Journal of the Midwest Association for Information Systems

Volume 2026 | Issue 1 Article 1

Date: 01-31-2026

The Implications of AI and Generative AI for Teaching and Learning

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

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

Abstract

GAI has received significant investment and adoption across various businesses over a relatively short time period. Just as the Internet and the World Wide Web did several decades ago, GAI will substantially alter how teaching and learning are conducted and how institutions of higher education will use this technology and related tools. A function of higher education institutions is to prepare the workforce of the future. However, given that incorporating GAI in teaching and learning creates both opportunities and challenges, a major question is how faculty, students, and higher education institutions should properly adopt this rapidly changing and transformative technology. In this position paper, we propose a multi-factor framework that addresses the ethical and appropriate use of this technology in teaching and learning for faculty and students.

Keywords: GAI in teaching and learning, 21st-century educational competency

DOI:10.17705/3jmwa.000096

1. Introduction

The implications of Generative AI (GAI) in teaching and learning are not only broad but also significantly impactful (Wang, Jing, and Shen, 2025). Just as the Internet and the World Wide Web did several decades ago, GAI will substantially alter how teaching and learning are conducted and how institutions of higher education will use this technology and related tools. Numerous articles have been published on how to prepare students for the transition to "Industry 4.0" (Cantú-Ortiz et al., 2020, p. 1195). Many institutions are already advocating for the ethical, transparent, and proper uses of this technology. McDonald, Johri, Ali, and Collier (2025) analyzed documents from 116 US universities and found that the majority of them (63%) encourage appropriate uses of GAI in teaching and learning. External accreditation agencies such as AACSB and ACBSP are advocating the proper uses of GAI in teaching and learning and for uses in the preparation and documentation reports regarding meeting and maintaining accreditation standards.

This technology has significant benefits, such as increasing students' engagement, eliminating or at least reducing the learning inequity, improving critical thinking, and helping students' career readiness (Honigsberg, Watkowski, & Drechsler, 2025), and improving faculty productivity. It also creates technical, ethical, societal and economic, regulatory, and governance challenges (Nah et al., 2023). Some universities, such as Carnegie Mellon, Harvard, and Massachusetts Institute of Technology, have developed courses and programs that are focused on ethical uses of AI (Southworth, 2023). Nah et al. (2023) suggest considering three major characteristics of GAI – "content generation, generalization ability, and reinforcement learning using human feedback" to better understand how this technology may be used in teaching and learning (p. 249). This technology is currently being used at all educational levels, including elementary, middle, and high school, and at the higher education level (Southworth et al., 2023).

Regarding career readiness, AI literacy, its integration in our daily lives (Luckin, et al., 2022; NG, et al., 2021), and a curriculum suitable for the 21st. century, some universities have initiated processes to require AI competency for all undergraduates as a part of graduation requirements in the near future. The Ohio State University (OSU) https://oaa.osu.edu/ai-fluency, State University of New York (SUNY) https://www.suny.edu/suny-news/press-releases/1-25/1-7-25/general-education.html, University of Michigan (UM) https://genai.umich.eu, and UC Berkeley https://www.berkeley.edu/ai/, New Jersey Institute of Technology (NJIT) https://ldi.njit.edu/ai-core-competencies, Jacksonville State University (JSU) https://catalog.jsu.edu/undergraduate/library-services/ai-literacy-microcredential/, San Jose State University (SJSU) https://www.sjsu.edu/cfeti/teaching-resources/ai-pedagogies/ai-literacy.php, and University of South Carolina (USC) https://sc.edu/about/offices_and_divisions/provost/academicpriorities/undergradstudies/interdisciplinary-certificates/ai-literacy.php are some examples. A similar AI competency plan is happening in other countries. For example, in the European Union, the Asia-Pacific Region, and China (Kandlhofer et al., 2016; Su et al., 2022; Dai et al., 2020; Cantú-Ortiz et al., 2020).

The University of Florida (UF) has been planning similar requirements since 2023. Southworth et al. (2023) report that: "The AI Across the Curriculum" initiative being developed at UF will make AI education a cornerstone opportunity for all students," p. 1. The UF has received infrastructure support from NVIDIA (Merrit, 2021) to achieve its goal of developing AI competency across its entire curriculum. To accomplish these important goals, the UF established the Artificial Intelligence Academic Initiative Center (Southworth et al., 2023) in 2022. Due to the importance of AI technology and its implications, it is critical to point out that AI needs to be incorporated across all disciplines, not just any particular area. With the integration of AI and Quantum computing, which facilitates predicting the 3D-structure of proteins, scientists hope to be able to cure cancer soon (Shontell 2020). Hence, the idea of AI competency and its ethical use for all disciplines, including non-STEM (Science, Technology, Engineering, and Mathematics), is essential (Su et al., 2022).

In a recent AACSB career connection article, Williams (2025) stated that: "If we keep preparing students for the workplace of yesterday, we'll keep failing the workforce of tomorrow. To remain relevant, business schools must do more than update content; they must build cultures of continuous learning. They must reimagine their programs and create interconnected systems that support student choice, faculty innovation, employer partnerships, and real-time skill

development,"(p.3). Both ACBSP (https://acbsp.org/page/ai) and AACSB (https://www.aacsb.edu/educators/accreditation/business-accreditation/ai-use-cases-for-accreditation)

have developed GAI-related websites to support faculty and higher education institutions. The ACBSP site is a "collaborative Padlet" in support of faculty for appropriate uses of GAI. The AACSB site includes "actionable guidance on leveraging AI to meet and maintain AACSB accreditation standards."

A major question is how faculty, students, and higher education institutions should properly adopt this rapidly changing and transformative technology. As far as incorporating AI in the curriculum is concerned, Hashmi and Bal (2024) propose a "two-dimensional toolbox framework" for "AI human enhancing" and "AI human replacing" for "existing" and "new" skills in the curriculum (p. 612). Using a framework seems logical, particularly given the rapid development and fast-changing nature of this technology and related tools. We believe that more factors need to be considered when incorporating this technology in teaching and learning.

In this position paper, we propose a multi-factor framework that addresses the use of this technology in teaching and learning for faculty and students, and addresses implications for higher education institutions.

2. A Multi-factor Framework for Incorporating GAI in Teaching and Learning

To better understand why and how GAI could be used in teaching and learning, we first need to clarify and make some distinctions. Not all disciplines will use GAI the same way when incorporating it in teaching and learning. For example, there are significant differences in using GAI to teach a digital marketing class versus a chemistry class. Hence, one factor is the discipline's nature and characteristics. The other difference, obviously, is how faculty and students use it in teaching and learning.

Our focus will be on both faculty and learners, and what each group needs to pay attention to using this transformative technology. We will look at the following factors:

- Ways to ethically and transparently use GAI to improve the overall teaching and learning processes and outcomes
- How should faculty rethink their pedagogy and use this technology effectively?
- What kinds of guidelines, such as Ethical Use, Limitations, and Bias Training, are required for faculty and students to properly use this technology?
- What are the prompt engineering skills and abilities faculty and students need to have to be able to effectively use this technology?
- What knowledge and skills should students be taught about the GAI to help their career readiness?
- What are the limitations and areas faculty, students, and higher education institutions need to avoid when incorporating this technology?
- What types of jobs are likely to be created due to the existence of GAI technology? What types of jobs are likely to change or disappear?

2.1 Ways to Ethically and Transparently Use GAI to Improve the Overall Teaching and Learning Processes and Outcomes

GAI technology itself is evolving. Any applications of it need to consider this reality. For many disciplines, it is not a valid argument to prevent faculty and students from using GAI to prevent plagiarism. The advantages that the appropriate uses of this transformative technology, which some researchers are calling "Industry 4.0" (Cantú-Ortiz et al., 2020, p. 1196), far exceed its potential negative effects. Faculty need to focus on how to use this technology to improve the efficiency and effectiveness of their teaching. Faculty can use this technology to significantly increase the learners' engagement with the course contents and personalize their education to the needs of each student in the class. They can use their time more effectively by automating routine teaching tasks and spending more time with each student to understand their strengths and potential weaknesses, allowing them to better focus on the learning preferences of each learner. Another critical contribution is for the faculty to determine the

career readiness of each of the students and determine what is the best way to use GAI and provide guidance for students to ethically use GAI to support their own career readiness.

As described and documented in the introduction, many educational institutions are already advocating and planning to make AI competency a part of their graduation requirement. Focusing on the career readiness of each learner and customizing the AI competency further enhances the learners' career readiness capabilities.

2.2 How Should Faculty Rethink their Pedagogy and use this Technology Effectively?

One of the major implications of the availability of GAI to both faculty and students is the fact that the traditional pedagogy needs to be modified to ethically and effectively use this technology in teaching and learning. The traditional pedagogy consists of i) lesson planning, ii) content presentation, iii) explanation and clarification, iv) practice and reinforcement, v) assessment, vi) review and remediation. The GAI technology can be utilized in all of these steps to enhance presentation quality, effectiveness, and productivity. Effective use of this technology enables faculty members to enhance engagement with learners, thereby gaining a deeper understanding of their individualized learning preferences and customizing the learning experience for each of their learners. This allows them to tailor their pedagogy to each learner, rather than relying on one-size-fits-all instruction. As described in the VARK model (Othmana & Amiruddin, 2010), more visual learners would prefer images, videos, charts, and graphs. GAI can easily be used to create high-quality charts and graphs. For auditory learners, podcasts and recorded lessons can be a valuable resource. For verbal learners, lecture summaries can be prepared by faculty or students themselves. For kinesthetic learners who prefer hands-on learning, practical assignments, experiments, and role-playing exercises would be preferred. For the social learners, group assignments and collaborative discussions will be more effective. The GAI technology can be used to more effectively modify the prepared lesson plans to fit different learning styles.

Using GAI can enhance the assessment process by making it more personalized based on the learners' learning preferences. A critical note here is that the assessment process needs to be changed in any course where the use of GAI is allowed by the faculty. It is no longer appropriate to rely on essay-writing types of assignments as a partial or full method of assessing learners. Instead, individual or group projects and presentations can be used.

2.3 What Kinds of Guidelines, such as Ethical Use, Limitations, and Bias Training, are required for Faculty and Students to Properly use this Technology?

The most critical training for both faculty and students is related to the ethical uses of this technology. Both faculty and students need to be familiar with what is acceptable when using this technology. Transparency and appropriate disclosures are important and are an easy way to articulate when developing the training and guidelines. Faculty need to clearly articulate in their course syllabi if the use of GAI is acceptable in their courses and for what purpose or assignments it is allowed. Students need to clearly and accurately state where and how they are using this technology in any assignment.

As it is well documented, this technology and related platforms are known to have generated bias and inaccurate output based on hallucination. Hence, it is critical to carefully review and verify the accuracy of any output produced as well as requesting and checking references. Another safeguard is to use different platforms and compare the output.

Another major focus and training needs to be related to general skill training and ethics.

2.4 What are the Prompt Engineering Skills and Abilities Faculty and Students Need to Have to be Able to Effectively use this Technology?

An important skill all GAI users need to learn very early is how to effectively and accurately craft their prompts. Because of the lack of adequate reasoning capability of the large language models, a more accurate prompt may prevent potential hallucination-embedded output. Various, but similar definitions of prompt engineering exist. IBM (https://www.ibm.com/think/topics/prompt-engineering) defines it as: "the process of structuring or crafting an

instruction in order to produce better outputs from a generative artificial intelligence model." Wikipedia (https://en.wikipedia.org/wiki/Prompt_engineering) defines it as: "the process of structuring or crafting an instruction in order to produce better outputs from a generative artificial intelligence model." Liu et al. (2023) suggest prompt engineering involves cleansing inputs to GAI platforms to generate high-quality outputs.

The main purpose of a prompt is to ask GAI platforms a question and expect an accurate response. As such, making sure the prompt is as precise as possible (Gimpel et al. 2024; Dwivedi et al., 2023) is critical. Given that these platforms can produce inaccurate and even biased responses, it is essential to include citation requests in any prompt and to double-check and verify the accuracy of the information provided.

Tolzin, Knoth, and Janson (2024) conducted an experimental study involving about 250 students and found that exposing students to some example-based prompting significantly improved their ability to refine their prompts that resulted in more accurate output. One essential skill for any user of GAI platforms is to learn to write better and more refined prompts. This can be achieved by looking at some examples and providing more accurate input.

Various prompt engineering training and short courses exist for all users. Coursera, for example, offers such a training known as Prompt Engineering Mastery "(https://www.coursera.org/learn/prompt-engineering-mastery). The essential characteristics of prompt engineering OpenAI (openai.com), ChatGPT (chatgpt.com), and ChatGPT Business (chatgpt.com/business) are:

Clarity: Be as specific as possible when developing your prompt.

Context: Provide adequate and specific background, limitations, or constraints

Structure: Specifically, pay attention to your use of any formatting, examples, steps, or lists.

Iteration: One of the best practices is to refine your prompt as many times as needed to improve output quality. Another important step is to try to use multiple platforms, inputting similar prompts, and compare the output quality.

Role Assignment: Assign a role to the model as a part of your prompt. For example, if you are seeking language translation, assign the role of interpreter.

As faculty and students develop a broader repertoire of more sophisticated prompt strategies (e.g., Schmidt et al., 2023), the gap between student and faculty perceptions of the utility of GAI and the workplace skills needed for successful careers is likely to shrink.

2.5 What Knowledge and Digital Skills Should Students be Taught about the GAI to Help their Career Readiness?

The current and future generations of learners need adequate and ethical AI literacy to not only incorporate this technology in their daily lives but also be ready for the opportunities that this technology brings to the workplace. They also need to learn and know how this technology is potentially supporting employers and employees in their selected field of study and what knowledge, skills, and experience they need to be better career-ready. As Williams (2025) stated: "Our students' success will depend not only on how well we adopt curricula to AI, but on how much we teach a core skill that is too often overlooked" (p. 1). There is no doubt that the GAI technology is going to have major implications for many careers. This has already been demonstrated. The likely scenario is that some workers without the GAI technology will be replaced by those who have the required knowledge and skills. The employability rate (the percent of students who find employment within three months after graduation) for many fields is about 94%. This rate for AI programs is about 100% (Cantú-Ortiz et al., 2020).

Southworth et al. (2023) propose five AI-related areas with specific contents and learning outcomes. These are: "enabling AI, know & understand AI, use and apply AI, evaluate and create AI, and AI ethics" (p. 7). These and similar skill sets can be incorporated as a part of general education for almost all disciplines, with different emphasis based on the areas of study. It is important to point out that in any AI-related training and skill development, a major focus needs to be on the critical thinking, problem-solving, and communication development of the learners. Given the rapid pace of technology innovation and change, any GAI-related skill development and training needs to be regularly updated to meet the needs of employers to enhance the career readiness of learners.

2.6 What are the Limitations and Areas that Faculty, Students, and Higher Education Institutions Need to Avoid When Incorporating this Technology?

Incorporating GAI in teaching and learning creates opportunities but also some challenges. As we listed in previous sections, if used properly, GAI can support teaching and learning in a number of ways, including increasing students' engagement, personalizing learning, and helping with career readiness. Using GAI can also potentially produce misinformation and bias (Susarla et al., 2023). Faculty and students need to be aware of this possibility. One relatively simple approach to at least reduce the potential for bias is to specifically ask for sources in any prompt and then verify that the sources are real and check the accuracy of the contents with other sources.

2.7 What Types of Jobs are Likely to be Created Due to the Existence of GAI Technology? What Types of Jobs are Likely to Change or Disappear?

Automation changes the nature of many jobs. Over the years, we have observed the transformative nature of various technologies, in particular digital technologies. When the Internet and the Web came about, we saw the impact of these technologies. Educational pioneers took advantage of these technologies and developed a new way of providing education over distance that became known as online education. The impact of this innovation is now very clear given its use, particularly during COVID. The GAI technology's impact is significantly broader.

The 2025 Future of Jobs report from the World Economic Forum https://www.weforum.org/publications/the-future-of-jobs-report-2025/ states that "Technology-related roles are the fastest-growing jobs in percentage terms, including Big Data Specialists, Fintech Engineers, AI and Machine Learning Specialists, and Software and Application Developers" (p. 6). Other innovations include robotics, autonomous driving, computer vision, natural language processing, computer vision, speech recognition, agentic commerce, general delivery services, surveillance, healthcare, entertainment, and education, as examples of some important applications.

We have already seen the implementation of agentic AI in several areas. Some jobs related to retail transactions will likely be affected by GAI. Some clerical, executive assistance, and cashier jobs are expected to see significant change and decline due to GAI implementation. On the other hand, big data and analytical abilities will continue to be highly in demand.

3. Conclusions

We can say without reservation that appropriate and ethical uses of AI and GAI technologies, in particular, will have significant and impactful implications not only for higher education, businesses, organizations, but also the entire society. It is essential to point out and remind all users that only appropriate and ethical uses of these technologies by all users, including higher education institutions, are recommended and are acceptable.

4. Overview of the Contents of this Issue

This issue of the journal includes three other articles. Steven Schilhabel, in his interesting article, writes about the use of artificial intelligence (AI) systems in consumer-facing types of decision-support systems. The article describes AI-assisted betting in general and in sports betting, in particular. The article analyzes survey data and classifies three distinct behavioral profiles and concludes that betters can engage in accountable adoption.

Royce Fichtner and Troy Strader, in their timely article, articulate both positive and negative implications of deepfake technology. The authors looked at all state-level legislation for the last seven years and documented how states have used criminal laws to challenge harmful deepfake applications.

Vishal Shah and Javad Norouzi Nia argue convincingly in their viewpoint article that information systems professionals need to advance their design practices to make meaningful use of insights drawn from user-generated content. They specifically introduce the Lead UX framework as a part of their articulation.

We appreciate and wish to acknowledge the contributions of reviewers for this issue of the journal, including Foluso Ayeni (Metro State University), Gaurav Bansal (Ohio University), Queen Booker (Metro State University), Yi "Maggie" Guo (University of Michigan-Dearborn), Jeffery Merhout (Miami University, Oxford, Ohio), Alanah Mitchell (Drake University), Troy Strader (Drake University), and Jake Young (Bradley University).

5. References

Cantú-Ortiz, F. J., Galeano S´ anchez, N., Garrido, L., Terashima-Marin, H., & Brena, R. F. (2020). An artificial intelligence educational strategy for the digital transformation. International Journal on Interactive Design and Manufacturing, 14, 1195–1209. https:// doi.org/10.1007/s12008-020-00702-8

Dai, Y., Chai, C. S., Lin, P. Y., Jong, M. S. Y., Guo, Y., Qin, J. (2020). Promoting students' well-being by developing their readiness for the artificial intelligence age, Sustainability, 12 (16), p. 6597, 10.3390/su12166597

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). 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, Article 102642.

Gimpel, H., Gutheil, N., Mayer, V., Bandtel M., Büttgen, M., Decker, S., Eymann, T., Feulner, S., Kaya, M.F., Kufner, M., Kühl, N., Lämmermann L., Mädche, A., Ruiner, C., Schoop, M., Urbach, N. (2024). (Generative) AI Competencies for Future-Proof Graduates. Inspiration for Higher Education Institutions. Stuttgart, Germany: University of Hohenheim, February 19, 2024. https://doi.org/10.5281/zenodo.10680210
Hashmi, N. and Bal, A. S. (2024). Generative AI in higher education and beyond. *Business Horizons,* Volume 67, Issue 5, 607-614, doi.org/10.1016/j.bushor.2024.05.005.

Honigsberg, S., Watkowski, L., & Drechsler, A. (2025). Generative Artificial Intelligence in Higher Education: Mediating Learning for Literacy Development. *Communications of the Association for Information Systems*, 56, 1044-1076. https://doi.org/10.17705/1CAIS.05640

Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., Huber, P. (2016). Artificial intelligence and computer science in education: From kindergarten to university, IEEE Frontiers in Education Conference (FIE), pp. 1-9, 10.1109/FIE.2016.7757570

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9), 1–35.

Luckin, R., Cukurova, M., Kent, C., & du Boulay, B. (2022). Empowering educators to be AI-ready. Computers & Education: Artificial Intelligence, 3. https://doi.org/10.1016/j. caeai.2022.100076. Article 100076.

