. *Proceedings of MWAIS 2024*. http://arxiv.org/abs/2407.00818

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv Preprint ArXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Donahue, J., Krähenbühl, P., & Darrell, T. (2016). Adversarial feature learning. *ArXiv Preprint ArXiv:1605.09782*. http://arxiv.org/abs/1605.09782

Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., & Courville, A. (2016). Adversarially Learned Inference. *ArXiv Preprint ArXiv:1606.00704*. https://doi.org/10.48550/arXiv.1606.00704

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elfeki, A., Masoud, M., & Niyazi, B. (2017). Integrated rainfall–runoff and flood inundation modeling for flash flood risk assessment under data scarcity in arid regions: Wadi Fatimah basin case study, Saudi Arabia. *Natural Hazards*, *85*(1), 87–109. https://doi.org/10.1007/s11069-016-2559-7

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June 10). Generative Adversarial Networks. *Proceedings of Advances in Neural Information Processing Systems*. http://arxiv.org/abs/1406.2661

Guo, Z., Zhang, R., Tong, C., Zhao, Z., Gao, P., Li, H., & Heng, P.-A. (2025). Can We Generate Images with CoT? Let's Verify and Reinforce Image Generation Step by Step. *ArXiv Preprint ArXiv:2501.13926*. http://arxiv.org/abs/2501.13926

Haan, K. (2023, April 25). *24 Top AI Statistics and Trends In 2024*. Forbes Advisor. https://www.forbes.com/advisor/business/ai-statistics/

Hao, Y., Chi, Z., Dong, L., & Wei, F. (2023). Optimizing Prompts for Text-to-Image Generation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Proceedings of Advances in Neural Information Processing Systems* (pp. 66923–66939). Curran Associates, Inc. http://arxiv.org/abs/2212.09611

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Proceedings of Advances in Neural Information Processing Systems*, 6480–6851. https://doi.org/10.48550/arXiv.2006.11239

Hossain, S., Ibrahim Protik, T., & Sazid, G. (2024). Potential Risk of Faking Microscopic Images in the Characterization of Nanomaterials Through Artificial Intelligence. In *Proceedings of International Conference on Physics*. https://doi.org/10.13140/RG.2.2.25498.45766

Kakara, R. S., Moreland, B. L., Haddad, Y. K., Shakya, I., & Bergen, G. (2021). Seasonal variation in fall-related emergency department visits by location of fall – United States, 2015. *Journal of Safety Research*, *79*, 38–44. https://doi.org/10.1016/j.jsr.2021.08.002

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *ArXiv Preprint ArXiv:1312.6114*. https://doi.org/10.48550/arXiv.1312.6114

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Li, D.-C., Hu, S. C., Lin, L.-S., & Yeh, C.-W. (2017). Detecting representative data and generating synthetic samples to improve learning accuracy with imbalanced data sets. *PLOS ONE*, *12*(8), e0181853. https://doi.org/10.1371/journal.pone.0181853

Li, T., Tian, Y., Li, H., Deng, M., & He, K. (2024). Autoregressive Image Generation without Vector Quantization. *ArXiv Preprint ArXiv:2406.11838*. http://arxiv.org/abs/2406.11838

Meng, C., Rombach, R., Gao, R., Kingma, D., Ermon, S., Ho, J., & Salimans, T. (2023). On Distillation of Guided Diffusion Models. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14297–14306. https://doi.org/10.1109/CVPR52729.2023.01374

Mescheder, L., Nowozin, S., & Geiger, A. (2017). Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks. *Proceedings of International Conference on Machine Learning*, 2391–2400. https://doi.org/10.48550/arXiv.1701.04722

Mills, B., Andrey, J., Doherty, S., Doberstein, B., & Yessis, J. (2020). Winter Storms and Fall-Related Injuries: Is It Safer to Walk than to Drive? *Weather, Climate, and Society*, *12*(3), 421–434. https://doi.org/10.1175/WCAS-D-19-