McDonald, N., Johri, A., Ali, A., and Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans,* Volume 3, doi.org/10.1016/j.chbah.2025.100121.

Merrit, R. (2021). AI Vision guides University of Florida's rise in college rankings, https://blogs.nvidia.com/blog/2021/09/14/university-of-florida-rankings-ai/, Accessed November 10, 2025.

Nah, F. F.-H., Cai, J., Zheng, R., & Pang, N. (2023). An activity system-based perspective of generative AI: Challenges and research directions. AIS Transactions on Human-Computer Interaction, 15(3), pp. 247-267. DOI: 10.17705/1thci.00190 Available at http://aisel.aisnet.org/thci/vol15/iss3/1

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers & Education: Artificial Intelligence, 2. https://doi.org/10.1016/j.caeai.2021.100041. Article 100041.

Othmana, N. & Amiruddin, N. O. (2010). Different Perspectives of Learning Styles from VARK Model. *Procedia Social and Behavioral Sciences,* 7(C) pp. 652–660

Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Towards a catalog of prompt patterns to enhance the discipline of prompt engineering. *ACM Sig Ada Letters*, 42(2) pp. 43-51.

Shontell, A. (2025). Fortune 500 digest, November 1, view.mail.fortune.com viewed November 1, 2025.

Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, *4*, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. Computers & Education: Artificial Intelligence, 3. https://doi.org/10.1016/j.caeai.2022.100065. Article 100065.

Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research*, 34(2), 1–18

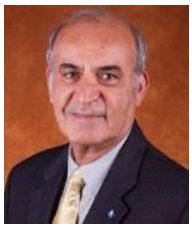
The World Economic Forum, Future of Jobs Report 2025. https://www.weforum.org/publications/the-future-of-jobs-report-2025/, accessed December 20th, 2025.

Tolzin, A., Knoth, N., & Janson, A. (2024). Leveraging Prompting Guides as Worked Examples for Advanced Prompt Engineering Strategies. ICIS 2024 Proceedings. 1. https://aisel.aisnet.org/icis2024/learnandiscurricula/learnandiscurricula/1

Wang, P., Jing, Y., and Shen, S. (2025). A systematic literature review on the application of generative artificial intelligence (GAI) in teaching within higher education: Instructional contexts, process, and strategies. *The Internet and Higher Education*, Volume 65, doi.org/10.1016/j.iheduc.2025.100996.

Williams, K. (2025). The autonomy advantage in business education. AACSB Career Connection, https://www.aacsb.edu/insights/articles/2025/08/the-autonomy-advantage-in-business-education, accessed, November 29, 2025

Author Biographies



Rassule Hadidi is Dean of the College of Business and Management, Metro State University, Minneapolis, Minnesota. His research areas of interest include 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.

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.

This page intentionally left blank

Journal of the Midwest Association for Information Systems

Volume 2026 | Issue 1 Article 2

Date: 01-31-2026

AI-Assisted Bettors: Analyzing AI-Driven Betting Behavior through Cluster Analysis

Dr. Steven A. Schilhabel
University of Wisconsin Oshkosh, schilhabels@uwosh.edu

Abstract

The use of artificial intelligence (AI) systems in consumer-facing decision-support systems (DSS) such as predictive analytics and automated recommendation platforms is growing in popularity in numerous domains, including sports betting. The degree to which users interact with the output of automated systems, calibrate trust, and exhibit automation bias–consistent behavior is largely unknown. In this study, we investigate the behavioral segments formed by bettors using AI-powered predictive sports betting DSS based on their shifts in confidence, risk-taking behavior, and bankroll management practices.

We use survey data from a sample of 200 U.S.-based bettors and SPSS TwoStep Clustering to identify three distinct behavioral profiles: Traditional Bettors, AI-Influenced Confident Bettors, and AI-Adopting Risk-Takers, each with their own unique set of interactions with and through predictive DSS. The findings show that bettors can engage in responsible adoption through strategic bankroll management practices and a tempering of AI trust, while overreliance behaviors can be mitigated or amplified, respectively, by counter or co-aligning with individual differences.

Framing betting platforms as real-time, in-the-wild, and consumer-deployed DSS contributes to IS research on algorithmic decision environments, user trust, and human–computer interaction. The results advance IS theory by contributing to the discussion of how cognitive biases, human decision behaviors, and confidence amplification in and through automated systems manifest in such domains. We conclude the paper with implications for responsible DSS design and deployment as well as practical guidelines for user segmentation in predictive analytics DSS environments.

Keywords: Design Science, AI-Assisted Betting, AI-Driven Decision-Making, Gambling Psychology, Machine Learning in Betting.

DOI:10.17705/3jmwa.000097

1. Introduction

Artificial intelligence (AI) has seen a swift uptake as an integral component of decision-support systems (DSS), facilitating enhanced processing power, accuracy, and speed across a variety of domains and industries, including banking, healthcare, and logistics (Power, 2002; Davenport & Harris, 2007). Betting tools have been extended to the consumer with the integration of artificial intelligence, and the intersection of real-time odds calculations, predictive modeling, and personalized risk assessments reflects several key DSS components. As bettors increasingly rely on predictive platforms to inform wagering decisions, they are engaging with consumer-facing DSS characterized by data-driven recommendations that leverage structured data and probabilistic forecasting to guide their betting decisions. In other words, users of AI-powered tools interface with AI-assisted DSS for their decision-making tasks, an IS-related feature observed in professional settings.

The gamification of gambling has also created and sustained two opposed bettor categories. Recreational bettors participate in sports to bet on them. They are aware of the poor return on investments for most bets; however, they enjoy the risk-taking as a component of their entertainment (Humphreys & Perez, 2012). The social norms of this culture involve rec bettors. These gamblers enjoy the social aspect of team loyalty and playing with their "friends." Rec bettors also rely on general football knowledge as a precursor to wagers, often employing an intuitive approach in their betting strategies (Humphreys & Perez, 2012). In contrast, sharp bettors use statistical modeling to find value and make bets for long-term profit. To achieve profitability, sharp bettors play the sports betting market as a long-term investment opportunity, implementing bankroll management strategies and algorithms to uncover market inefficiencies (Wong, 2001; Šír & Lábaj, 2021; Donahue, 2022). Market inefficiencies occur when an investment, like a wager, has a higher expected value than the original cost of the investment (Donahue, 2022). Smart bettors use money management and hedging techniques to minimize risks and lower the volatility of their bankrolls.

In more recent years, a new hybrid category has emerged, AI-Assisted Bettors. AI-assisted bettors utilize AI in their betting decision-making, but not to the extent of fully implementing all the components used by sharp bettors (Galekwa et al., 2025; Skrill, 2024). To fully understand the basis of each bettor category, one must look at the decision support with which each category of bettors interacts. Rec and sharp bettors have begun using more AI-empowered betting tools, and there is growing evidence that bettors using these tools show a higher volume of bets per day and overconfidence in their betting decisions (Elder et al., 2022). Decision theory and human–computer interaction (HCI) research offer insight into the relationship between user behavior and the use of DSS. Excessive trust in the output of machine learning systems, overconfidence, and the over-reliance on automation without independent verification are known as automation bias (Mosier & Skitka, 1996; Elder et al., 2022). Cognitive and social psychology research has been used to explain the decision-making processes that bias humans when making financial choices in situations with uncertain outcomes (Kahneman & Tversky, 1979; Nickerson, 1998; Tversky & Kahneman, 1971). The behavioral tendencies of people making wagers that rely on AI DSS could be a useful area of exploration. How do bettors trust AI-powered prediction systems? The AI-powered betting platforms function as consumer-deployed DSS in these types of use cases and can be analyzed with IS principles for this work. This raises an interesting IS research question: how do people make decisions in predictive systems where they balance human and AI augmentation?

The use of IS to understand how to develop artifacts, how people use artifacts, and how artifacts can be integrated and networked for decision-making has a long history. A behaviorally informed IS would apply relevant human and HCI research to how users of predictive betting platforms think, act, and relate to one another. The use of information systems in gamification technologies has a significant history, but the specific use of DSS to inform gambling decisions within betting apps has not been a focus of prior work.

Gambling on sports has a range of purposes in American life. It can serve as entertainment, a hobby, a source of income, or an identity-defining process for professional bettors. The accessibility of betting in most American states has fostered a diverse user base that would be the prime subjects of clustering analysis in IS. As user-generated content, online betting forums, such as Reddit, have been used to capture some initial descriptions of bettor types with the use of latent semantic analysis (Salari et al., 2022). However, the adoption of DSS has not been used to investigate bettor profiles. This raises an interesting question for IS research. With significant consideration given to the integration of decision environments, how do people think when choosing from predictive DSS, and what role does decision-making theory play in these environments? We know that people tend to fall into categories of sharp or recreational. We also know that bettors are using AI-powered systems to inform betting decisions. However, the intersection of identity with these platforms has not been explored. This study approaches these issues by investigating a US-based sample of 200

active AI bettors to study and segment their use of AI-based betting platforms using SPSS Two-Step Cluster Analysis.

Aim: In this paper, we investigate the heterogeneity in confidence, risk-taking, and strategic engagement among bettors who employ AI-augmented decision-support tools. We seek to answer the following IS research question: "What are the patterns of confidence, risk-taking, and AI reliance that lead to different user segments in bettors using predictive decision-support systems?"

Method: We leverage survey-based data from 200 US bettors and unsupervised clustering methods to identify unique profiles of AI interaction – bettors who are using AI as a delegate for their decisions, confidence amplifier, or a secondary resource. We discuss implications for understanding how users trust and calibrate automated recommendations in high-stakes decision contexts and for designing and deploying responsible DSS in consumer-facing settings.

2. Literature Review

One of the original foci of IS scholarship has been on the intersection of AI and decision support systems (DSS). The classic DSS literature has analyzed how structured decision environments "augment" human judgment with data-driven decision support, predictive modeling, and ex ante scenario evaluation (Power, 2002; Davenport & Harris, 2007). At the same time, the expanded availability of consumer-facing AI has pushed the DSS "ecology" beyond its traditional enterprise use cases (Jarke et al., 2019; Rai, 2020). Commercially-deployed DSS that directly support consumer micro-decisions now include various types of AI-powered betting systems, which provide both real-time forecasts and probabilistic confidence estimates, automated "expert" recommendations, and multiple-choice scenario selection. Betting platforms thus present a valuable real-world use case for examining DSS theory at the intersection of prediction markets, AI, and consumer behavior.

A more recent IS research stream has highlighted how users engage with machine-produced advice and interact with AI models, focusing particularly on the calibration of algorithmic trust (Glikson & Woolley, 2020; Jussupow et al., 2022). If a user lacks the ability or motivation to understand a particular AI output, or if they misperceive the capabilities of the system as being more or less accurate than it actually is, their level of trust will be miscalibrated. Excessive trust in an AI system has also been conceptualized as "automation bias" (Mosier & Skitka, 1996; Lee & See, 2004). In IS, several experimental studies have demonstrated that over-trust in algorithms leads to less critical processing, delegation of effort and attention, and even responsibility-shifting from the human user to the machine (Logg et al., 2019; Castelo et al., 2022). By contrast, people can also avert their trust in machine suggestions in the presence of AI errors, even when the system is demonstrably more accurate than most humans (Dietvorst et al., 2015). As a final caveat, DSS are perhaps more likely to be adopted for consumer-oriented decision making where users have more limited capacity to check system accuracy and access system explanations.

In a parallel literature within HCI and IS design science traditions, a small number of recent studies have turned attention to individual differences in how users respond to AI tools. This work highlights the role of moderators in how end users think and act in decision environments—extending the basic propositions about calibration, confidence, and mental effort (Shin, 2021; Ehsan & Riedl, 2020). In addition to varying by system type, user experience, and self-perceived expertise, AI-enabled decision-making can also boost confidence when a system makes a recommendation—even if those predictions are uncertain or probabilistic in nature (Jussupow et al., 2022). This so-called "confidence boost effect" has been previously documented in IS research on various types of financial and medical decision support (Söllner et al., 2021; Longoni et al., 2019; Pourebrahim et al., 2023). Empirical work in this area has not yet been done on fast-paced, high-stakes domains like sports betting, despite high levels of heterogeneity in bettor behavior, psychology, and sophistication (Hägg et al., 2022).

Betting behavior has of course been extensively studied in the domain of behavioral finance, usually by reframing conventional cognitive biases like Prospect Theory (Kahneman & Tversky, 1979), overconfidence bias (Elder et al., 2022), or the Gambler's Fallacy (Tversky & Kahneman, 1971) in the context of sports betting (Hur et al., 2018; Bechmann et al., 2023; Yildiz & Gurel, 2014). The work in this thesis augments this research by examining how digital technologies can shift how those biases are expressed, and how trust and confidence are calibrated with respect to personalized, algorithmic predictions. As users shift more of the cognitive load to betting platforms, their cognitive framing of gains and losses, perceived risk, and sensitivity to feedback can all change. For example, algorithmic forecasts may introduce or amplify confirmation bias, as bettors seek out system outputs that confirm their prior beliefs (Nickerson, 1998). Personalized recommendations may also help reframe decisions from deliberative to reactive, if users act on expert suggestions less critically or treat the system as the default choice (Shin, 2021).

Finally, a large subfield of IS research on user segmentation and personalization may also be relevant to this study.

Prior work has performed cluster analyses to segment users by behavior, cognitive style, or trust dispositions across multiple use cases—including e-learning systems (Gupta & Pathak, 2020), insurance claims (Leidner et al., 2017), and even algorithmic pricing tools (Castelo et al., 2022). These frameworks all provide a lens for empirically understanding heterogeneity in how users approach AI-powered decision environments. In the case of AI-assisted betting, user clustering can be useful not only for distinguishing between frequent and infrequent users, heavy and light users, or system-dependent and system-independent users, but also for profiling user types that adopt different strategic postures (e.g., risk vs. reward orientation) and self-regulation tactics (e.g., confidence calibration, bankroll management).

In sum, the current IS literature on AI and DSS offers a rich theoretical framework for understanding both the generic mechanisms and individual differences in how AI-powered betting systems function as consumer-facing DSS. The literatures also justify the present research and methodological need to empirically test for segmentation into different types of bettors based on a combination of behavioral and psychological measures—especially in an algorithmic setting where end users are presented with a high degree of confidence, placed at risk with real money, and lack full system transparency.

3. Methodology

Survey Instrument

Items were selected from a larger survey instrument used in a larger research project on generative AI and sports betting behavior. Items were mapped to four primary behavioral constructs of interest for this analysis:

1. AI reliance,
2. Confidence shift post-AI adoption,
3. Risk-taking behavior, and
4. Bankroll management practices.

These constructs were selected based on prior research on decision-support technology (Mosier & Skitka, 1996), automation bias (Elder et al., 2022), and gambling studies (Kahneman & Tversky, 1979; Donahue, 2022).

Respondents provided ratings to a series of items using a five-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Many items were included in a pre/post format to identify changes in behavior post-adoption of generative AI tools, such as ChatGPT. For instance, users were asked to provide agreement ratings to paired statements such as:

- Pre-AI: “I followed a strict bankroll management plan.”
- Post-AI: “I followed a strict bankroll management plan.”

Responses to these items were used to create change-score variables for cluster analysis (e.g., Δ confidence, Δ bankroll discipline)

Post-AI variables also included:

- Confidence amplification: “I feel more confident in sports betting after using AI.” (Q15_4)
- Increased betting activity: “I increased the number of bets I make.” (Q15_1), “I increased the average size of bets I make.” (Q15_2)
- AI interaction type: “I use generative AI to analyze raw data,” “I use paid or free AI analytics platforms to make decisions.” (Q11 block)
- Self-identification: “I consider myself more of a recreational, semi-sharp, or sharp bettor after using AI.” (Q16 block)

Demographics, AI experience, and fantasy app use items were also included, along with attention-check items to screen for data quality. The instrument was IRB-approved by the University of Wisconsin–Oshkosh and fielded on the

CloudResearch platform to pre-screen U.S. adults.

The survey's constructs and representative items are listed in Table 1.

Table 1 - Survey Constructs and Representative Items

Construct	Measurement Source	Sample Items (Likert Scale)	Pre/Post Format?
AI Reliance	Developed for this study (Q11, Q12)	"I use generative AI tools to analyze raw data for sports bets. "I used paid analytics to inform bets."	Post only
Confidence Shift	Developed for this study (Q15_4)	"I feel more confident in sports betting after using AI."	Post only
Bankroll Management	Adapted from prior gambling strategy literature (Q8_6 / Q9_6)	"I followed a strict bankroll management plan."	Pre & Post
Risk-Taking Behavior	Adapted from financial decision literature (Q15_1, Q15_2)	"I increased the number of bets I make. "I increased the average size of bets."	Post only
Sharp/Semi-Sharp Identity	Adapted from Donahue (2022); custom developed (Q16)	"I consider myself more of a sharp bettor. "I consider myself more of a semi-sharp bettor."	Post only
AI Perceived Competence	Developed for this study (Q12_1, Q12_2)	"I understand what generative AI is. "I understand how to use generative AI."	Post only
Betting Frequency	Behavioral indicator (Q10_1)	"How many sports bets did you place per week on average in 2023?"	Post only

Notes: All items were measured on a 5-point Likert scale unless otherwise noted. Pre- and post-difference scores were used to assess changes in constructs such as bankroll management and strategic alignment.

Data Collection

The Connect platform from CloudResearch served as the recruitment system for participants, enabling high-quality data collection (Hartman et al., 2023). The research initially recruited 310 U.S. adult participants. However, after screening for AI usage and other quality checks (attention check, time to complete, etc), 200 participants who actively used AI in their sports betting decisions were retained for analysis. The final sample included participants between 21 and 65 who identified as 62.3% male, 36.8% female, and 0.9% non-binary or other genders. Participants received a nominal payment from CloudResearch after finishing the survey, which required 10 to 15 minutes. The survey gathered data regarding bettors' dependence on AI systems and monitored their betting frequency and changes in confidence and financial control practices.

Cluster Analysis Approach

The research utilized SPSS TwoStep Cluster Analysis to categorize bettors according to their AI usage patterns and confidence shifts in betting behaviors, as this method automatically identifies the optimal number of clusters and handles both categorical and continuous data types (SPSS, 2001; Bacher, Wenzig, & Vogler, 2004; IBM, 2025). Three distinct clusters emerged from the analysis, categorizing bettors according to their use of AI systems for decision-making and their tendency to take risks. The study examined variables of AI dependency, betting frequency, confidence variations, and bankroll management principles. The Bayesian Information Criterion (BIC) analysis proved that the three-cluster solution provided the optimal data fit.

Validation Tests

The research evaluated the clustering solution's robustness by applying chi-square tests to measure categorical variations between clusters while using ANOVA to examine continuous variable mean differences across clusters. Statistical tests demonstrated significant differences between groups in AI reliance, betting behavior, and confidence

shifts ($p < 0.05$), which validates the cluster identification.

4. Results

Resulting Clusters:

The cluster analysis identified three groups of bettors based on their AI usage, betting confidence, and risk-taking tendencies. Traditional AI Bettors showed limited engagement with AI insights, maintaining lower betting frequency and small changes in confidence levels after adopting AI technology. These bettors preferred to trust their instincts and basic evaluation methods instead of using AI analytics. Bettors who moderately used AI tools reported higher confidence in their bets due to the insights provided by AI technology. Through their structured betting system, they continued to be cautious by combining AI-generated data with traditional research methods to guide their decisions. AI-driven risk-takers exhibited the highest reliance on AI-generated insights. They used AI regularly to inform their betting decisions. Members of this category placed bets more frequently and at higher stakes, exhibiting heightened confidence following AI integration. The increased risk-taking tendencies of these bettors reduced their ability to maintain disciplined bankroll management compared to AI-assisted bettors, who follow more strategic approaches.

Cluster One – Traditional Bettors

Traditional Bettors are more risk-averse and cautious participants in the betting world. Traditional Bettors use AI tools less because they base their betting choices on intuition, personal experience, and publicly available information. Traditional Bettors participate in wagering activities only occasionally because they treat betting as an informal hobby instead of a strategic or analytical exercise. Their trust in betting decisions does not increase significantly following the introduction of AI tools because they either omit AI from their strategies or implement it only marginally. Traditional Bettors avoid experimenting with AI-generated insights and do not modify their betting approaches based on external analytical data. This group maintains its resistance to technological advances by adhering to traditional sports betting methods that do not rely on data-driven techniques.

Cluster Two – AI-Influenced Confident Bettors

AI-Influenced Confident Bettors form a bridge between Traditional Bettors who resist technological change and aggressive AI adopters who fully embrace AI tools for betting. They moderately use AI tools during their betting processes by applying AI insights to improve their decision-making while avoiding complete reliance on automated systems. Incorporating AI tools into their strategies lets these bettors evaluate matchups and value bets more effectively while reducing uncertainty, which results in higher confidence levels than Traditional Bettors. These bettors preserve a systematic betting strategy by combining AI analysis with traditional methods, including personal research and expert viewpoints. This group practices careful bet sizing and bankroll management to avoid reckless gambling despite using AI, unlike high-risk AI-adopting bettors. These bettors demonstrate strong potential to develop into AI-assisted bettors through their balanced use of AI technology and disciplined betting practices.

Cluster Three – AI-Adopting Risk Takers

The most aggressive betting group regarding AI applications and wagering patterns consists of AI-adopting risk-takers. The group depends on AI to make betting choices without verifying predictions through manual examination. AI-powered insights boost their confidence, which results in a noticeable rise in the number of bets they place while increasing their wager volume. The increased confidence from AI-generated predictions results in these bettors taking bigger risks by making substantial wagers without traditional research or expert advice.

AI-Influenced Confident Bettors maintain structured decision-making processes, whereas AI-Adopting Risk-Takers display impulsive betting patterns that deviate from bankroll management standards. Gamblers who rely too heavily on AI tools develop unwarranted trust in AI predictions, as they believe these predictions cannot make mistakes.

5. Key Findings

To better understand the differences among AI-assisted bettors, Table 2 provides a comparative overview of the three

identified clusters: Three distinct groups of bettors emerge when studying AI-assisted betting patterns: Traditional Bettors, who use minimal AI support; AI-Influenced Confident Bettors, who blend AI insights with their decisions; and AI-Adopting Risk-Takers, who fully embrace AI technology. The differentiation between clusters relies on their use of AI, level of betting confidence, approach to risk-taking behavior, bankroll management strategies, and decision-making

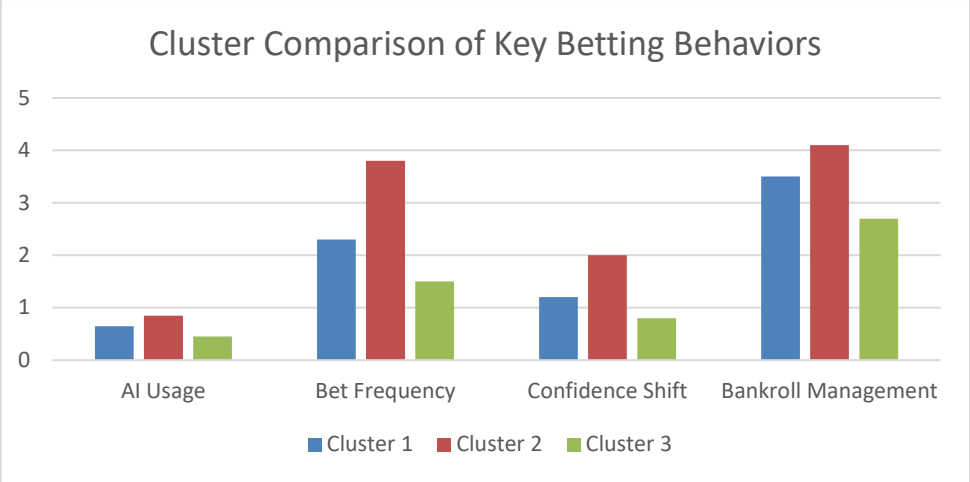


Figure 1 - Cluster Comparison of Key Betting Behaviors

Notes: Figure 1 displays the predictor importance scores from the SPSS TwoStep Cluster Analysis, indicating the relative contribution of each variable to the formation of the three-cluster solution. Higher values signify a stronger influence on cluster differentiation.

methodologies. Traditional Bettors utilize AI tools minimally because their approach is mainly driven by intuitive judgment. AI-Influenced Confident Bettors use AI analysis to guide their decisions but remain strategic and balanced in their betting approach. AI-Adopting Risk-Takers rely entirely on AI-generated forecasts, which increases their confidence but often leads to reckless betting practices. Identifying these distinct groups highlights AI's impact on betting behaviors and how various bettors utilize AI within their betting strategies. The primary attribute of these bettor clusters involves bankroll management, which represents the methods people employ to handle their betting money (Intellias, 2025). These bettors who practice solid bankroll management distribute their betting money carefully while establishing betting boundaries and varying stakes according to risk factors. Bettors with poor bankroll management often place bets driven by overconfidence, engaging in high-risk wagering strategies.

Table 2 - Clusters Defined

Cluster	**AI Usage**	**Betting Confidence**	**Risk-Taking Behavior**	**Bankroll Management**	**Decision-Making Approach**
Traditional Bettors	Low	Minimal change after AI adoption	Risk-averse, occasional betting	Cautious, conservative	Intuition-based, relies on personal experience and public data
AI-Influenced Confident Bettors	Moderate	Increased confidence with AI insights	Balanced risk-taking, controlled bets	Disciplined, strategic	Uses AI insights but verifies with traditional research and expert opinions
AI-Adopting Risk-Takers	High	Overconfident due to AI reliance	Frequent betting, larger stakes	Poor, often reckless	Fully depends on AI-generated predictions without manual verification

The use of AI-driven confidence significantly influences predictions about AI-assisted behavior. Following AI adoption, bettors who showed increased confidence levels shifted towards data-driven betting strategies, utilizing insights generated from AI systems to make their decisions. Bettors who relied solely on AI-generated confidence failed to achieve strategic success because they neglected disciplined betting practices. Not all AI users become AI-assisted bettors. AI-assisted bettors distinguish themselves through their capability to integrate elevated AI confidence levels with effective bankroll management techniques. A segment of AI users engaged in reckless betting practices, while another group used methodical decision-making patterns to demonstrate the critical nature of disciplined financial planning for success. These key findings are summarized in Table 3.

Table 3 - Summary of Key Findings

Key Finding	Description
AI-Driven Confidence	Increased confidence correlates with data-driven betting shifts.
Bankroll Management as a Differentiator	Strategic bankroll management separates AI-assisted bettors.
Paid AI Tools as a Stronger Predictor	Paid AI users are more likely to transition into AI-assisted bettors.

This study's conclusions provide essential guidance for IS creators and organizations that utilize AI-based decision-support systems. IS developers need to create features that encourage users to critically assess AI recommendations due to the demonstrated overconfidence in AI-assisted betting. AI-powered decision environments like financial forecasting and automated investment platforms face similar challenges. When organizations create AI systems that improve user comprehension of probabilistic results and risk assessment, they reduce automation bias risks and enhance decision-support capabilities.

Cluster Validation and Model Selection

Researchers analyzed the clustering solution with ANOVA and chi-square tests, demonstrating that the groups have distinct behavioral patterns. The ANOVA analysis presented in Table 4 reveals significant differences in Perceived AI Competence (10_1) and Betting Frequency (15_5) among clusters, where p-values are less than 0.001, indicating that these continuous variables effectively separate different bettor segments. Chi-square tests (Table 5) validate the cluster structure by showing significant associations ($p < .001$) across the categorical variables Trust in AI Recommendations (11_3), Risk-Taking Behavior (15_1), and use of AI Tools (16_2).

The SPSS Two-Step Clustering process utilized the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) to establish the best number of clusters, resulting in a three-cluster solution that optimally balances model complexity with interpretability. Figure 1 demonstrates which key predictors drive cluster formation and identifies AI reliance (16_2), Confidence Shift (15_4), and Betting Frequency (15_1) as the top influential variables.

Table 4 - ANOVA Results

Variable	Cluster 1 Mean (SD)	Cluster 2 Mean (SD)	Cluster 3 Mean (SD)	F-Value	p-Value	Effect Size (Eta²)
Perceived AI Competence (10_1)	3.27 (1.51)	4.63 (1.51)	5.38 (1.41)	30.474	**<.001**	0.238
Betting Frequency (15_5)	4.20 (1.10)	5.05 (0.73)	5.44 (0.92)	28.533	**<.001**	0.226

Table 5 - Chi-Square Test Results

A	Chi² Value	df	p-Value
Trust in AI Recommendations (11_3)	60.543	8	**<.001**
Risk-Taking Behavior (15_1)	124.715	8	**<.001**
Betting Strategy Type (15_4)	138.072	8	**<.001**
Use of AI Tools (16_2)	142.151	8	**<.001**
Willingness to Adopt AI (16_3)	105.59	8	**<.001**

The predictor importance scores quantify the extent to which each variable contributes to distinguishing the identified clusters. SPSS TwoStep Cluster Analysis calculates scores between 0 and 1 for each variable's impact on cluster formation. Scores that approach 1 show a variable's greater impact on cluster formation. Variables receive lower scores (approaching 0) because they contribute minimally to cluster separation.

The scores shown in Figure 2 illustrate the contribution of each variable towards creating the three-cluster solution. The variables AI reliance (16_2), Confidence Shift (15_4), and Betting Frequency (15_1) achieved the highest scores, which demonstrated their significance in differentiating bettor types.

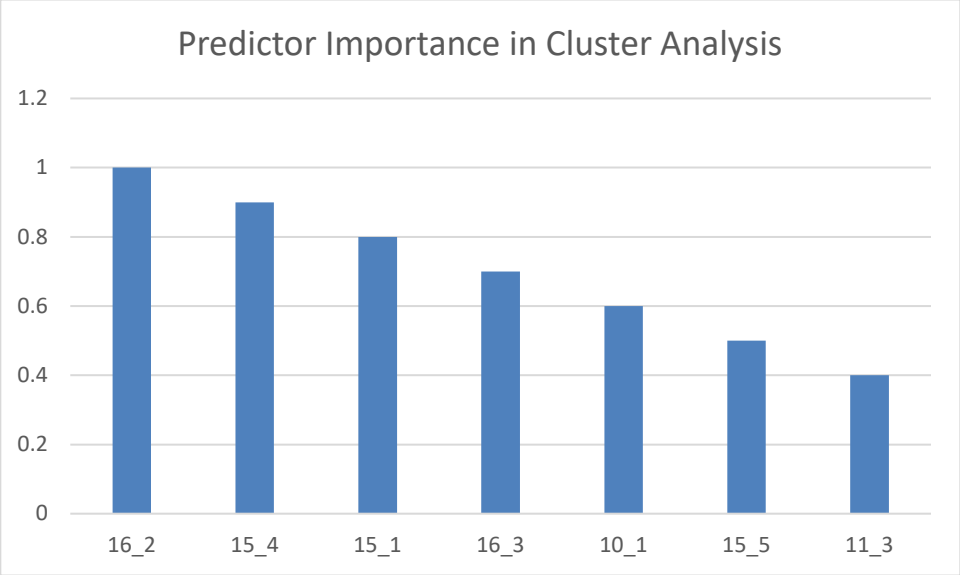


Figure 2 - Predictor Importance in Cluster Analysis

Notes: displays the predictor importance scores from the cluster analysis, highlighting the relative influence of key behavioral variables in distinguishing between bettor segments. The variables include Use of AI Tools (16_2), Betting Strategy Type (15_4), Risk-Taking Behavior (15_1), Willingness to Adopt AI (16_3), Perceived AI Competence (10_1), Betting Frequency (15_5), and Trust in AI Recommendations (11_3). Higher scores indicate a stronger role in shaping the three-cluster solution, emphasizing differences in AI reliance, betting behaviors, and strategic decision-making.

6. Discussion

Information systems research sees a significant evolution through the adoption of AI-powered predictive analytics in sports betting for data-driven decision-making. Research in Information Systems demonstrates that decision-support systems improve organizational decision-making capabilities by utilizing structured data analysis and predictive modeling techniques (Davenport & Harris, 2007). AI delivers these functionalities to decision-makers through instant

market analysis and customized betting recommendations. The research findings illustrate how AI-powered information systems applications shape bettor behavior by increasing their decision confidence and creating segmentation patterns comparable to those in business intelligence systems for enterprise analytics.

This research advances human-computer interaction knowledge in information systems by studying user interactions with decision tools powered by artificial intelligence. Studies in information systems have demonstrated that decision-making in critical situations is influenced by automation bias and dependence on AI recommendations, as noted by Mosier and Skitka (1998). Bettors show increased trust in AI forecasts, which causes them to change their risk-taking behavior and betting strategies. This study examines how AI influences decision-making processes, which adds to IS literature by exploring HCI elements, algorithmic trust concepts, and how users adapt their behavior.

Theoretical Contributions

This study makes several contributions to the IS theory literature on DSS, algorithmic trust calibration, and user segmentation in AI-augmented decision environments.

First, the current work contributes to DSS theory by utilizing the contextual domain of predictive, AI-augmented, high-stakes, real-time decision support tools to investigate the deployment of such decision support principles in a new context: namely, AI-augmented betting platforms for consumer decision-makers. While much of the extant DSS literature has focused on formally constrained and regulated enterprise decision environments (Power, 2002; Arnott & Pervan, 2014), the current work provides novel evidence that the DSS constructs of probabilistic modeling, recommendation logic, and real-time processing now power betting apps and websites that are used by individual consumer actors, and by users in a behavioral environment that is far less regulated and less predictable than the decision context for which DSS systems were initially developed. This study thus helps to close a gap in the IS consumerization of enterprise technology literature (Alder et al., 2018) by empirically articulating the core principles of DSS as they are currently implemented to support individual, real-time decisions in an environment where user behavior is more complex and variable than originally theorized.

Second, the current study contributes to IS theories of algorithmic trust calibration and automation bias by empirically identifying three latent user segments that map onto three unique orientations towards the system. Traditional Bettors, in showing aversion to using the system in the first place, are indicative of the DSS phenomenon of algorithm aversion (Dietvorst et al., 2015). AI-adopting risk-takers, by placing their trust in the system's recommendations, are demonstrating a lack of calibration and exhibiting automation bias (Mosier & Skitka, 1996). AI-Influenced Confident Bettors, in contrast, have shown well-calibrated trust in the predictions of the system (Jussupow et al., 2022). The differences among these three user groups thus provide an important behavioral validation of a set of extant IS theories that describe user trust in algorithmic systems as non-uniform, dynamic, context-specific, and psychologically mediated (Glikson & Woolley, 2020).

Third, by employing unsupervised clustering to detect latent subgroups of users based on differences in AI usage, shifts in confidence levels, and self-regulation of gambling behavior, the study contributes to IS literature on user segmentation and personalization in technology-mediated decision environments. This methodology builds on prior IS work on user segmentation, which in this context is used to model heterogeneity in system adoption decisions, systems interaction style, and risk preferences (Gupta & Pathak, 2020; Leidner et al., 2017). In doing so, it provides a replicable method for DSS and HCI scholars interested in how different types of users engage with and experience semi-autonomous, complex decision support systems.

In this way, this study advances IS theory by providing empirical evidence that connects key aspects of user cognition in AI-augmented systems—trust in AI and self-regulatory behavioral differences—to real-world engagement with complex decision-support technologies. The resulting theoretical contributions provide an explanatory lens with which to view current user behavior in DSS contexts, as well as normative design implications for the more responsible development of AI-based systems.

Practical Implications

Results demonstrate numerous implications for the operation of sportsbooks and betting platforms. Betting platforms

can develop AI-driven tools that cater to AI-assisted bettors, improving their strategic decision-making abilities while promoting responsible gambling practices. Sportsbooks can create premium AI analytics services for bettors who use AI in a structured manner to place confident bets.

Sportsbooks should modify their betting odds or enforce limits to address the heightened risk-taking tendencies of bettors using AI technologies. Sportsbooks should consider implementing personalized betting limits and educational tools about bankroll management for groups known for aggressive betting practices. Platforms can optimize engagement methods through better segmentation analysis while balancing user experience and risk management.

Limitations & Future Research

This study makes important contributions but also presents numerous limitations. The dataset depends on self-reported survey responses, which could lead to response biases, especially in evaluating confidence and risk-taking behaviors. Actual betting data should be used in future research to validate these findings. The research captures bettor behavior during one specific moment while betting strategies develop dynamically. Longitudinal research would provide valuable insights into how AI-assisted bettors adjust or refine their betting strategies over time.

Future research should focus on regulatory issues as a major area of investigation. The increasing sophistication of AI-driven betting tools could prompt policymakers to pass new regulations influencing bettors' behavior. Research into how prospective AI regulations affect bettor segmentation will shed light on the changing dynamics between AI technology and gambling decision processes. While this study focuses exclusively on AI-using bettors, future research should compare AI and non-AI users to determine whether AI adoption significantly alters betting behavior or enhances pre-existing tendencies. Future investigations require analysis of AI-assisted bettors using new theoretical models to understand the effects of AI adoption on extended betting methods and segmentation of bettors. Subsequent research should examine how Generative AI influences betting decisions to establish a more robust theoretical base for understanding AI-based betting activities.

7. Conclusion

This study contributes to IS research by demonstrating how AI-powered information systems shape decision-making behaviors, reinforcing the importance of IS frameworks in predictive analytics, decision support, and Human-Computer Interaction (HCI). Advanced betting behaviors driven by AI technology are establishing new categories of bettors that extend beyond conventional classifications. This research demonstrates that AI-assisted bettors are primarily differentiated by their confidence levels, Betting Frequency, and dependence on AI support. AI-adopting risk-takers utilize technology to expand their betting activities and risk limits, whereas AI-influenced confident bettors maintain a systematic strategy that supports disciplined bankroll control.

AI advancements in sports betting demand a comprehensive analysis of bettor behavioral changes and the elements that create lasting, successful betting methods. Future investigations should examine the relationship between sportsbooks and AI systems in conjunction with changing bettor profiles, and analyze the extended financial and psychological outcomes of betting choices enhanced by AI technology. A deeper understanding of these dynamics will enable the development of responsible betting structures while also enhancing bettors' comprehension of AI-based decision-making processes.

8. References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Arnott, D., & Pervan, G. (2014). A critical analysis of decision support systems research. *Journal of Information Technology, 29*(4), 261–293. https://doi.org/10.1057/jit.2014.17

Bookmakers Review. (2025). *How offshore sportsbooks handle action from sharp bettors*. Retrieved from https://www.bookmakersreview.com/industry/how-offshore-sportsbooks-handle-action-from-sharps/

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical*

information-theoretic approach (2nd ed.). Springer-Verlag.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2022). Algorithm aversion and algorithm appreciation: Understanding user behavior in algorithmic decision-making. *Journal of Consumer Research, 48*(3), 455-471. https://doi.org/10.1093/jcr/ucab014

Daschner, S., & Obermaier, R. (2022). Algorithm aversion? On the influence of advice accuracy on trust in algorithmic advice. *Journal of Decision Systems, 31*(2), 193-210. https://doi.org/10.1080/12460125.2022.2070951

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126. https://doi.org/10.1037/xge0000033

Elder, L., Canfield, C., Shank, D. B., Rieger, T., & Hines, C. (2022). Knowing when to pass: The effect of AI reliability in risky decision contexts. *Human Factors, 66*(2), 348-362. https://doi.org/10.1177/00187208221100691

Galekwa, R. M., Tshimula, J. M., Tajeuna, E. G., & Kyandoghere, K. (2025). A systematic review of machine learning in soccer betting: Techniques, challenges, and future directions. *Preprints*. https://doi.org/10.20944/preprints202501.2060.v1

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Gupta, A., & Pathak, G. S. (2020). Clustering e-learners using clickstream data to optimize learning environments. *Information Systems Frontiers, 22*(6), 1333-1346. https://doi.org/10.1007/s10796-019-09956-3

Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). Introducing Connect by CloudResearch: Advancing Online Participant Recruitment in the Digital Age. *OSF Preprint*. https://doi.org/10.31234/osf.io/ksgyr

Humphreys, B. R., & Perez, L. (2012). Who bets on sports? A look at the characteristics of sports bettors. *Estudios de Economía Aplicada, 30*(2), 579-598.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2020). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *arXiv*. https://arxiv.org/abs/2010.07487

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. (Ensure cited if used in the manuscript)

Mosier, K. L., & Skitka, L. J. (1996). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 6*(2), 101-116. (Make sure page numbers are correct if you cite this; if not, use the 1998 version you have.)