0099.1

Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. *ArXiv Preprint ArXiv:1411.1784*. http://arxiv.org/abs/1411.1784

Nichol, A., & Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models. *Proceedings of International Conference on Machine Learning*, 8162–8171. https://doi.org/10.48550/arXiv.2102.09672

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of International Conference on Machine Learning*, 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. In *ArXiv*. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. In *Journal of Machine Learning Research* (Vol. 21). http://jmlr.org/papers/v21/20-074.html.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. *ArXiv Preprint ArXiv:2204.06125*. https://doi.org/10.48550/arXiv.2204.06125

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. *Proceedings of International Conference on Machine Learning*, 8821–8831. https://doi.org/10.48550/arXiv.2102.12092

Ren, S., Yu, Y., Ruiz, N., Wang, F., Yuille, A., & Xie, C. (2024). M-VAR: Decoupled Scale-wise Autoregressive Modeling for High-Quality Image Generation. *ArXiv Preprint ArXiv:2411.10433*. http://arxiv.org/abs/2411.10433

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695. https://doi.org/10.48550/arXiv.2112.10752

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Proceedings of Advances in Neural Information Processing Systems*, 36479–36494. https://doi.org/10.48550/arXiv.2205.11487

Shaji George, A., Hovan George, A. S., & Gabrio Martin, A. S. (2023). A Review of ChatGPT AI's Impact on Several Business Sectors. *Partners Univers Int Innov J*, *1*(1), 9–23. https://doi.org/10.5281/zenodo.7644359

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *Proceedings of International Conference on Machine Learning*, 2256–2265. https://doi.org/10.48550/arXiv.1503.03585

Valeriya, G., John, V., Singla, A., Yamini Devi, J., & Kumar, K. (2024). AI-Powered Super-Workers: An Experiment in Workforce Productivity and Satisfaction. *BIO Web of Conferences*, *86*, 01065. https://doi.org/10.1051/bioconf/20248601065

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, & R. Fergus (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 6000–6010). Curran Associates Inc. http://arxiv.org/abs/1706.03762

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F.-Y. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, *4*(4), 588–598. https://doi.org/10.1109/JAS.2017.7510583

Warren, S. G. (2019). Optical properties of ice and snow. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *377*(2146), 20180161. https://doi.org/10.1098/rsta.2018.0161