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Power, D. J. (2002). *Decision support systems: Concepts and resources for managers*. Quorum Books.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

https://doi.org/10.1214/aos/1176344136

Shin, D. (2021). The effects of explainability and causability on human–machine trust in AI. *Computers in Human Behavior, 123*, Article 106878. https://doi.org/10.1016/j.chb.2021.106878

Skrill. (2024). *AI and the bookie: How artificial intelligence is helping transform sports betting*. Skrill News. Retrieved from https://www.skrill.com/en-us/skrill-news/gaming/ai-and-the-bookie-how-artificial-intelligence-is-helping-transform-sports-betting/

Söllner, M., Bitzer, M., & Janson, A. (2021). Digital nudging in decision support systems: The effect of interface design on the use of analytical information. *Journal of the Association for Information Systems, 22*(4), 1102-1136. https://doi.org/10.17705/1jais.00702

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110. https://doi.org/10.1037/h0031322

Vélez Jiménez, R. A., Lecuanda Ontiveros, J. M., & Possani, E. (2023). Sports betting: An application of neural networks and modern portfolio theory to the English Premier League. *arXiv preprint*. https://arxiv.org/abs/2307.13807

Author Biography

Dr. Steven A. Schilhabel is an Assistant Professor of Information Systems at the University of Wisconsin Oshkosh, where he teaches courses in IS Essentials, IS Strategy and Cybersecurity. His research focuses on digital transformation, AI-enabled infrastructure, and design science approaches to instructional and collaborative systems.

His recent work includes peer-reviewed publications and conference presentations on human-centered AI, research collaboration frameworks, and educational technology, contributing to the field's understanding of applied and pedagogical innovations in IS.

Journal of the Midwest Association for Information Systems

Volume 2026 | Issue 1 Article 3

Date: 01-31-2026

Criminalizing Deepfakes: A Comprehensive Review of US Legislative Responses to Synthetic Media Abuse

J. Royce Fichtner
Drake University, royce.fichtner@drake.edu

Troy J. Strader
Drake University, troy.strader@drake.edu

Abstract

Deepfake technology can create fabricated media that realistically depicts individuals saying or doing things they never did. While deepfake technology has many positive applications, it also has many negative applications, such as child pornography, nonconsensual sexual imagery, political misinformation, and fraud. These negative applications raise profound legal, social, and ethical concerns, prompting calls for stronger regulation. This study reviews all state-level legislation in the United States between January 1, 2018, the approximate date the term "deepfake" was popularized, and May 19, 2025, the date Congress passed the Take It Down Act, its first law criminalizing negative deepfake applications. It aims to identify how states have utilized criminal law to confront harmful deepfake applications and to document the scope and content of legislative efforts. Through comprehensive database searches and thematic coding, the authors identified 89 pieces of enacted legislation in 41 states and categorized the prohibited activities into four areas: (1) child sexual abuse material deepfakes, (2) nonconsensual sexual deepfakes, (3) election deepfakes, and (4) other emerging uses such as commercial impersonation and fraud. Further analysis reveals bipartisan sponsorship, recent rapid legislative expansion, and significant variation in statutory definitions and scope. These findings highlight the fragmented nature of US deepfake criminal law and underscore the need for consistent definitions, expanded protections, and enhanced enforcement capacity. By documenting legislative trends and shortcomings, this study informs policymakers, legal practitioners, and scholars seeking to strengthen legal frameworks against synthetic media abuse.

Keywords: deepfake, regulatory framework, deepfake law enforcement

DOI:10.17705/3jmwa.000098

1. Introduction

In 2016, special effects forerunners from Industrial Light & Magic famously used digital technology to make a youthful Carrie Fisher appear to reprise her role as Princess Leia in a Star Wars film without any direct participation from the nearly sixty-year-old actress (Cotter, 2019). It was one thing for Hollywood filmmakers to utilize this futuristic technology, but it was quite another when less than a year later an anonymous Reddit user posting under the name "deepfakes" utilized an open-source machine learning framework to digitally overlay new Star Wars actress Daisy Ridley's face—along with faces of other celebrities—onto adult film actors' bodies in explicit video content (Harris, 2018; Spivak, 2018; Chesney & Citron, 2018). This was not done with a large Hollywood production budget; instead, readily accessible machine learning processes applied advanced algorithms to facial images of these celebrities to make it seem like they did something they never actually did. Months later, an app was shared on the same Reddit platform allowing users to create their own videos using "face swapping" (Chawla, 2019). The manipulation of images was not a new phenomenon, but the ability to do it so realistically in the more complex form of a video with relatively little effort made this remarkable.

A new trend, and the catchy term "deepfake," was born (Cole, 2018). Sexually explicit deepfakes proliferated very rapidly and studies show that the volume of sexually explicit deepfake content continues to grow at an alarming rate (Hawkins et al., 2025), with an 87 percent increase reported between 2022 and 2023 (ActiveFence, 2023). This trend is having a significant impact on minors. In the US 15 percent of high school students have heard of a sexual deepfake involving someone from their school (Wong, 2024). In the United Kingdom 13 percent of teenagers report having been the subject of a sexually explicit deepfake image (Internet Matters, 2024). But there are many negative applications beyond sexual deepfakes, where nefarious actors can use deepfake technology to manipulate public opinion, tarnish reputations, introduce false evidence in the courtroom, and perpetuate financial frauds. There are countless examples of malicious deepfake applications over the past several years. A Flemish political party posted a deepfake of President Donald Trump urging Belgians to withdraw from the Paris Climate Agreement (Burchard, 2018). A cybercriminal used audio deepfake technology to successfully deceive a CEO into authorizing a financial transfer of $243,000 (Kshetri et al., 2023). Strikingly realistic, unauthorized music generated using the voices of well-known artists have been appearing across the internet (Statt, 2020). A deepfake of Ukrainian President Zelenskiy urging surrender was broadcast during the early stages of the war between Russia and Ukraine (Twomey et al., 2023). And an AI-generated image of a fake explosion near the Pentagon caused a brief dip in the US stock market (O'Sullivan, 2023).

One of the major topics for information systems research is the many ways in which information technology positively and negatively impacts individuals, organizations, the economy, and society. Deepfakes are a relatively new information technology issue and, while they may have legitimate uses, they are too often used for malicious purposes that can harm individuals and organizations. Some information systems studies focus on technology applications that can reduce problems by detecting and/or eliminating deepfakes. One study reviewed more than a hundred published articles from 2018 through 2020 and found that deepfake detection technology may be based on a variety of methods encompassing machine learning, deep learning, statistical methods, and blockchain (Rana et al., 2022). While detection may reduce some of the potentially negative impacts, another perspective on the problem that has received less attention in information systems research is to assess the legal environment that may discourage deepfake misuse. This is the focus for the present study.

From the beginning, legal scholars warned that existing US civil and criminal laws were poorly equipped to address the unique challenges posed by this new synthetic media (Blitz, 2018; Harris, 2018; Hall, 2018). While a victim of deepfake technology could potentially bring a private civil lawsuit for claims such as defamation, false light, or intentional infliction of emotional distress, these remedies are limited by First Amendment free-speech protections and the immunities conferred by Section 230 of the Communications Decency Act, which largely shields online platforms from liability for distributing and sharing user-generated content (Chesney & Citron, 2019; Spivak, 2019). Beyond the legal hurdles to a successful civil lawsuit, there are practical challenges as well. First, it can be very difficult to determine who created and/or distributed the deepfake content because perpetrators might be known to the victim, like an ex-partner, coworker, or former classmate, or be complete strangers that merely found the video or image and then manipulated it through a readily available app on their phone. Even if the perpetrator can be identified, the victim might not have the financial means to hire an attorney to begin litigation (Flynn et al., 2021). Even worse, a successful plaintiff

might find little monetary remedy if the defendant did not have the financial means to pay a large civil judgment. Finally, "given the [humiliating] nature of many ... deepfakes, the adverse publicity that a public trial would generate will scare off many potential plaintiffs from filing a lawsuit." (Venema & Geradts, 2020)

Scholars have also noted that public, rather than private, enforcement mechanisms face their own set of challenges (Flynn et al., 2021). Criminal statutes targeting fraud, impersonation, harassment, or identity theft were not designed with synthetic media in mind and therefore contain technical loopholes to avoid conviction (Jodka, 2024). As noted by one scholar, "Privacy-based laws against distributing actual intimate images of people do not always cleanly apply to distributing synthetic images of them" (Grimmelmann, 2025). In order to apply many of these criminal laws to deepfake abusers the prosecution would need to prove that the perpetrator created and distributed the deepfake media with a malicious intent to cause harm. This is a high bar for successful prosecution as perpetrators may be able to argue they co-opted the victim's image for satire or their own personal enjoyment, rather than any desire to harass or harm the person (Venema & Geradts; Williams, 2024). Deepfake technology also created a new possible defense for those accused of possession of child pornography, which is more commonly referred to within the academic literature as "child sexual abuse material." Child sexual abuse material laws were written to protect real children, not necessarily synthetic images of children. A defendant charged with possessing child sex abuse material could potentially undermine prosecutorial efforts by arguing that they believed the material was artificially generated and did not depict real children. In many states, this loophole presents a challenge for law enforcement and prosecutors in proving harm when no identifiable child victim could be confirmed (Gorin, 2025; Willis, 2024).

On a more practical level, local law enforcement entities are ill-equipped to pursue potential deepfake criminal violations because many agencies lack the technical skill and resources to do so (Moloney et al., 2022; International Association of Chiefs of Police, 2025). This is compounded by the fact that the individual that created or distributed the deepfake could very likely reside outside the jurisdiction of the local agency (Chesney & Citron, 2019).

Recognizing these limitations, scholars have called for legal reform, including the creation of deepfake-specific criminal statutes (Kirchengast, 2020). Since the beginning the public has also clamored for legislative action in this area. A 2019 Pew Research Center poll found that the American public saw deepfakes as a major problem, with 77% of US adults favoring steps to restrict altered videos and images intended to mislead (Gottfried, 2019). Despite the pervasive calls for action, Congress has been very slow to pass any significant legislation addressing harmful deepfake applications. For the first seven years since the term deepfake entered the public discourse, the only legislation addressing negative deepfake applications enacted by Congress was the National Defense Authorization Act for Fiscal Year 2020 directing the Director of National Intelligence to submit a report to Congress discussing "the potential national security impacts of machine-manipulated media (commonly known as 'deepfakes')" and "the actual or potential use of machine-manipulated media by foreign governments to spread disinformation or engage in other malign activities" (U.S. Congress, 2019).

State legislatures have been much more proactive, with many moving to pass civil and criminal legislation regulating deepfakes within their own jurisdictions. Scholars have called for a more comprehensive national framework, covering AI model developers, providers, platforms, and creators, to curb fraud, political manipulation, nonconsensual sexual content, and other deepfake harms (Miotti & Wasil, 2024). In May of 2025 Congress finally enacted a notable piece of legislation addressing negative deepfake applications. The "Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act,'' otherwise known as the "Take It Down Act" (2025) addresses the nonconsensual sharing of intimate images generated or altered using deepfake technology. This act criminalizes the publication of such imagery and requires platforms to remove it within 48 hours of notification. While this act was an important first step, it was not the comprehensive federal law desired by many scholars. As a result, it merely added to the patchwork of existing state laws addressing negative deepfake applications.

Because the US is at a historic point in time where Congress has taken its first steps to criminalize at least one type of negative deepfake application, this study will pause to document how far state legislators have already gone to criminalize the use of deepfake technology. While legislators have also passed laws establishing civil liability for malicious deepfakes, this paper limits focus to legislation establishing actual criminal liability. This was done, in large part, because studies show a weak correlation between civil tort liability and deterrence (Simard, 2014; Cardi et al., 2012). Specifically,

this paper will investigate what legislation has been passed in the US to criminalize the use of deepfake technology through a systematic analysis of all deepfake-related state legislative enactments between January 1, 2018, the approximate date the term "deepfake" entered the public lexicon, and the passage of the Take It Down Act in May of 2025. It reviews the trajectory of state and federal criminal legislation targeting deepfake misuse, assesses the scope and focus of the legislation, and identifies legislative trends and remaining gaps. In doing so, it aims to assist scholars, legislators, and legal practitioners in navigating and strengthening the evolving legal framework around synthetic media.

The article is structured into six parts. After this introduction, section 2 provides background information, first explaining the basics behind deepfake technology, positive and negative applications of deepfake technology, and then discussing the recently enacted Take It Down Act, which serves as the first piece of US federal legislation criminalizing the use of deepfake technology. Section 3 is a literature review, highlighting studies about the negative consequences of deepfake technology, why deepfakes spready so rapidly through digital platforms, who is most likely to be deceived by deepfakes, technological efforts to detect deepfakes, and the regulation of deepfakes. Section 4 discusses the underlying methodology behind this study. Section 5 presents the results, with a chronological view of the legislation, followed by a brief discussion of any partisan trends in the sponsorship of enacted legislation, a grouping of the legislation into categories based upon the type of criminalized applications, and a discussion of the varying definitions of deepfake activities used within the enacted legislation. Section 6 provides a discussion and future research directions. The article ends with a brief conclusion.

2. Background

2.1 Deepfake Technology

Deepfakes can be simply defined as "videos, images, or other media that have been manipulated to appear as if the subject of the medium is speaking or partaking in an action that they did not actually undertake" (Mullen, 2022). Although scholars debate the precise definition of deepfake, a broad definition is that it constitutes audio-visual manipulation (Altuncu et al., 2024) using "a spectrum of technical sophistication . . . and techniques" (Brady, 2020). For many scholars the key component of a deepfake is the use of artificial intelligence, or deep learning based approaches (Altuncu et al., 2024). Deep learning is a type of machine learning where a computer analyzes datasets to find patterns (Altuncu et al., 2024). It is a branch of machine learning that models artificial neural networks on the structure of the human brain using unsupervised techniques, meaning it learns from unstructured data like audio and video to train generative models (Mubarak et al., 2023). Deepfake technology commonly uses machine learning to map the face or body of one individual onto the face or body of another individual (Suslavich, 2023).

Generative adversarial networks (GANs) are one method to generate deepfakes. GANs involve two neural networks working in tandem: one (the generator) creates an image, and the other (the discriminator) evaluates whether that image is real or fake (Davis, 2020). As described by Goodfellow et al. (2014), "[t]he generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency." As a result, "[c]ompetition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles" (Goodfellow et al., 2014). The discriminator scores the output's authenticity, prompting the generator to refine its output until the discriminator can no longer reliably tell the difference (Spivak, 2019). This technique supports "identity replacement," where one person's face is swapped onto another's body while preserving expressions and movements, and "identity reenactment," which directly alters the original subject's facial expressions and gestures (Spivak, 2019).

Initially, creating deepfakes required technical expertise and a powerful computer (Ajder & Schick, 2021). But accessibility has since grown significantly, and deepfake creation tools are now widely available, even via free smartphone apps (Masood et al., 2023). The emergence of user-friendly generation tools has enabled even non-experts to produce deepfakes, fueling an increase in harmful content, particularly involving nonconsensual sexual imagery

(Hawkins et al., 2025). Content is easy to obtain, often taken from social media profiles and images that are scraped without user consent. As a result, almost anyone with a computer or smartphone can access deepfake tools and train them on publicly available photos or voice data. The manipulation of images to cause harm is not a new phenomenon, but generative AI significantly reduces the effort required to produce highly realistic material and social media platforms amplify the scale of distribution.

2.2 Positive and Negative Applications

Despite their risks, deepfakes are not inherently harmful. Deepfake technology presents several promising benefits and real-world applications throughout many industries.

In healthcare, deepfake technology can enhance diagnostic accuracy, telemedicine, and mental health therapy through the use of synthetic data, virtual simulations, and AI-powered avatars. Its applications range from generating realistic medical images for AI training, improving surgical education with virtual reality platforms, providing emotionally responsive virtual therapists, to enabling grief counseling and memory reconstruction (Kaur et al., 2024). It can also help Alzheimer's patients, where animated representations of loved ones can offer comfort and emotional connection (Westerlund, 2019).

Deepfake technology also has many positive applications in film and television. It can reduce production costs and enable multilingual dubbing with greater ease and realism. For example, in a public service campaign promoting malaria awareness and supporting efforts to discover a cure, artificial intelligence-driven deepfake technology was employed to have international soccer star David Beckham appear to speak several different languages (Griffin, 2019). Deepfakes have also been used to resurrect deceased actors, where deepfake technology allowed a filmmaker to once again have Carrie Fisher reprise her role as Princess Leia in a Star Wars film, this time posthumously (Fiske, 2025).

Perhaps the greatest potential lies in education, where deepfake technology can build realistic training scenarios and make learning more interactive and engaging through techniques such as animating historical figures or generating realistic video lectures featuring well-known personalities (Nguyen et al., 2022). AI-generated avatars have also been used for accessibility tools, such as translating spoken language into sign language (Saunders et al., 2020) and it can also support language acquisition by simulating conversations with native speakers (Abernathy, 2024).

However, despite these promising uses, deepfake technology has become synonymous with harmful applications—particularly the spread of misinformation through social media. As noted by researcher Joshaua Habgood-Coote (2023), the 2017 Reddit incident sparked strong public narrative that deepfake videos were a significant and unprecedented threat to our epistemic practices. From election interference and financial fraud to nonconsensual pornography and digital impersonation, deepfakes have been weaponized to exploit the public's trust in audiovisual media. For example, a deepfake video featuring Tom Hanks falsely endorsing a dental scam circulated online, exploiting the actor's trusted image to deceive consumers (Guardian, 2023). This risk is amplified by the decentralized and viral nature of social media platforms, which allow deceptive content to spread globally in a matter of seconds.

Although the headlines tend to highlight negative deepfake applications directed at celebrities, private citizens are also at risk for social media abuse. For example, a fourteen-year-old student in New Jersey was victimized when innocuous images she had posted online were manipulated into nude images and then circulated around her high school (Hadero, 2023). In another case an Indian woman's identity was misappropriated by her ex-boyfriend, who allegedly used it for revenge to create "Babydoll Archi" an AI influencer that generated over 1 million rupees in earnings featuring seductive deepfake videos and images—all without her knowledge (Jaiswal, 2025). Deepfake social media posts can also be used to manipulate financial markets through fabricated yet convincing videos of CEOs or CFOs. Even a brief appearance of a fake executive issuing a false statement—such as a resignation, earnings warning, or acquisition announcement—can trigger market volatility (Bateman, 2022). For cybercriminals, this short window of disruption is enough to execute lucrative trades, such as selling inflated shares or shorting stock before the deception is uncovered. The rapid impact of such tactics undermines market integrity and investor trust, while allowing perpetrators to disappear with illicit profits (de Rancourt-Raymond, 2023).

Negative applications can also occur outside of social media. In courtroom proceedings, deepfakes pose a threat not only by introducing fabricated evidence, but also by casting doubt on the authenticity of legitimate materials—forcing parties to prove that real evidence is not fake (Bellas, 2025). In one case outside of the US, a wife in a custody dispute tried to use an audio recording to portray her husband as threatening, but she was thwarted by an expert witness that

concluded the metadata in the recording showed signs of deepfake manipulation (Swerling, 2020; Court, 2020). There are numerous examples of creative, yet malicious uses of deepfake media. Political operatives used AI-generated audio to mimic US President Joe Biden's voice in robocalls, discouraging voters from participating in the New Hampshire primary (Swenson & Weissert, 2024). In another example, artificial intelligence–based "deep voice" technology was used to mimic a CEO's voice, enabling the theft of $35 million from a Japanese company operating in Hong Kong (Brewster, 2023).

2.3 Take It Down Act

In this section we will address the first US federal law enacted to directly address negative deepfake applications. Although deepfake technology began drawing widespread attention in early 2018, Congress did not pass any meaningful legislation regulating deepfakes until the Take It Down Act was signed into law on May 19, 2025. The Act was backed by First Lady Melania Trump and her fight against "revenge porn," receiving overwhelming support in both houses of Congress (Ortutay, 2024).

The Take It Down Act criminalizes the nonconsensual distribution of authentic and AI-generated sexually explicit images. Even though it uses the term "deepfake" in its long-form name, it does not actually use the name throughout its text. Instead, it uses the term "digital forgeries"— which it defines as AI-generated or technologically manipulated depictions of identifiable individuals, including adults and minors, that portray nudity or sexually explicit conduct (Take It Down Act, § 2). To be actionable, prosecutors must first prove these forgeries are realistic enough that a reasonable observer cannot distinguish them from authentic images. Prosecutors must then prove the defendant's improper intent. When the content involves adults, federal prosecutors must prove that the defendant intended to cause or did cause psychological, financial, or reputational harm. In cases involving minors, the government must show that the image was published either to humiliate or harass the minor or to satisfy the sexual desires of any person (Take It Down Act, § 2).

The Take It Down Act also goes beyond prohibiting the publication of harmful deepfake content. It also targets its use in coercive behavior such as blackmail and digital abuse. It criminalizes intentional threats to publish digital forgeries when the threat is done to intimidate, harass, or as a form of extortion. Penalties under the act vary depending on the nature of the offense and the age of the victim. For adults, the unauthorized publication or threat to publish either deepfakes or authentic intimate imagery may result in fines and imprisonment for up to 2 years (or 18 months in the case of threatened publication). If the offense involves a minor, the penalties are even more severe with potential imprisonment of up to 3 years for publication and up to 30 months for a threat to publish the material (Take It Down Act, § 2).

Section 3 of the Act also imposes new obligations on digital platforms, which it defines broadly as public websites, online services, and mobile applications that primarily facilitate user generated content or specialize in publishing nonconsensual intimate imagery. It establishes a mandatory "notice-and-removal" framework whereby digital platforms must implement a mechanism that allows individuals to report nonconsensual sexually explicit "visual depictions" of themselves and request their removal. Once the platform receives the request, it must act promptly to remove the content and undertake "reasonable efforts" to identify and eliminate any identical duplicates within 48 hours. Noncompliance constitutes a violation of the Federal Trade Commission Act (Take It Down Act, § 3) and the Federal Trade Commission (FTC) can step in to impose fines or order changes (Grimmelmann, 2025).

Some commentators have pointed to potential shortcomings in the Take It Down Act. While it mandates platform cooperation, it does not establish clear federal mechanisms for detecting or investigating synthetic media offenses (Ortutay, 2025). Nor does it allocate funding for law enforcement training or technological infrastructure (Kimery, 2025). These gaps may hinder the law's effectiveness in practice.

3. Literature Review

A large body of literature has analyzed the range of malicious applications associated with deepfake technology and raised concerns about its impact on public discourse, institutional trust, and personal security. The seminal study on the consequences of deepfake technology and the broader misinformation crisis was authored by Chesney and Citron (2019). They articulated five categories of harm associated with deepfake misuse: exploitation of individuals (such as blackmail or reputational sabotage), societal harm through misinformation, erosion of trust in institutions, threats to national security, and the broader undermining of public confidence in truth itself. They also explored the concept of the "liar's

dividend," whereby the mere possibility of media manipulation en ables bad faith actors to discredit legitimate evidence by falsely asserting it is fabricated. Recent studies describe more specific risks to businesses such as fraud, damage to image, reputation, and more specific risks to individuals such as bullying, defamation, harassment, identity theft, intimidation, and revenge porn.

Deepfake researcher Aviv Ovadya captures the severity of the technology's potential impact through the term "Infopocalypse" which he used to describe a future fundamentally altered by synthetic media where there is "a catastrophic failure of the marketplace of ideas" (2018). He emphasizes that the threat goes beyond mere individual misuse, instead he underscores their broader implications for truth, trust, and democracy and characterizes deepfakes as "the distortion of reality itself." Empirical studies reinforce these concerns. For example, Vaccari and Chadwick (2020) showed that watching a deepfake video left people uncertain about what is real and what is not, which in turn reduced overall trust in social media and the news. Lollia (2023) highlights how deepfakes threaten democratic stability by shaping public opinion and manipulating sentiment, particularly during critical moments such as elections.

Other research documents how rapidly deepfakes spread through digital platforms, circumventing conventional journalistic oversight, and enabling harmful and deceptive actors to distribute synthetic content with minimal examination (Hameleers et al., 2022). Fenn et al. (2019) note that the presence of an image significantly increases the likelihood that a fake news story will be shared, highlighting why deepfakes—being highly visual and often more realistic—are particularly effective at spreading more widely than other forms of misinformation. Rayhan and Rayhan (2023) similarly note the ease with which AI generated media can be distributed online, posing serious risks to democratic processes. Their research emphasizes the difficulty of identifying and correcting such falsehoods once they are released into the public domain. They call for greater attention to detection tools and legal responses to ensure information integrity in political contexts.

A psychological dimension of deepfake susceptibility is explored by Thaler (2024), who examines the role of motivated reasoning in public perception. Thaler found that individuals are more likely to accept and trust content, deepfake or otherwise, that aligns with their pre-existing political beliefs. This cognitive bias not only increases the perceived credibility of deepfakes, but also reduces the public's ability to discern factual content from manipulated media, thus compounding the dangers of synthetic misinformation. Although not specifically directed at deepfakes, Van Bavel et al.'s (2024) study of the spread of misinformation is also instructive. This study shows that misinformation is more likely to be accepted when it is consistent with the recipient's social identity. In other words, "[p]artisanship can override accuracy concerns, resulting in politically biased beliefs and the spread of misinformation" (Van Bavel et al., 2024).

There is also an arm of research studying individual ability to recognize deepfakes and investigating who is particularly vulnerable to being fooled by deepfakes. Bray et al. (2023) asked participants to differentiate between genuine and deepfake facial images and found that overall detection accuracy was just about chance, even though the participant's confidence in their ability to correctly identify deepfakes was high. Other studies confirm that people are overly confident regarding their detection ability (Köbis et al., 2021; Lago et al., 2021). As to whom is more likely to be fooled by deepfakes, Sütterlin et al. (2023) found little correlation between gender and deepfake detection but did find that individuals were less likely to recognize a deepfake if it was viewed on a mobile phone or tablet as compared to a stationary computer. Caramancion (2021) found that older individuals are especially vulnerable to deepfakes, and Lovato et al. (2024) found that younger adults were better able to distinguish between authentic video content and deepfake video content. Ahmed and Chua (2023) found that individuals with high cognitive ability were less likely to perceive deepfakes as accurate and less likely to share them.

As deepfake technology becomes more sophisticated, distinguishing deepfakes from authentic videos is growing increasingly challenging (Romero-Moreno, 2025). Korshunov and Marcel (2021) found that while human perception is very different from machine perception, both are successfully fooled by deepfakes but in different ways and that "algorithms struggle to detect the deepfake videos that humans find to be very easy to spot." Recent studies highlighted significant variation in human abilities to detect artificially generated media across countries, underscoring the global

and cross-cultural implications of deepfake threats (Herbert et al., 2024). Studies by Ha et al. (2024) as well as Nightingale and Farid (2022) found that artificial media were often perceived as so authentic that some viewers found AI-generated content more trustworthy than human-generated material.

A significant body of research in computer science and engineering has been devoted to the detection of deepfakes through digital forensics (Tampubolon, 2024). This research focuses on building algorithmic tools capable of identifying synthetic content based on inconsistencies in facial movements, lighting, eye blinking patterns, or audio-visual synchronization (Ahmed et al., 2024). Romero-Moreno (2025) describes the "arms-race" between deepfake developers and forensic analysts, noting that as detection technologies improve, so too do the methods to evade them.

Most pertinent to this study is the vein of research discussing the regulation of deepfakes. Research in this area often compares and comments on deepfake regulatory approaches across multiple nations (Geng, 2023; Zheng et al., 2025), discusses regulation inadequacies (Darma et al., 2025), and provides suggestions for future legislation (Vyas, 2024; Gupta, 2023). Recently, some have begun to document the extent and scope of legislative proposals (Birrer & Just, 2024). Burniston's (2025) law review article reviewed enacted legislation regulating deepfakes in political advertising. A recent article by Ugwuoke & Sanfillippo (2025) analyzed 319 deepfake-related state bills that were proposed between 2019 and 2024 but noted that only 74 were ultimately enacted into law. This paper carries forward this stream of inquiry through a comprehensive analysis of what state legislators in the US have done to address negative deepfake applications through enacted legislation. Rather than focus on proposed regulation, this paper is centered on enacted legislation that provides specific criminal punishments for negative deepfake applications.

The following section will discuss the methodology behind this study, noting how the authors pursued this comprehensive analysis of recent legislation criminalizing the malicious use of deepfake technology.

4. Methodology

This study reviews all state-level legislation in the US between January 1, 2018, and May 19, 2025, searching for any legislation that assesses a criminal penalty for deepfake applications. The dates of this study were chosen because, although the earliest forms of deepfake technology existed prior to 2018, it is widely accepted that the term deepfake was popularized and entered the public discourse at the very end of 2017 (Office of the Director of National Intelligence, 2021). The research aims to identify how states have utilized criminal law to confront harmful deepfake applications and to document the scope and content of legislative efforts prior to the enactment of the federal Take It Down Act on May 19, 2025.

Despite the growing scholarly and legal attention, there is no universally accepted definition of the term "deepfake" (Birrer & Just, 2024), which is sometimes alternatively spelled "deep fake" (Fernandez, 2021). Merriam Webster (2025) dictionary defines a deepfake as "an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said." Many scholars would contend this definition is too broad, as it misses the fundamental aspect of a deepfake, the use of artificial intelligence to generate the media (Fernandez, 2021). However, the use of artificial intelligence to create the deepfake is not universally accepted among scholars (Birrer & Just, 2024). Though some try to distinguish "cheapfakes" or "shallowfakes," which are manipulations not utilizing artificial intelligence, from deepfakes, there is no clear boundary between what is and what is not considered a deepfake (Altuncu, 2024). As noted in the analysis below, even a cursory glance at enacted deepfake legislation reveals that many legislators are not concerned with the specific technological tools used to create a deepfake, they just want to curtail deepfake abuse. Therefore, to capture all legislative enactments that came after the term was coined, this paper focused on legislative enactments that met the broader definition outlined in the Merriam Webster quotation noted above.

To locate relevant legislation, the study utilized key terms to search the Westlaw Campus Research Enacted Legislation database, which was then cross-referenced with the results of additional searches in the following legal and policy research databases to form an initial list of enacted bills: the National Conference of State Legislatures (NCSL), the Legislative Use of Artificial Intelligence Surveys (2023 and 2024), and the MultiState AI Policy Tracker (2024).

These searches were conducted using various combinations of the following keywords: "artificial intelligence," "deepfake," "deep fake," "synthetic media," "deceptive media," "computer generated media," and "machine learning."

Each of the pieces of legislation revealed through these queries were then reviewed to identify those that were both related to the specific use of deepfake technology and criminal in nature. The analysis focused only on legislation that imposed criminal penalties for deepfake-related conduct, thereby excluding civil remedies and regulatory initiatives that lacked criminal penalties. The resulting 89 criminal deepfake bills were then coded using thematic analysis, a qualitative research technique that identifies patterns and themes within data. Following the methodology outlined by Braun and Clarke (2006), the research took an inductive approach, meaning that the analysis was not guided by a predefined framework but allowed themes to emerge organically from the content of the legislation itself. The goal was to detect patterns that reflect how state legislatures are responding to the risks and realities of synthetic media.

5. Analysis

This section outlines the key findings of the legislative review and thematic analysis conducted for this study, including the scope, content, chronology, and political sponsorship of deepfake-related criminal laws. In total, the research identified 89 enacted bills addressing inappropriate application of deepfake technology through criminal law prior to the passage of the Take It Down Act. The following sections will describe: 1) the chronology of these enactments, 2) partisan trends in introducing the legislation, 3) categories or groupings of the types of criminalized applications, and 4) inconsistent legislative definitions of deepfake activity. Citations to each enactment referenced throughout the balance of this paper are listed by state, in alphabetical order, in Appendix A.

5.1 Chronology

As noted above, the term "deepfake" was officially coined at the end of 2017 and the public quickly began to realize its potential negative applications through high profile examples of image-based sexual abuse. However, it took until 2019 for any state to pass legislation to criminalize its use. Virginia was the first state to do so, passing a bill that made it a crime to publish or threaten to publish, with bad intent, a sexual image of another. This statute did not specifically use the term deepfake, but it was clearly aimed at sexual media that was created or manipulated without consent. Virginia's 2019 House Bill 2678 empowered Virginia authorities to punish any person who "maliciously disseminates or sells" with the "intent to coerce, harass, or intimidate" a sexual video or image of another person "created by any means whatsoever." Importantly, the "another person" referenced in the bill "includes a person whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person and who is recognizable as an actual person by the person's face, likeness, or other distinguishing characteristic."

Maryland quickly followed suit that same year by addressing the issue in the realm of child sexual abuse material, once again not specifically using the term deepfake, but instead criminalizing the possession of "a film, videotape, photograph, or other visual representation showing an actual child or a computer-generated image" of a child engaging in sexual conduct "that is indistinguishable from an actual and identifiable child" (House Bill 1027). The statute then went on to specify that "'[i]ndistinguishable from an actual child' includes an actual minor or a computer-generated image that has been created, adapted, or modified to appear as an identifiable child."

Texas was the only other state to criminalize deepfakes in 2019, but it went after an entirely different application, prohibiting the creation and distribution of deepfake videos intended to harm candidates for public office or to influence elections. Texas specifically used the term "deep fake," becoming the first state to do so in a criminal statute (Senate Bill 751). The law provided that "[a] person commits an offense if the person, with intent to injure a candidate or influence the result of an election: (1) creates a deep fake video; and (2) causes the deep fake video to be published or distributed within 30 days of an election." This bill also defined the term "deep fake" broadly as "a video, created with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality." One commentator notes that this definition is so broad it could even include merely slowing down a person's voice or actions with an intent to deceive the audience into believing the depicted individual was intoxicated (Tandy, 2024).

These early efforts signaled emerging legislative awareness of deepfake capabilities but also revealed three distinct,

yet divergent priorities. First, Virginia updated its criminal law concerning the malicious publication of nonconsensual sexual media, commonly referred to as "revenge porn," to include computer generated images. For clarity, this paper will refer to this legislative priority as nonconsensual sexual deepfake legislation. Second, Maryland updated its child sexual abuse material laws to include deepfake applications. This was necessary because most child sexual abuse material laws were written to protect real children, not necessarily synthetic images of children. Texas ushered in the third major legislative priority, protecting election integrity from deepfakes. Over the next few years, all legislation criminalizing deepfakes fit into one of these three categories, with new priorities surfacing in 2024 and 2025.

In 2020, Georgia was the only state that took an action to criminalize the inappropriate use of deepfake technology. Georgia focused on nonconsensual sexual deepfake legislation by amending the invasion of privacy statute to include "a falsely created videographic or still image" (Senate Bill 337). When updating the existing law, the legislature also made sure to add language exempting the "interactive computer service" from liability if the content was put there by a third party.

In 2021, three states—Hawaii, Utah, and Wyoming—criminalized nonconsensual sexual deepfakes. In doing so, Utah used the novel term "counterfeit intimate image" which it defined as "any visual depiction, photograph, film, video, recording, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means that has been edited, manipulated, or altered to depict the likeness of an identifiable individual" (House Bill 193).

The pace of new legislation remained modest over the next two years. In 2022, Florida, Indiana, and South Dakota passed legislation criminalizing nonconsensual sexual deepfakes. In 2023, New York and Minnesota passed nonconsensual sexual deepfake legislation, with Minnesota also criminalizing election deepfakes. Three states also updated their existing criminal deepfake laws in 2023. Virginia and Maryland made minor changes to their existing nonconsensual sexual deepfake laws and Texas became the first state to enact legislation addressing all three categories when Senate Bill 1361 criminalized nonconsensual sexual "deep fake video" and House Bill 2700 added "computer-generated visual material that was created, adapted or modified using an artificial intelligence application or other computer software" to its laws regulating child sexual abuse material.

The number of states regulating deepfakes exploded in 2024, when 29 states passed a total of 48 pieces of legislation regulating deepfakes. For example, on April 29, 2024, in two separate bills, Florida outlawed election deepfakes and deepfake material depicting child sexual abuse. Delaware did the same, passing its only two deepfake-related criminal bills on the same day in 2024, addressing child sexual abuse material and election deepfakes. In total, during the 2024 legislative session 21 states addressed deepfake material depicting child sexual abuse, 9 addressed nonconsensual sexual deepfakes, and 8 passed legislation criminalizing election deepfakes. Two states also passed legislation addressing different deepfake applications outside the major categories of child sexual abuse, nonconsensual sexual deepfakes, and election deepfakes. Tennessee criminalized deepfakes of an individual's voice or likeness when done for commercial purposes (House Bill 2091; Peretti & Rice, 2024) and New Hampshire passed legislation criminalizing deepfakes that harassed, defamed, entrapped, extorted, or financially harmed an identifiable person (House Bill 1432).

In 2025, prior to the May 19th passage of the Take It Down Act, 15 states passed 23 pieces of legislation criminalizing deepfakes. Of those bills, 10 addressed child sexual abuse material, 7 addressed nonconsensual sexual deepfakes, and 4 addressed election deepfakes. Three states added new categories, with New Jersey enhancing the punishment for a long list of criminal offenses if deepfakes technology was used to carry out the crime, Arizona Senate Bill 1295 added "using a computer-generated voice recording, image or video of another person with the intent to defraud other persons" to the list of actions constituting the offense of "criminal impersonation," and Virginia criminalized the use of any synthetic digital content for the purpose of committing any criminal offense involving fraud (House Bill 2124).

Figure 1 illustrates the criminal laws enacted between 2019 and the passage of the Take It Down Act in May of 2025.

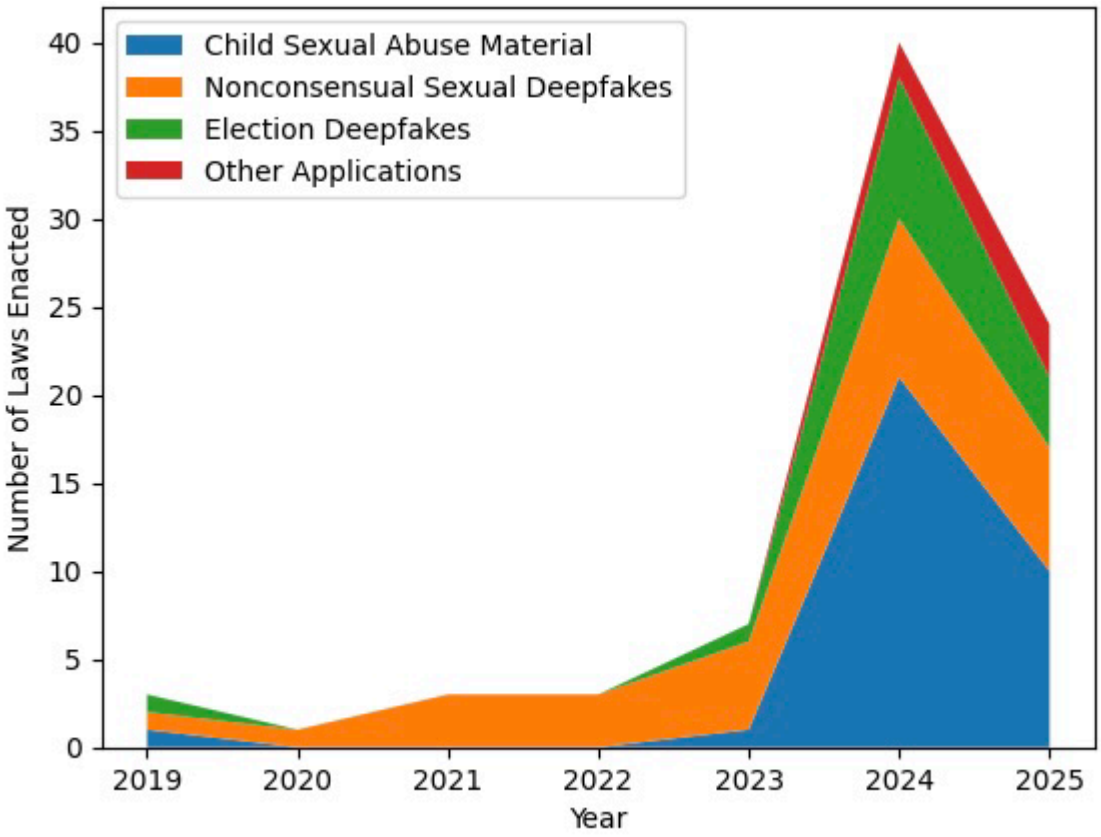


Figure 1. State Deepfake Criminal Legislation by Category (Jan. 1, 2019–May 19, 2025).