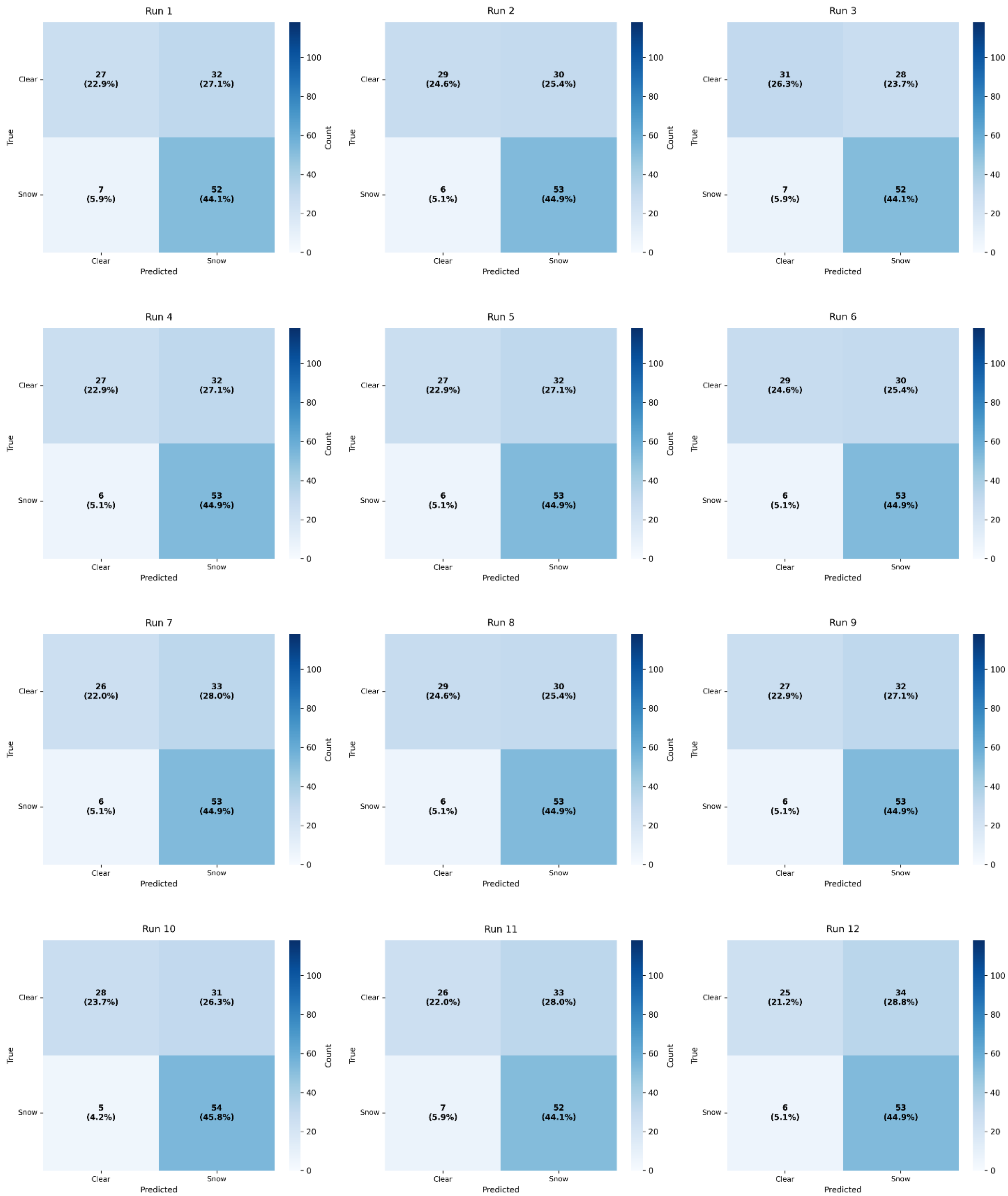
Zhang, T., Wang, Z., Huang, J., Tasnim, M. M., & Shi, W. (2023). A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions. *ArXiv Preprint ArXiv:2308.13142*. https://doi.org/10.48550/arXiv.2308.13142

Zhou, Y., Sahak, H., & Ba, J. (2023). Training on Thin Air: Improve Image Classification with Generated Data. *ArXiv Preprint ArXiv:2305.15316*. https://doi.org/10.48550/arXiv.2305.15316