5.2 Bipartisan Support and Sponsorship

This section considers whether political affiliation played a role in the introduction of criminal deepfake bills. Analysis of who sponsored or introduced each of these enacted bills reveals that criminalizing negative deepfake applications has largely been a bipartisan endeavor. Of the 89 pieces of legislation, 55 were sponsored by bipartisan coalitions or committees. Only 21 of the bills were sponsored solely by Republicans and only 15 were sponsored solely by Democrats. No consistent partisan pattern emerged in terms of the types of harm targeted.

For the 42 bills enacted to criminalize child sexual abuse deepfakes, 3 were sponsored by Democrats, 9 were sponsored by Republicans, and 30 were introduced with co-sponsors from each party.

For the 28 bills enacted to criminalize nonconsensual sexual deepfakes, 7 were sponsored by Democrats, 8 were sponsored by Republicans, and 13 were introduced with co-sponsors from each party.

For the 16 bills enacted to criminalize election deepfakes, 5 were sponsored by Democrats, 2 were sponsored by Republicans, and the remaining 9 were introduced with co-sponsors from each party.

For the 5 bills enacted to criminalize deepfakes outside of these major categories, 1 was sponsored by a Democrat, 2 were sponsored by Republicans, and 2 were introduced with co-sponsors from each party.

In total, the sponsorship data suggests a broad, cross-party recognition of deepfakes as a serious issue requiring attention. There were widespread concerns about deepfake media infiltrating elections, and both parties worked together to address the issue through criminal legislation.

Category of Criminalization	Total Bills	Democratic-only	Republican-only	Bipartisan
Child sexual abuse material	42	3	9	30
Nonconsensual sexual deepfakes	28	7	8	13
Election deepfakes	16	5	2	9
Other deepfake applications	5	1	2	2
Total	**89***	**16**	**21**	**55**

Table 1. Sponsorship Patterns by Category of Deepfake Harm.
*A handful of bills contained criminal penalties for multiple categories.

5.3. Categorization

In total, 41 of the 50 states passed legislation criminalizing negative deepfake applications between January 1, 2018, and May 19, 2025. These uses have been grouped into four categories, which are discussed below. These four categories reflect the primary ways deepfakes can harm victims, with each category defined by the nature and scope of the impact. A state-by-state summary of legislative activity in each category is provided in Appendix B.

5.3.1 Child Sexual Abuse Material

Child sexual abuse material generated via artificial intelligence is growing at an alarming rate. One source notes that reports of AI-generated child sexual abuse imagery quadrupled between 2023 and 2024 (Internet Watch Foundation, 2025). Thirty-one states passed legislation to criminalize this type of deepfake activity. Most of these legislative enactments simply added language unequivocally listing computer or AI-generated images to the definition of child pornography or child sexual abuse material and extended the criminal penalties to include fines and jail time for people who create, possess, or distribute this type of media.

Wisconsin's 2023 enactment is instructive. Prior to the passage of the new legislation, Wisconsin criminalized the possession of visual media depicting a child engaged in actual or simulated sexually explicit conduct (Staubach 2025). To secure a conviction, prosecutors were required to show that the defendant knew or reasonably should have known that the individual depicted in the identified media was under 18. Senate Bill 314 expanded the definition of child sexual abuse material to include "digital" or "computer-generated" images that appeared to depict a real child engaged in sexually explicit conduct, even if no actual child was used in producing the image. This change allowed prosecutors to pursue charges for the possession or distribution of synthetic or AI-generated child sexual abuse material, so long as the content of the image met the legal definition of obscenity.

Likewise, in 2024 Pennsylvania sought to close legal loopholes that could make it difficult to prosecute cases involving AI-generated images that did not involve actual children. Rather than use the term deepfake, Senate Bill 1213 used the term "artificially generated child sexual abuse material" which it defined to mean media "that appears to authentically depict a child under 18 years of age engaging in a prohibited sexual act or in the simulation of such act that did not occur in reality" and "the production of which was substantially dependent upon technical means, including artificial intelligence or photo editing software, rather than the ability of another person to physically impersonate the child."

5.3.2 Nonconsensual Sexual Deepfakes

Prior to the popularization of the term deepfake, most every state had "revenge porn" laws criminalizing the distribution of nonconsensual graphic intimate images of an individual (Suslavich, 2023). However, prosecution of nonconsensual sexual deepfakes was hampered because deepfake media may not show the actual intimate portions of the victim, instead depicting mere fabrications or intimate images from another source. This study found that 22 states passed legislation

specifically addressing nonconsensual sexual deepfakes. Generally, these laws prohibit the malicious creation, posting, or distribution of manipulated sexual images depicting an identifiable individual without their consent.

For example, Arkansas created the criminal offense of unlawfully creating or distributing "deepfake visual material" when someone knowingly creates or distributes deepfake visual material that "[i]s generated, modified, or adapted using technology to falsely depict a person's appearance, voice, or conduct; and appears to a reasonable person to be an authentic depiction of a person" in a state of nudity or engaging in sexual acts, without that person's consent (House Bill 1529). Minnesota made it a crime to intentionally and without consent distribute a deepfake image that realistically depicts an identifiable person's intimate parts or an identifiable person engaging in a sexual act (House File 1370).

Alabama addressed this loophole by amending its existing law to state that it includes "[a] recording that a reasonable person would believe actually depicts an identifiable individual, regardless of whether any portion of the recording depicts another individual or is artificially generated" (House Bill 161). The first violation is a misdemeanor, and any subsequent conviction is a felony. This bill also provided a specific liability exemption noting "[n]o developer or provider of technology shall be held to have violated this section solely for providing or developing technology used by another person to violate this section."

In some cases, prosecutors must demonstrate that the defendant acted with the intent to harass, harm, or intimidate the victim. For example, Idaho makes it a crime to disclose "explicit synthetic media" if there is "intent to annoy, terrify, threaten, intimidate, harass, offend, humiliate, or degrade an identifiable person portrayed in whole or in part in the explicit synthetic media" or if it is done "with the intent to obtain money or other valuable consideration" (House Bill 575).

5.3.3 Election Deepfake Laws

Sixteen states enacted legislation to criminalize the use of manipulated media in public elections. For example, Delaware established a new election crime, the use of deepfake technology to influence the outcome of an election. House Bill 316 made it a crime to distribute within 90 days of an election any audio or visual content manipulated or created by "generative adversarial network (GAN) techniques" or "other digital technology" if done with the intent to harm a political party or candidate or to deceive voters. Punishment for the crime is enhanced if it was done with the intent to cause violence or bodily harm. The bill also specifically exempts any "interactive computer service" from liability.

Similarly, Hawaii made it a crime to recklessly disregard the risk of harming a candidate's reputation, electoral prospects, or changing voter behavior by distributing "materially deceptive media" between the first working day of February and the general election in an even-numbered year. Senate Bill 2687 specifies that the materially deceptive media must have been created by "[g]enerative adversarial network techniques or another technique that translates a source image into another image using machine learning, deep learning techniques, and convolutional neural networks" or "artificial intelligence" or "digital technology."

Many bills provided an exemption from criminal liability if the manipulated election media contains a disclosure or disclaimer. The technical requirements, such as the font size of the disclaimer, the speed of an audio disclaimer, and length of time the disclaimer must be visible, varied from state-to-state. For example, New Mexico requires that the "text of the disclaimer shall appear in a size that is easily readable" (House Bill 182), and Alabama requires that the disclaimer be written in "letters in a size that is easily readable by the average viewer" (House Bill 172). Some states require general warnings that the media is not authentic, while others require specific verbiage in the disclaimer.

Verbiage requirements vary greatly. Washington's legislation requires that the disclosure must state that the media has been manipulated and adhere to specific visual and audio requirements (House Bill 1999), while Hawaii states that manipulated media must include a "disclaimer informing the viewer that the media has been manipulated by technical means and depicts appearance, speech, or conduct that did not occur" (Senate Bill 2687). Other states are even more specific. Delaware requires a label that says, "This (image/video/audio) has been altered or artificially generated" (House Bill 316). New Hampshire's election deepfake criminal law does not apply if there is a disclaimer stating, "This _____ has been manipulated or generated by artificial intelligence technology and depicts speech or conduct that did not occur" (House Bill 1596).

Similarly, many states give requirements for how disclaimers must appear on manipulated audio media. Delaware's requirement for manipulated audio media states that "the disclosure must be read in a clearly spoken manner and in a

pitch that can be easily understood by the average listener, at the beginning of the audio, at the end of the audio, and, if the audio is greater than 2 minutes in length, interspersed within the audio at intervals of not more than 2 minutes each" (House Bill 316). New Hampshire's requirements for manipulated audio state that the disclosure must be "read in a clearly spoken manner and in a pitch that can be easily heard by the average listener, at the beginning of the audio, at the end of the audio, and, if the audio is greater than 2 minutes in length, interspersed within the audio at intervals of not greater than 2 minutes each" (House Bill 1596). New Mexico is less specific but still states that disclaimers on manipulated audio need to be "clearly spoken during the advertisement" (House Bill 182).

5.3.4 New Legislative Priorities

Child sexual abuse material deepfakes, nonconsensual sexual deepfakes, and election integrity deepfakes garnered the most legislative attention. However, five states have recently enacted legislation beyond these narrow applications.

In March of 2024, Tennessee criminalized the unauthorized commercial use of someone's voice or likeness—a response to AI impersonation in advertising and endorsement fraud. The "Ensuring Likeness Voice and Image Security (ELVIS) Act" (House Bill 2091) provides both criminal penalties and civil liability for unauthorized use of Tennessee artists' names, voices, likenesses and images through artificial intelligence.

In July of 2024, New Hampshire's governor signed House Bill 1432 into law, which stipulates that an individual is guilty of a felony "if the person knowingly creates, distributes, or presents any likeness in video, audio, or any other media of an identifiable individual that constitutes a deepfake for the purpose of embarrassing, harassing, entrapping, defaming, extorting, or otherwise causing any financial or reputational harm to the identifiable person." The bill defines a "Deepfake" as "a video, audio, or any other media of a person in which his or her face, body, or voice has been digitally altered so that he or she appears to be someone else, he or she appears to be saying something that he or she has never said, or he or she appears to be doing something that he or she has never done." Specific applications are exempted from criminal liability such as when the deepfake was satire, parody, or news reporting. The law also specifically exempts third party interactive computer services that provide access to the deepfake from criminal liability.

In March of 2025, Virginia made it a Class 1 misdemeanor for any person to use any synthetic digital content for the purpose of committing any criminal offense involving fraud (House Bill 2124). The law makes such action a separate and distinct offense with punishment separate and apart from any punishment received for the commission of the primary criminal offense.

In April of 2025, New Jersey made it illegal to create and/or distribute a deepfake when the deepfake is used to commit a long list of potential crimes beyond promoting the sexual exploitation of minors or improperly influencing political processes (Assembly Bill 1432). The list of crimes or offenses includes sex-related crimes, theft and property crimes, forgery, fraud, deceiving law enforcement, obstruction of justice, endangering the welfare of a child, creating false public alarms, harassment, hazing, and intimidation because of race, color, religion, gender, or disability. This law essentially elevates the penalty for the underlying offense, so that by creating or distributing such deceptive media for criminal purposes results in further penalties that will be imposed in addition to any sentence for the related underlying crime. Importantly, the law punishes both those who create or generate the deepfake for criminal purposes and those who solicit, disclose, or use the deepfake for criminal purposes. The law includes a long list of exemptions for the legitimate use of deepfake technology, which includes satire, parody, news reporting, teaching, and research.

In May of 2025, Arizona passed Senate Bill 1295 amending the existing crime of "criminal impersonation" by specifically identifying the that the use of a "computer generated voice recording, image or video of another person with the intent to defraud other persons" constituted this crime. The use of deepfake technology also enhances the penalties associated with the crime, as compared to impersonation committed without the use of deepfake technology.

These laws represent an important expansion of legislative concern—acknowledging that deepfakes can be used in a

wide array of harmful contexts beyond sexual exploitation and election deception. The next section will consider the varying types of terms and definitions used throughout these bills to refer to the concept of a deepfake.

5.4 Legislative Definitions

The legislative enactments identified in this paper do not consistently use the same terms to describe what type of media is prohibited. For example, Utah used the term "synthetic media" which it defines as audio or visual media "substantially produced by generative artificial intelligence" (Senate Bill 131). Washington prohibits "fabricated depiction[s] which includes, but is not limited to, creation or alteration of any visual or printed matter by use of artificial intelligence" (House Bill 1999). Utah used the term "counterfeit intimate image" (House Bill 148).

Ten states specifically used the term deep fake or deepfake. However, just as there is disagreement within the academic community as to the proper definition of the term deepfake, there is disagreement amongst the legislatures defining the term.

Four of the ten states—Mississippi, Montana, South Dakota, Tennessee—specified the use of artificial intelligence within the definition of the term deepfake or deep fake. For example, South Dakota noted that a deepfake is "any image, audio recording, or video recording created or manipulated with the use of artificial intelligence or other digital technology that is so realistic, a reasonable person would believe it depicts the speech or conduct of an actual individual who did not in fact engage in the speech or conduct" (Senate Bill 164).

Delaware did not use the term artificial intelligence but instead defined a deep fake as a depiction "created or intentionally manipulated with the use of generative adversarial network techniques or other digital technology in a manner to create a realistic but false image, audio, or video" (House Bill 316).

The remaining five states utilizing the term deepfake or deep fakes—Minnesota, New Hampshire, Arkansas, Louisiana, Texas—made no mention of artificial intelligence, generative adversarial networks, or any other specific technological tool when defining the term, instead keeping the nature of the manipulation vague. For example, New Hampshire defines a deepfake as "a video, audio, or any other media of a person in which his or her face, body, or voice has been digitally altered so that he or she appears to be someone else, he or she appears to be saying something that he or she has never said, or he or she appears to be doing something that he or she has never done" (House Bill 1432). Louisiana's child sexual abuse material law defines a deepfake as "any audio or visual media in an electronic format, including any motion picture film or video recording, that is created, altered, or digitally manipulated in a manner that would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual or replace an individual's likeness with another individual and depicted in the recording" (Senate Bill 175). Finally, Minnesota merely states that the political media needs to be created in a way that is "substantially dependent upon technical means, rather than the ability of another individual to physically or verbally impersonate [another] individual" (House File 1370).

For the most part, in each state the criminal rule applies to the three primary types of media: image, audio, and video. However, some are more limited. For example, Idaho uses the term synthetic media, which it defines as "any image or video created or altered using technical means, such as artificial intelligence . . ." but noticeably does not include audio media (House Bill 575). Another example is Texas, which only specifically criminalizes "deep fake video[s]" for election deepfakes (Senate Bill 751) and nonconsensual sexual deepfakes (Senate Bill 1361).

6. Discussion

This study demonstrates that deepfakes constitute a complex and evolving technological and social phenomenon that will continue to pose significant challenges for the foreseeable future. The rapid expansion of state and federal legislation reflects growing awareness of the harms associated with synthetic media, particularly in the contexts of child sexual abuse material, nonconsensual sexual exploitation, and election interference. At the same time, the findings reveal that the legal response to deepfakes remains fragmented, incomplete, and unevenly enforced. Legislation alone is insufficient to meaningfully reduce the negative impacts of deepfake technology. Instead, an effective response requires coordinated action among multiple stakeholder groups, including lawmakers, law enforcement, researchers, information technology developers, and cybersecurity professionals.

6.1 Contributions and Implications for State Legislators

Despite the large number of legislation enactments, substantial gaps remain. For example, deepfakes used for

interpersonal harassment or nonsexual reputational sabotage are not clearly addressed by existing criminal laws. Some laws require proof of intent to harm, which can be difficult to establish. State-level legislation demonstrates that bipartisan support exists for criminalizing deepfake misuse, particularly in the context of sexual exploitation and election deception. Yet this patchwork approach leads to uneven protections, with some states criminalizing a broad range of conduct and others addressing only the narrowest categories of abuse. The variation in statutory definitions, scope, and enforcement mechanisms leaves many victims without adequate recourse and allows offenders to hide behind jurisdictional lines. Legislators are encouraged to enact or update laws that will fill these gaps. These findings suggest that state legislators should focus on filling substantive gaps, harmonizing definitions, and broadening coverage to reflect the full range of harms posed by synthetic media.

6.2 Contributions and Implications for Federal Legislators

The patchwork of state laws results in inconsistencies that complicate enforcement and limit protection for victims whose harm transcends state borders. The Take It Down Act provides a baseline federal framework, but further action is needed to harmonize definitions, expand protections, and improve investigative infrastructure. While the passage of the Take It Down Act marks a crucial federal intervention, its narrow scope limits its efficacy. The Act only addresses a subset of nonconsensual sexual deepfake harms and does not provide a comprehensive solution to the many forms of synthetic media abuse. It also does not fully equip law enforcement with the tools needed to investigate and prosecute offenders. The law's takedown provisions represent an important first step, but further regulatory and technical support is necessary to ensure meaningful change and enforcement.

Looking forward, a more comprehensive and coordinated strategy is needed. First, federal legislation should be expanded to address a wider range of harms, including commercial impersonation, synthetic defamation, and fraudulent uses that currently fall outside most legal frameworks. Deloitte's Center for Financial Services predicts that generative AI could enable fraud losses in the United States to expand from $12.3 billion in 2023 to $40 billion in 2027 (Lalchand, 2024). Second, clear and consistent definitions of "deepfake" should be developed to guide enforcement and close jurisdictional gaps.

Federal policymakers should consider complementary civil remedies, additional platform accountability measures, transparency in AI model development, and enhanced user protections. A multifaceted and future-oriented legal framework is necessary for an effective response to the evolving challenges of synthetic media.

6.3 Implications for Law Enforcement Capacity and Deterrence

The study highlights substantial enforcement challenges. Detection and enforcement capabilities for deepfake-related crimes remain underdeveloped, as many local agencies lack the resources, infrastructure, and specialized skills necessary for effective cybercrime investigation (Moloney et al., 2022; Lin, 2025). Support for technical training, digital forensic infrastructure, and public education should be prioritized to improve detection and reduce vulnerability. Enforcement is crucial. Increasing the risk of apprehension and conviction is more influential in reducing crime than raising the expected severity of punishment (Bun et al., 2020). In a large meta study on deterrence, the probability of legal sanctions was found to have a greater deterrent effect than the severity of legal sanctions (Rupp, 2008) and as noted by one scholar, "there is far more empirical support for the deterrent effect of changes in the certainty of punishment than changes in the severity of punishment" (Nagin, 2013). Therefore, emphasis should be placed on making perpetrators aware that they could be caught. Strengthening investigative capacity will help improve detection and reduce overall vulnerability to deepfake threats.

6.4 Contributions and Future Directions for Legal and Criminal Justice Research

This study documents the large number and variety of state statutes that now criminalize negative deepfake applications. Criminal justice researchers could follow through with studies to determine which legal sanctions have been most effective in reducing deepfake misuse. Future research could also focus on subsequent legislative developments since passage of the Take It Down Act. Researchers could study how online platforms have responded to state and federal legislation, including voluntary content moderation policies and compliance challenges. More specific investigation into the implementation of the notice and takedown provisions set forth in the Take It Down Act would be instructive. Legal researchers could also analyze how deepfake laws are being challenged or upheld in US courts. In addition, comparative analyses of legal responses across different countries may yield valuable insights into best practices

and inform the development of more effective and cohesive regulatory frameworks.

6.5 Implications for Information Technology Developers

While legal research provides some insights into the problems associated with deepfakes, information technology can offer additional solutions through continued advances in methodologies for deepfake detection. For example, the breadth of potential detection technologies is described in a study by Rana, et al. (2022) where they conducted a systematic literature review summarizing 112 relevant articles from 2018 to 2020 and identified four categories for deepfake detection methodologies: deep learning-based techniques, classical machine learning-based methods, statistical techniques, and blockchain-based techniques. They found that there was no one best solution at the time of their review, so information technology researchers are encouraged to continue work to create methods to quickly and effectively detect deepfakes (Rana et al., 2022). The hope is that this can complement work by other stakeholders to reduce the overall negative impact on the victims. This is particularly important because it is expected that systems will continue to be produced that make it easier to create deepfakes leading to a more widespread problem.

6.6 The Role of Cybersecurity Professionals and Security Awareness Training

While this study and prior sections discuss improving general education and technical detection capabilities, social engineering–based deepfake attacks—such as voice cloning used to impersonate executives—are likely to increase as well. These attacks exploit human trust rather than technical vulnerabilities and pose significant risks to organizations and individuals. Cybersecurity professionals will need to play a critical role to mitigate the harm of such deepfakes. For example, integrating realistic deepfake examples into security awareness training can help employees recognize and respond to synthetic media threats. Such training would not only strengthen organizational defenses against fraud and impersonation but could also produce broader societal benefits by increasing overall public recognition of deepfake risks. Because cybersecurity awareness programs reach millions of employees across numerous sectors, this has the potential to help reduce vulnerability to overall deepfake misuse and foster digital literacy.

6.7 Need for Stakeholder Collaboration

Finally, as deepfake technology continues to advance, its potential to erode trust, manipulate perception, and inflict harm will only grow. Lawmakers, technologists, and civil society must proactively collaborate to mitigate these risks and ensure that digital innovation is accompanied by legal safeguards that protect individuals, institutions, and democratic integrity. The legislative landscape remains uneven, with substantial variation in scope, definitions, and enforcement mechanisms. It is clear that a multifaceted approach is needed to address this growing problem.

6.8 Limitations

Several limitations affect the generalizability of this study. First, the results demonstrate a snapshot of a moment in time—the state of enacted criminal deepfake legislation as of the date of passage of the Take It Down Act on May 19, 2025. Some states have likely updated their laws since its passage. Second, this study focused on legislation enacted between January of 2018 and May of 2025. If forward-thinking legislators had passed broad legislation criminalizing the use of computer-generated images before they became as realistic as we see today, those legislative enactments would not have been captured in this study. This paper is also limited to criminal legislation and does not include new legislation that creates private civil remedies for those who were injured by deepfakes. Finally, although all efforts were used to capture every possible legislative enactment criminalizing the use of deepfake technology, it is possible that legislators could have used other terminologies within their respective legislation that went beyond the common terms noted above.

7. Conclusion

This study provides a comprehensive analysis of the evolution of US criminal law responses to deepfake technology, documenting the rapid expansion, diversification, and bipartisan nature of state and federal legislative efforts. By mapping and analyzing enacted statutes, the manuscript demonstrates that US efforts to criminalize deepfakes have largely followed a cautious, harm-specific approach rather than a comprehensive regulatory strategy, selectively targeting high-risk applications such as child sexual abuse material, nonconsensual sexual deepfakes, and election manipulation. While state lawmakers have made progress in these areas, the resulting legal framework remains fragmented and incomplete. The findings underscore the limits of criminal law as a response to rapidly evolving generative technologies and demonstrate that legislation alone is insufficient to address the multifaceted harms posed by synthetic media. Effective mitigation will require coordinated legal reform, enhanced enforcement capacity, continued technological

innovation, expanded cybersecurity awareness, and sustained interdisciplinary collaboration. Together, these efforts provide a foundation for more coherent, evidence-based policy development that seeks to preserve constitutional protections and technological innovation while minimizing the potential for harm.

8. References

Abernathy, P. C. (2024). AI tools as supplementary support in language acquisition. DOI: 10.20944/preprints202407.2213.v1

ActiveFence. (2023). *AI Surge in Non-Consensual Intimate Imagery (NCII):* New Insights. https://www.activefence.com/research/ai-surge-ncii/

Ahmed, N. U. R., Badshah, A., Adeel, H., Tajammul, A., Duad, A., & Alsahfi, T. (2024). Visual deepfake detection: Review of techniques, tools, limitations, and future prospects. *IEEE Access*.

Ahmed, S., & Chua, H. W. (2023). Perception and deception: Exploring individual responses to deepfakes across different modalities. *Heliyon*, *9*(10).

Ajder, H., & Schick, N. (2021, March 30). *Deepfake apps are here and we can't let them run amok*. WIRED. https://www.wired.com/story/deepfakes-security/

Altuncu, E., Franqueira, V. N., & Li, S. (2024). Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*, 1400024. doi.org/10.3389/fdata.2024.1400024

Bateman, J. (2020). *Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios Cyber Policy Initiative Working Paper Series | "Cybersecurity and the Financial System" #7*. https://carnegie-production-assets.s3.amazonaws.com/static/files/Bateman_FinCyber_Deepfakes_final.pdf

Bellas, G. (2025). *Deepfakes in the Courtroom: Problems and Solutions | Illinois State Bar Association*. Isba.org. https://www.isba.org/sections/ai/newsletter/2025/03/deepfakesinthecourtroomproblemsandsolutions

Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*, *0*(0). https://doi.org/10.1177/14614448241253138

Blitz, M. J. (2018). Lies, line drawing, and deep fake news. *Okla. L. Rev.*, *71*, 59.

Brady, M. (2020, September 1). A new disinformation threat?: Understanding the potential harms and what is being done to address the threat (Publication No. 1). https://democracyreporting.s3.eu-central-1.amazonaws.com/images/20842020-09-01-DRI-deepfake-publication-no-1.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect 'deepfake'images of human faces. *Journal of Cybersecurity*, *9*(1).

Brewster, T. (2021, October 14). *Fraudsters Cloned Company Director's Voice In $35 Million Bank Heist, Police Find*. Forbes. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/

Bun, M. J., Kelaher, R., Sarafidis, V., & Weatherburn, D. (2020). Crime, deterrence and punishment revisited. *Empirical economics*, *59*(5), 2303-2333. https://doi.org/10.1007/s00181-019-01758-6

Burchard, H. von der. (2018, May 21). *Belgian socialist party circulates "deep fake" Donald Trump video*. POLITICO. https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/

Burniston, M. M. (2025). State and Regulatory Agency Approaches to Limiting Deepfakes in Political Advertising. *Vanderbilt Journal of Entertainment & Technology Law*, *27*(4), 797.

Busacca, A., & Monaca, M. A. (2023). Deepfake: Creation, purpose, risks. In *Innovations and economic and social changes due to artificial intelligence: the state of the art* (pp. 55-68). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33461-0_6

Caramancion, K. M. (2021, April). The demographic profile most at risk of being disinformed. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (pp. 1-7). IEEE. Doi: 10.1109/IEMTRONICS52119.2021.9422597.

Cardi, W. J., Penfield, R. D., & Yoon, A. H. (2012). Does Tort Law Deter Individuals? A Behavioral Science Study. *Journal of Empirical Legal Studies*, *9*(3), 567-603.

Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, *4*(6), 4-8.

Chesney, B. & Citron, D. (2018, February 21). *Deepfakes: A Looming Crisis for National Security, Democracy and Privacy?* Lawfare. https://www.lawfareblog.com/deepfakes-looming-crisis-national-security-democracy-and-privacy

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753..

Cole, S. (2018, March 1). *AI-assisted fake porn is here and we're all fucked*. VICE. https://www.vice.com/en/article/ai-assisted-fake-porn-is-here-and-were-all-fucked/

Cotter, P. (2019, February 25). *How Rogue One: A Star Wars Story Brought Back Princess Leia*. ScreenRant. https://screenrant.com/rogue-one-princess-leia-actress-cgi-carrie-fisher/

Court. (2020). *AIAAIC - Dubai deepfake court evidence*. Aiaaic.org. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/dubai-deepfake-court-evidence

Cullins, A., & Cullins, A. (2017, May 1). *Carrie Fisher, "Star Wars" and the Legal Issues of Dead But In-Demand Actors*. The Hollywood Reporter. https://www.hollywoodreporter.com/movies/movie-news/carrie-fisher-star-wars-legal-issues-dead-but-demand-actors-997335/

Darma, M., Gisymar, N. A., Purwadi, A., Zubaedah, P. A., & Judijanto, L. (2025). Legal Implications of Deepfake Technology Misuse in Digital Content on Social Media. *Science of Law*, *2025*(3), 98-103.

Davis, R. (2020, Spring). *Technology Factsheet: Deepfakes.* Belfer Center for Science and International Affairs. https://www.belfercenter.org/publication/technology-factsheet-deepfakes

Department of Homeland Security. (2023). Increasing Threat of Deepfake Identities. In *Department of Homeland Security*. Department of Homeland Security. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

de Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime*, *30*(4), 1066-1077.

Fenn, E., Ramsay, N., Kantner, J., Pezdek, K., & Abed, E. (2019). Nonprobative Photos Increase Truth, Like, and Share Judgments in a Simulated Social Media Environment. Journal of Applied Research in Memory and Cognition, 8, 131–138.

Fernandez, A. (2022). "Deep fakes": disentangling terms in the proposed EU Artificial Intelligence Act. *UFITA Archiv für Medienrecht und Medienwissenschaft*, *85*(2), 392-433.

Fiske, J. (2025). *A Star Wars Trilogy: Examining the Legal Implications of Disney's Three Posthumous Uses of AI and CGI*. Brooklaw.edu. https://sports-entertainment.brooklaw.edu/media/a-star-wars-trilogy-jpf/

Flynn, A., Clough, J., & Cooke, T. (2022). Disrupting and preventing deepfake abuse: Exploring criminal law responses to AI-facilitated abuse. In *The Palgrave handbook of gendered violence and technology* (pp. 583-603). Cham: Springer International Publishing.

Frank, J., Herbert, F., Ricker, J., Schönherr, L., Eisenhofer, T., Fischer, A., Dürmuth, M. & Holz, T. (2024, May). A representative study on human detection of artificially generated media across countries. In *2024 IEEE Symposium on Security and Privacy (SP)* (pp. 55-73). IEEE.

Geng, Y. (2023). Comparing "Deepfake" Regulatory Regimes in the United States, the European Union, and China. *Geo. L. Tech. Rev.*, *7*, 157.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf

Gorin, D. (2021). *Is AI-Generated Child Pornography a Crime?* Egattorneys.com. https://www.egattorneys.com/ai-generated-child-pornography

Gottfried, J. (2019, June 14). About Three-Quarters of Americans Favor Steps to Restrict Altered Videos and Images. Pew Research Center Report.

Griffin, M. (2019, September 4). *The DeepFake tech behind David Beckham's viral malaria campaign explained*. 311 Institute. https://www.311institute.com/the-tech-behind-david-beckhams-viral-deepfake-malaria-campaign-explained/

Grimmelmann, J. (2025). Deconstructing the Take It Down Act. *Communications of the ACM*, *68*(8), 28-30. https://doi.org/10.1145/3747203

Guardian Staff, (2023, October 2). Tom Hanks says AI version of him used in dental plan ad without his consent. *The Guardian*. https://www.theguardian.com/film/2023/oct/02/tom-hanks-dental-ad-ai-version-fake

Gupta, K. (2023). The future of deepfakes: Need for regulation. *Nat'l LU Delhi Stud. LJ*, *5*, 130.

Ha, A. Y. J., Passananti, J., Bhaskar, R., Shan, S., Southen, R., Zheng, H., & Zhao, B. Y. (2024, December). Organic or diffused: Can we distinguish human art from ai-generated images?. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (pp. 4822-4836).

Habgood-Coote, J. (2023). Deepfakes and the epistemic apocalypse. *Synthese*, *201*(3), 103. https://doi.org/10.1007/s11229-023-04097-3

Hadero, H. (2023, December 2). *Teen girls are being victimized by deepfake nudes. One family is pushing for more protections*. AP News. https://apnews.com/article/deepfake-ai-nudes-teen-girls-legislation-b6f44be048b31fe0b430aeee1956ad38

Hall, H. K. (2018). Deepfake videos: When seeing isn't believing. *Cath. UJL & Tech*, *27*, 51.

Hameleers, M., van der Meer, T. G., & Dobber, T. (2022). You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media+ Society*, *8*(3), 20563051221116346. https://doi.org/10.1177/20563051221116346

Harris, D. (2018). Deepfakes: False pornography is here and the law cannot protect you. *Duke L. & Tech. Rev.*, *17*, 99.

Hawkins, W., Russell, C., & Mittelstadt, B. (2025). Deepfakes on Demand: the rise of accessible non-consensual deepfake image generators. *arXiv preprint arXiv:2505.03859*.

International Association of Chiefs of Police. (2025). Law enforcement in the era of deepfakes. Police Chief Magazine. https://www.policechiefmagazine.org/law-enforcement-era-deepfakes

Internet Matters Team. (2024, October 22) Report estimates half a million UK teenagers have encountered AI-generated nude deepfakes [Press release]. Internet Matters. https://www.internetmatters.org/hub/press-release/new-report-uk-teenagers-encountered-ai-generated-nude-deepfakes/

Internet Watch Foundation. (2025). *New AI child sexual abuse laws announced following IWF campaign*.https://www.iwf.org.uk/news-media/news/new-ai-child-sexual-abuse-laws-announced-following-iwf-campaign/

Jaiswal, N. (2025, July 14). *Babydoll Archi's ex-boyfriend turns revenge into AI racket, earns money by faking influencer's identity online*. Indiatimes. https://www.indiatimes.com/trending/babydoll-archis-ex-boyfriend-turns-revenge-into-ai-racket-earns-money-by-faking-influencers-identity-online/articleshow/122786353.html

Jodka, S. (2024, February). *Manipulating reality: the intersection of deepfakes and the law*. Reuters. https://www.reuters.com/legal/legalindustry/manipulating-reality-intersection-deepfakes-law-2024-02-01/

Kaur, A., Noori Hoshyar, A., Wang, X., & Xia, F. (2024, October). Beyond Deception: Exploiting Deepfake Technology for Ethical Innovation in Healthcare. In *Proceedings of the 1st International Workshop on Multimedia Computing for Health and Medicine* (pp. 70-78).

Kimery, A. (2025, May 16). *FTC chief warns deepfake takedown law requires urgent infrastructure overhaul*. Biometric Update | Biometrics News, Companies and Explainers; BiometricUpdate.com.

https://www.biometricupdate.com/202505/ftc-chief-warns-deepfake-takedown-law-requires-urgent-infrastructure-overhaul

Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308-323. Doi: 10.1080/13600834.2020.1794615.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *Iscience*, *24*(11).

Korshunov, P., & Marcel, S. (2021, June). Subjective and objective evaluation of deepfake videos. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2510-2514). IEEE.

Kshetri, N., DeFranco, J. F., & Voas, J. (2023). Is it live, or is it deepfake?. *Computer*, *56*(7), 14-16. https://doi.org/10.1109/MC.2023.3252059

Lago, F., Pasquini, C., Böhme, R., Dumont, H., Goffaux, V., & Boato, G. (2021). More real than real: A study on human visual perception of synthetic faces [applications corner]. *IEEE Signal Processing Magazine*, *39*(1), 109-116.

Lalchand, S., Srinivas, V., Maggiore, B., & Henderson, J. (2024, May 28). *Generative AI is expected to magnify the risk of deepfakes and other fraud in banking*. Deloitte Insights; Deloitte. https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html

Lollia, F. (2023). At the heart of social networks: influence, disinformation and societal risk. https://hal.science/hal-04134028v1/file/Heart%20of%20social%20network.pdf

Lovato, J., St-Onge, J., Harp, R., Salazar Lopez, G., Rogers, S. P., Haq, I. U., Hébert-Dufresne, L. & Onaolapo, J. (2024). Diverse misinformation: impacts of human biases on detection of deepfakes on networks. *npj Complexity*, *1*(1), 5. https://doi.org/10.1038/s44260-024-00006-y

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974-4026.

Merriam-Webster, deepfake. In Merriam-Webster.com dictionary. Retrieved October 17, 2025, from https://www.merriam-webster.com/dictionary/deepfake

Miotti, A., & Wasil, A. (2024). Combatting deepfakes: Policies to address national security threats and rights violations. *arXiv preprint arXiv:2402.09581*.

Moloney, C., Unnithan, P., & Zhang, W. (2022, April 6). *Assessing Law Enforcement's Cybercrime Capacity and Capability*. FBI: Law Enforcement Bulletin. https://leb.fbi.gov/articles/featured-articles/assessing-law-enforcements-cybercrime-capacity-and-capability-

Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dutse, I., Khan, S., & Parkinson, S. (2023). A survey on the detection and impacts of deepfakes in visual, audio, and textual formats. *Ieee Access*, *11*, 144497-144529. Doi: 10.1109/ACCESS.2023.3344653

Mullen, M. (2022). A new reality: Deepfake technology and the world around us. *Mitchell Hamline L. Rev.*, *48*, 210.

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, *154*, 113368.

Nagin, D. S. (2013). Deterrence: A review of the evidence by a criminologist for economists. *Annu. Rev. Econ.*, *5*(1), 83-105.

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q-V. & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, 103525.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8), e2120481119.

O'Sullivan, D. (2023, May 22). *See fake image of an "explosion" near the Pentagon that caused confusion | CNN Business*. Www.cnn.com. https://www.cnn.com/videos/business/2023/05/22/pentagon-fake-explosion-image-

twitter-osullivan-lead-vpx.cnn

Ortutay, B. (2025, April 29). *Congress approves Melania Trump's Take It Down Act. What is it?* AP News. https://apnews.com/article/take-it-down-deepfake-trump-melania-first-amendment-741a6e525e81e5e3d8843aac20de8615

Ovadya, A. (2018). What's Worse than Fake News? The Distortion of Reality Itself. *New Persp. Q.*, *35*, 43.

Peretti, K. (2024, June 10). *Tennessee Law Designed to Combat Deepfakes Set to Take Effect in July | Alston & Bird Privacy, Cyber & Data Strategy Blog*. Alston & Bird Privacy, Cyber & Data Strategy Blog. https://www.alstonprivacy.com/tennessee-law-designed-to-combat-deepfakes-set-to-take-effect-in-july/

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, *10*, 25494-25513.

Rayhan, S., & Rayhan, S. (2023). The role of AI in democratic systems: Implications for privacy, security, and political manipulation. Doi: 10.13140/RG.2.2.31121.61281.

Romero-Moreno, F. (2025). Deepfake detection in generative AI: A legal framework proposal to protect human rights. *Computer Law & Security Review*, *58*, 106162.

Rupp, T. (2008). Meta-Analysis of Crime and Deterrence: A Comprehensive Review of the Literature. Ph.D. Dissertation, Technische Universitat, Darmstadt. Doi:10.26083/tuprints-00001054

Saunders, B., Camgoz, N. C., & Bowden, R. (2020). Everybody sign now: Translating spoken language to photo realistic sign language video. *arXiv preprint arXiv:2011.09846*.