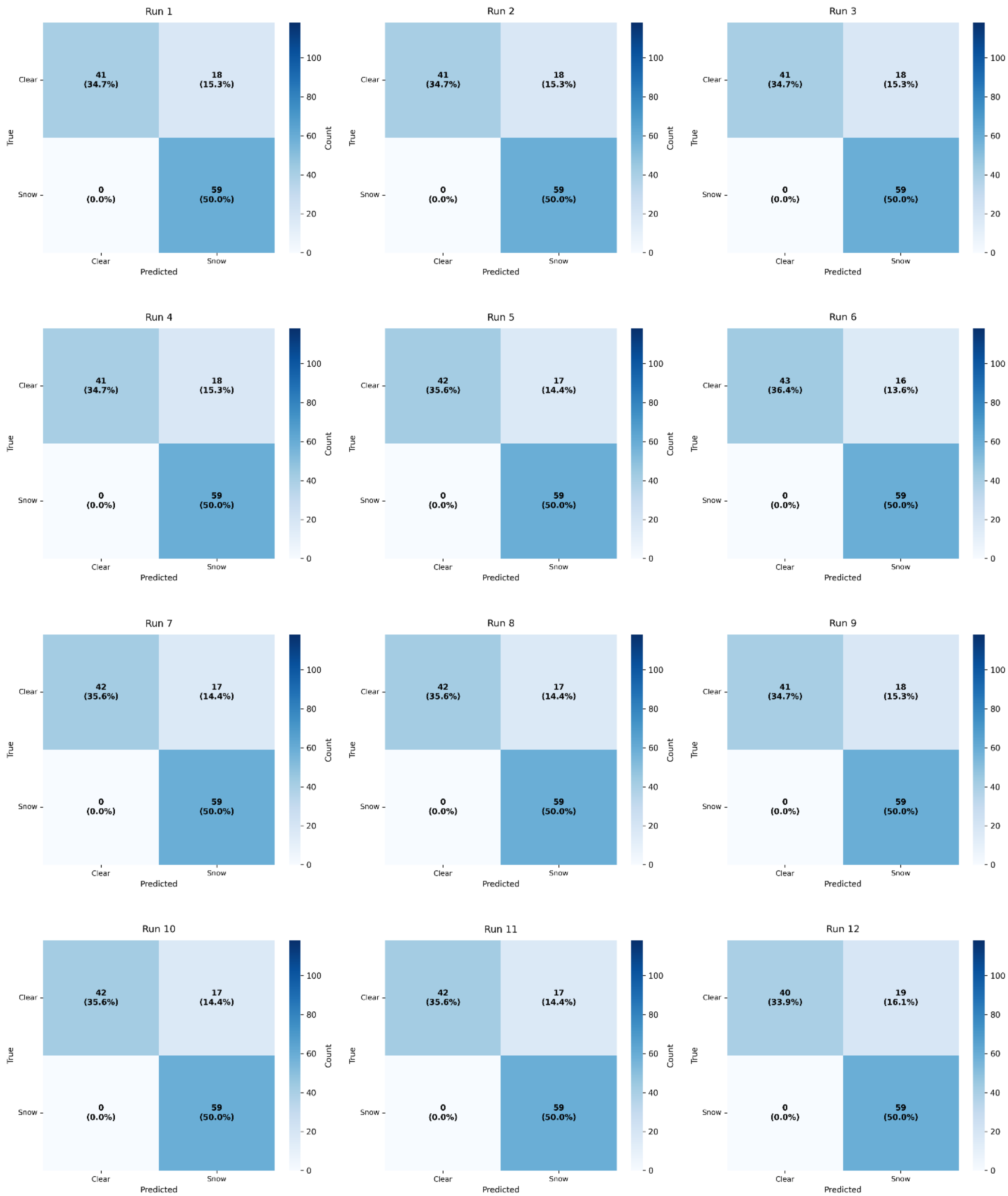
Zhuravleva, T. B., & Kokhanovsky, A. A. (2011). Influence of surface roughness on the reflective properties of snow. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *112*(8), 1353–1368. https://doi.org/10.1016/j.jqsrt.2011.01.004

Appendix

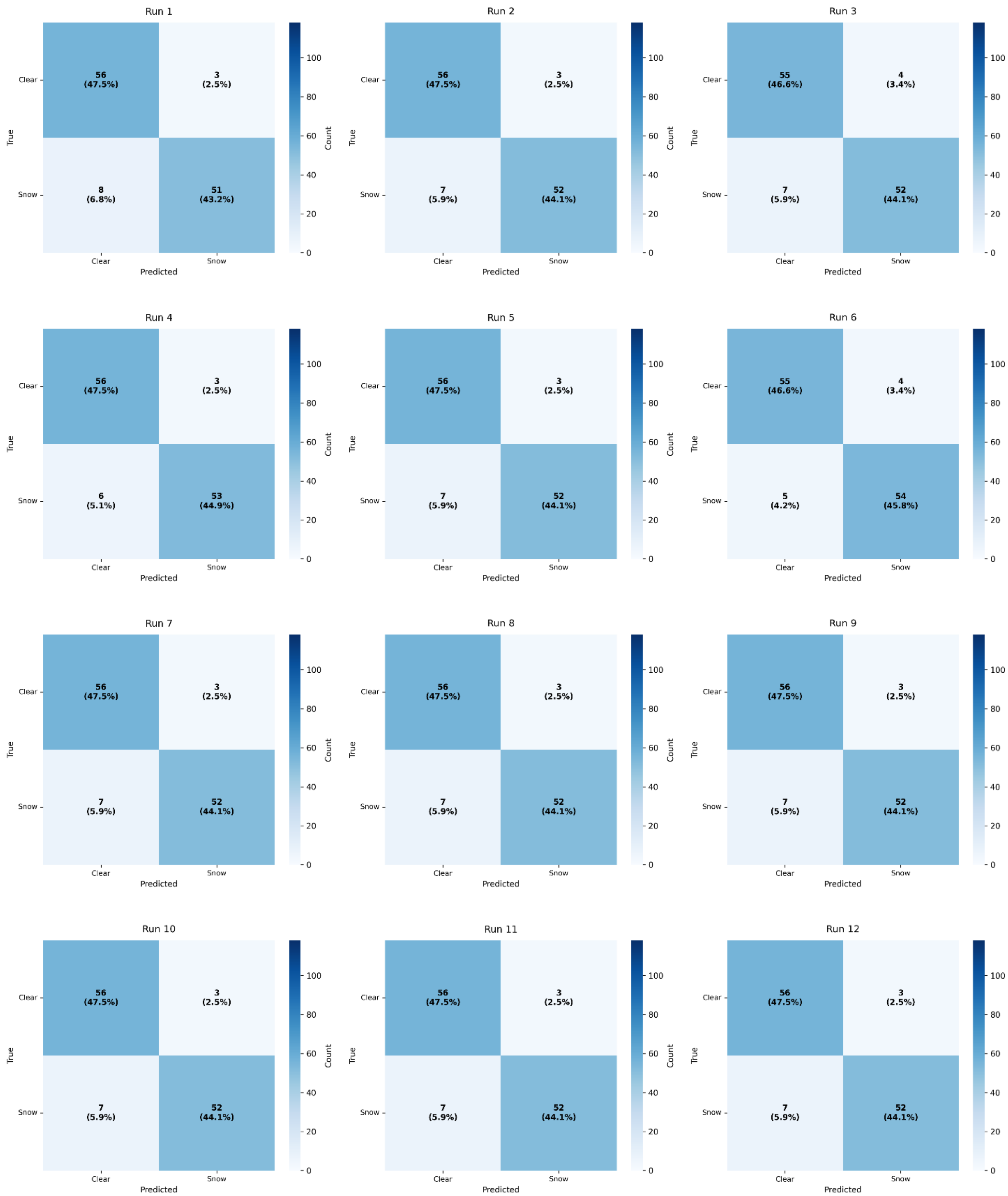
Appendix I. Model Confusion Matrices



Appendix I Fig. 1. DALL·E 3 Test Data Confusion Matrices



Appendix I Fig. 2. Aurora Test Data Confusion Matrices



Appendix I Fig. 3. Real-World Test Data Confusion Matrices

Author Biographies

Ricardo de Deijn is a data scientist and researcher based in Minnesota, with a Master of Science in Data Science from Minnesota State University, Mankato. Originally from the Netherlands, his fascination with snow and winter safety began after experiencing a Minnesota winter where snowfall reached up to his head. His research focuses on computer vision, real-time edge AI, and synthetic data generation for public safety. His work has been featured at academic conferences such as MWAIS, AIMLA, and CADSCOM, where he received a best paper award. He is also interested in exploring emerging and underutilized techniques, including mobile AI deployment and federated learning, for on-device inference in safety-critical environments.

Rajeev Bukralia is a tenured associate professor and the founding director of the Data Science and Artificial Intelligence programs in the Department of Computer Information Science at Minnesota State University, Mankato. He is a recipient of Minnesota's Tekne Tech Educator of the Year Award and the Minnesota State Board of Trustees' Outstanding Educator Award.