Simard, L. S. (2014). A view from within the Fortune 500: An empirical study of negative value class actions and deterrence. *Ind. L. Rev.*, *47*, 739.L.S. Lin, Organisational Challenges in US Law Enforcement's Response to AI-Driven Cybercrime and Deepfake Fraud. Laws 14(4) (2025) 46-64.

Spivak, R. (2018). "Deepfakes": The Newest Way to Commit One of the Oldest Crimes. *Geo. L. Tech. Rev.*, *3*, 339.

Statt, N. (2020). Jay z tries to use copyright strikes to remove deepfaked audio of himself from youtube. *The Verge*, *28*.

Statt, N. (2020, April 28). *Jay Z tries to use copyright strikes to remove deepfaked audio of himself from YouTube*. The Verge. https://www.theverge.com/2020/4/28/21240488/jay-z-deepfakes-roc-nationyoutube-removed-ai-copyright-impersonation/

Staubach, A. (2025, September 9) *Category: Laws And Legislation*. Wisconsin Justice Initiative Inc. https://www.wjiinc.org/blog/category/laws-and-legislation

Suslavich, B. T. (2023). Nonconsensual Deepfakes: A "Deep Problem" for Victims. *Alb. LJ Sci. & Tech.*, *33*, 160.

Sütterlin, S., Ask, T. F., Mägerle, S., Glöckler, S., Wolf, L., Schray, J., Chandi, A. Bursac, T. Khodabakhsh, A., Knox, B., Canham, M. & Lugo, R. G. (2023, July). Individual deep fake recognition skills are affected by viewer's political orientation, agreement with content and device used. In *International Conference on Human-Computer Interaction* (pp. 269-284). Cham: Springer Nature Switzerland. **https://doi.org/10.1007/978-3-031-35017-7_18**.

Swenson, A., & Weissert, W. (2024, January 22). *New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary*. AP News. https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocall-f3469ceb6dd613079092287994663db5

Swerling, G. (2020, January 31). Doctored audio evidence used to damn father in custody battle. *The Telegraph*. https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/

Tampubolon, M. (2024). Digital face forgery and the role of digital forensics. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, *37*(3), 753-767. http://repository.uki.ac.id/14200/1/DigitalFaceForgery.pdf

Tandy, B. (2024). Deepfakes: Identity Misappropriation in the Digital Age. *Belmont L. Rev.*, *12*, 271.

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in

news. *American Economic Journal: Microeconomics*, *16*(2), 1-38.

Take It Down Act, (2025). Pub L No 119-12, 139 Stat 321.

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *Plos one*, *18*(10), e0291668.

Ugwuoke, V., & Sanfilippo, M. R. (2025). The Current Landscape of Deepfake Legislation in the United States: Analysis of State-Level Responses. *Journal of Information Policy*, *15*.

U.S. Congress. (2019). National Defense Authorization Act for Fiscal Year 2020 (Public Law No. 116-92).

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, *6*(1), 2056305120903408.

Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*, *56*, 101787. https://doi.org/10.1016/j.copsyc.2023.101787

Venema, A. E., & Geradts, Z. J. (2020). Digital forensics, deepfakes, and the legal process. *The SciTech Lawyer*, *16*(4), 14-23.

Vyas, A. D. (2024). United States of Deepfake. *Tenn. L. Rev.*, *92*, 307.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).

Williams, K. (2024, September 12). *US States Struggle to Define "Deepfakes" and Related Terms as Technically Complex Legislation Proliferates*. Tech Policy Press. https://www.techpolicy.press/us-states-struggle-to-define-deepfakes-and-related-terms-as-technically-complex-legislation-proliferates

Willis, A. (2024, February 23). *Bills targeting AI-generated pornography advance in Alabama House*. Alabama Daily News. https://aldailynews.com/bills-targeting-ai-generated-pornography-advance-in-alabama-house/

Wong, M. (2024, September 26*). High school is becoming a cesspool of sexually explicit deepfakes*. The Atlantic https://www.theatlantic.com/technology/archive/2024/09/ai-generated-csam-crisis/680034/

Zheng, G., Shu, J., & Li, K. (2025). Regulating deepfakes between Lex Lata and Lex ferenda—a comparative analysis of regulatory approaches in the US, the EU and China. *Crime, Law and Social Change*, *83*(1), 1.

Appendix A: Referenced State Statutes

Alabama HB 161, 2024 Reg Sess (2024).
Alabama HB 172, 2024 Reg Sess (2024).
Arizona SB 1295, 57th Reg Sess, ch 184 (2025).
Arkansas HB 1529, 95th General Assembly (2025).
Delaware HB 316, 152nd Gen Assembly (2024).
Delaware HB 353, 152nd Gen Assembly (2024).
Florida HB 0919, 2024 Reg Sess, ch 126 (2024).
Florida SB 1680, 2024 Reg Sess, ch 118 (2024).
Florida SB 1798, 2022 Reg Sess, ch 212 (2022).
Georgia SB 337, 2019-2020 Reg Sess (2020).
Hawaii SB 2687, 2024 Reg Sess (2024).
Hawaii SB 309, 2021 Reg Sess (2021).
Idaho HB 575, 2024 Reg Sess, ch 105 (2024).
Indiana HB 1363, 2022 Reg Sess (2022).
Louisianna SB 175, 2023 Reg Sess (2023).
Maryland HB 1027, 2019 Reg Sess, ch 325 (2019).
Maryland SB 226, 2023 Reg Sess, ch 759 (2023).
Minnesota HF 1370, 93rd Reg Sess, ch 58 (2023).
New Hampshire HB 1432, 2024 Reg Sess, ch 243 (2024).
New Hampshire HB 1596, 2024 Gen Sess, ch 345 (2024).
New Jersey Bill A 3540, 2025 Reg Sess, ch 40 (2025).
New Mexico HB 182, 2024 Reg Sess, ch 57 (2024).
New York S 01042, 2023-2024 Gen Assembly, ch 513 (2023).
Pennsylvania SB 1213, 2023-2024 Reg Sess (2024).
South Dakota SB 120, 2022 Reg Sess (2022).
South Dakota SB 164, 2025 Reg Sess (2025).
Tennessee HB 2091, 11th Gen Assembly (2024).
Tennessee HB 2091, 11th Gen Assembly (2024).
Texas HB 2700, 88th Reg Sess (2023).
Texas SB 1361, 88th Reg Sess (2023).
Texas SB 751, 86th Reg Sess (2019).
Utah HB 193, 2021 Gen Sess (2021).
Utah HB 148, 2024 General Sess (2024).
Utah SB 131, 2024 Reg Sess (2024).
Virginia HB 2124, 2025 Reg Sess, ch 0398 (2025).
Virginia HB 2398, 2023 Reg Sess, ch 612 (2023).
Virginia HB 2678, 2019 Reg Sess, ch 490 (2019).
Washington HB 1999, 2023-2024 Reg Sess (2024).
Wisconsin SB 314, 2023-2024 Reg Sess (2024).
Wyoming HB 0085, 2021 Reg Sess, ch 85 (2021).

Appendix B: Deepfake Criminal Legislative Activity by Category (2019–May 19, 2025)

State	Nonconsensual Sexual Deepfakes	Election	Child Sex Abuse Material	Other
Alabama	✓	✓	✓	
Alaska				
Arkansas	✓		✓	
Arizona			✓	✓
California	✓		✓	
Colorado				
Connecticut			✓	
Delaware		✓	✓	
Florida		✓	✓	
Georgia	✓		✓	
Hawaii	✓	✓		
Idaho	✓		✓	
Illinois			✓	
Indiana			✓	
Iowa	✓		✓	
Kansas			✓	
Kentucky			✓	
Louisiana	✓		✓	
Maine				
Maryland	✓		✓	
Massachusetts	✓			
Michigan		✓		
Minnesota	✓	✓		
Mississippi		✓	✓	
Missouri				
Montana	✓	✓	✓	
Nebraska				
Nevada				
New Hampshire	✓	✓	✓	✓
New Jersey		✓	✓	✓
New Mexico		✓		
New York	✓			
North Carolina			✓	
North Dakota	✓	✓		
Ohio				
Oklahoma	✓		✓	
Oregon				
Pennsylvania			✓	
Rhode Island				
South Carolina	✓			
South Dakota		✓	✓	
Tennessee	✓		✓	✓
Texas	✓	✓	✓	
Utah		✓	✓	
Vermont	✓			
Virginia	✓		✓	✓
Washington			✓	
West Virginia			✓	
Wisconsin			✓	
Wyoming	✓			

Author Biographies

J. Royce Fichtner is the Ellis and Nelle Levitt Distinguished Professor of Business Law at the Zimpleman College of Business at Drake University. He received his Bachelor's Degree in Business Administration at the University of Northern Iowa and his Juris Doctorate from Drake University. His major research interests are information technology law, company law, and corporate governance.

Troy J. Strader is the Aliber Distinguished Professor of Information Systems in the Zimpleman College of Business at Drake University. He received his Ph.D. in Business Administration (Information Systems) from the University of Illinois at Urbana Champaign in 1997. His research interests include the impact of artificial intelligence on business and society, information technology (IT) ethics, online consumer behavior, and IT-related business strategy. Dr. Strader has edited three books and his research has been published in *Decision Support Systems, Electronic Markets*, the *Journal of Information Technology*, the *Journal of the Association for Information Systems,* the *International Journal of Electronic Commerce, Communications of the ACM, Electronic Commerce Research,* the *European Journal of Information Systems*, and other academic and practice publications. He is the founding Editor-in-Chief for the Drake Management Review.

Journal of the Midwest Association for Information Systems

Volume 2026 | Issue 1 Article 4

Date: 01-31-2026

From Users to Innovators: Making the Case for Lead UX in IS

Javad Norouzi Nia
Central Michigan University, norou1j@cmich.edu

Vishal Shah
Central Michigan University, shah3v@cmich.edu

Gustav Verhulsdonck
Central Michigan University, verhu1g@cmich.edu

Abstract

This viewpoint article argues that Information Systems (IS) professionals must evolve their design practices to effectively harness insights from user-generated content (UGC). While participatory and user-centered design remain foundational, they struggle to scale and anticipate emerging needs in today's dynamic digital ecosystems. We introduce the Lead UX framework, which integrates von Hippel's lead user theory with data-analytic approaches to systematically identify both innovation-driving lead users and key emergent lead topics within large-scale user discussions. By combining automated topic modeling and related natural language processing techniques with expert human review, Lead UX enables IS professionals to transform overwhelming volumes of UGC into actionable intelligence for proactive, innovation-driven design. We outline a hybrid workflow for operationalizing Lead UX, address challenges such as data quality and analytic bias, and propose future research directions, including comparative case studies, practitioner assessments, and algorithmic benchmarking. Finally, we offer actionable recommendations for educators and practitioners on embedding Lead UX into curricula and development processes, positioning users as co-designers in continuous IS innovation.

Keywords: Lead UX, user-generated content, information systems design, innovation analytics.

DOI:10.17705/3jmwa.000099

1. Introduction

Information Systems (IS) design must adapt to harness the innovative potential of user-generated content (UGC). This paper argues that IS professionals can enhance design processes by adopting Lead UX analytics, a data-analytic practice grounded in von Hippel's (2005) lead user innovation theory. Lead UX systematically identifies (1) lead users, individuals whose advanced needs and solutions anticipate broader markets, and (2) emergent lead topics, recurring themes in user discourse that point to high-impact design opportunities. This approach complements established methods as users have shifted from passive recipients to active content creators.

The proliferation of UGC across digital platforms has fundamentally reshaped user–system interactions. Millions of reviews, forum posts, and social media comments now reveal system performance, unmet needs, and creative workarounds (Kitchin, 2014). Traditional IS methods, such as user-centered design (UCD; Abras et al., 2004) and participatory design (PD; Muller, 2007), remain valuable for deep, contextual understanding but struggle with the volume, speed, and diversity of contemporary UGC. These approaches are resource-intensive (Schuler & Namioka, 1993), sensitive to power dynamics that may limit open participation (Bratteteig & Wagner, 2014) and often focus on current or average users rather than emerging behaviors and disruptive use cases (Koskinen et al., 2011). In contrast, integrating UGC into UX work can provide richer context than structured interviews (Schäfer, 2011), support tracking of evolving needs (Krumm, 2018), and inspire the development of new features (von Hippel, 2005).

Lead UX extends von Hippel's (2005) lead user theory, which posits that some users experience needs and devise solutions that later diffuse to the broader market. These users frequently encounter challenges that later become mainstream and possess both the incentive and capability to innovate (von Hippel, 2005). Classic examples include Chris Messina's (2007) proposal of the hashtag (#) on Twitter, which evolved into a core platform convention (Huddleston, 2020), and 1970s mountain biking enthusiasts in Marin County, California, who modified bicycles with oversized tires and reinforced frames to create "clunkers," anticipating a new market category years before manufacturers such as Specialized and Trek responded (von Hippel, 2006). Another example would be 3M products, developed through systematic lead-user engagement, which achieved annual sales of approximately $146 million after five years, about eight times those of products developed through traditional approaches (Lilien et al., 2002).

Lead UX adapts this perspective to large-scale digital environments. Rather than relying on small, curated samples, it employs computational tools to identify both individual- and collective-level signals. Lead topics, persistent recurring themes in user discussions, reveal emergent needs that may not be visible in traditional studies of average users. Analyses of social media interactions, for instance, show how early adopters introduce practices (such as hashtags) ahead of official platform support, shaping collective behavior and core functionalities (Zappavigna, 2015; Dvir et al., 2019). Lead UX seeks to detect these signals systematically.

The advent of big data analytics has expanded possibilities for understanding user needs at scale (Chen et al., 2015; Goes, 2014). Techniques such as topic modeling and sentiment analysis can reveal patterns in large text corpora that would be infeasible to analyze manually. These computational approaches augment, rather than replace, established user research methods by distilling actionable insights from large volumes of user comments and informing more targeted design decisions.

Accordingly, this paper addresses two central questions for IS design in the UGC era: (1) How can IS professionals systematically identify and amplify the perspectives of lead users whose insights may signal emergent needs yet are often overlooked in traditional large-scale analyses? (2) What analytic strategies facilitate the scalable extraction of actionable findings from large volumes of unsolicited user content? While Lead UX cannot capture all user perspectives, it aims to surface insights that drive innovation and shape future design directions.

We present Lead UX as both a conceptual framework and a practical methodology that synthesizes lead user theory with contemporary analytical tools. Unlike participatory approaches that broadly engage stakeholders, Lead UX targets participants whose authentic, unsolicited feedback indicates high innovation potential. Combining automated techniques with human interpretation enables IS professionals to turn significant data challenges into strategic opportunities for innovation.

The remainder of this viewpoint paper is structured as follows. Section 2 reviews the rise of UGC in IS and highlights the limitations of PD and UCD in scaling and anticipating emergent needs. Section 3 introduces the Lead UX framework,

situates it among established design methodologies, defines lead user characteristics, explains their strategic importance, and outlines hybrid methods for identifying them. Section 4 discusses the operationalization of Lead UX in practice, including implementation challenges and directions for future research. Section 5 offers actionable recommendations for the IS community to adopt and refine this approach.

2. Related Work

2.1 The Rise of User-Generated Content in Information Systems

The arrival of user-generated content (UGC) has fundamentally reshaped interactions between users and digital systems. With billions of reviews, forum contributions, and social media messages submitted daily, users have transitioned from passive consumers to active participants in the evolution of products and services (Chen, Mao, & Liu, 2014; Ren et al., 2007). Within the IS community, research increasingly emphasizes the importance of analyzing large-scale user contributions to inform system design and enhance engagement (Mitchell, 2018; Lebens, 2021).

UGC provides continuous, large-scale, and authentic feedback on system performance and unmet needs, far surpassing the snapshot nature of traditional usability studies (Preece & Shneiderman, 2009). It enables users to appropriate and redefine technology, a dynamic characterized as "user innovation" (von Hippel, 2005), which blurs conventional boundaries between designers and users. Emerging analytical techniques now allow researchers to capture and interpret this feedback at unprecedented scales, strengthening the link between user participation and system improvement (Klein, 2017).

Capturing value from this "social big data" requires integrating automated analytical strategies such as sentiment mining and natural language processing (Schäfer, 2011). Advances in computational methods, including collaborative filtering and crowdsourced recommendations, enable IS professionals to systematically extract patterns and actionable insights from massive UGC datasets (Solachidis et al., 2009). Building on these developments, recent work has proposed frameworks for mining, analyzing, and applying large volumes of user feedback to guide iterative improvements in information systems (Mitchell, 2018; Lebens, 2021). In this context, UGC functions both as a critical data resource and as evidence of a participatory cultural shift in IS development (O'Reilly, 2007; Jenkins, 2006; Markus & Silver, 2008).

2.2 Limitations of Traditional Design Approaches

User-centered design (UCD) and participatory design (PD) have historically empowered users in system development (Simonsen & Robertson, 2013; Bratteteig & Wagner, 2014). However, in the context of massive, continuously flowing UGC, these approaches show important shortcomings.

UCD often privileges mainstream preferences and averages, overlooking heavy users or those with unconventional needs who may drive disruptive innovation (Nielsen, 2024; Koskinen et al., 2011; Norman, 2013; Schreier & Prügl, 2008; von Hippel, 2005). Lüthje, Herstatt, and von Hippel (2006) illustrate the scale of this issue in their study of mountain biking user-innovators: 38% of users reported having developed ideas for product modifications, yet traditional UCD is structurally unable to systematically identify and prioritize these distributed innovations, particularly when they originate from a minority of highly advanced users.

Common UCD techniques, such as interviews, surveys, and usability tests, are effective in controlled, small-scale settings (Kujala, 2003; Maguire, 2001; Kuutti, 2009) but lack scalability to address the volume and velocity of modern digital contributions. Conducting in-depth studies with large user bases is economically infeasible, forcing organizations to rely on small, potentially unrepresentative samples that may systematically exclude lead users whose insights signal emerging trends (Wu & Huberman, 2007). These methods excel at generating deep contextual understanding but require complementary techniques to analyze large-scale, diverse user discussions.

Similarly, PD's collaborative workshops typically involve small groups and can reinforce the views of the average user while marginalizing outlier voices (Muller & Kuhn, 1993; Baldwin & von Hippel, 2011; Bødker, Kensing, & Simonsen, 2004). Wu and Huberman (2007) demonstrate that user contributions in online environments follow power-law distributions, with a small fraction of users generating a disproportionate share of content and innovation. Broad-sample PD may dilute or overshadow these minority perspectives, treating them as noise rather than as signals of innovation.

Both UCD and PD also tend to capture user needs at discrete points in time, making it difficult to adapt to rapidly changing requirements in dynamic digital ecosystems. In addition, they demand substantial investments of time and resources in recruitment, facilitation, and analysis, constraints that become prohibitive when organizations must monitor evolving user

feedback across multiple platforms.

Taken together, these limitations reveal a gap between the capabilities of established user research methods and the demands of the contemporary UGC landscape. Neither UCD nor PD fully leverages the scale, diversity, or velocity of user-generated data, nor do they systematically surface the most innovative insights, which typically originate from a small minority of highly engaged users (Wu & Huberman, 2007). The underdevelopment of analytic strategies for extracting, synthesizing, and applying these innovation-driving signals motivates the need for new frameworks that combine design principles with computational analysis. The following section introduces the Lead UX framework.

3. The Lead UX Framework: Harnessing Innovation from User-Generated Content

The Lead UX framework addresses gaps left by participatory design (PD), user-centered design (UCD), and conventional user-generated content (UGC) analytics in identifying and leveraging innovation-driving insights. By focusing on advanced, future-oriented users and emergent needs surfaced in digital contexts, Lead UX integrates the voices and foresight of strategically important contributors into information systems (IS) design (Baldwin & von Hippel, 2011; Schreier & Prügl, 2008).

This section positions Lead UX among established design approaches, defines the characteristics and strategic value of lead users, and outlines methods for identifying both lead users and lead topics.

3.1 Positioning Lead UX Among Design Approaches

Traditional PD distinguishes itself by directly involving a broad range of stakeholders through collaborative tools and workshops (Muller & Kuhn, 1993; Bratteteig & Wagner, 2014). UCD centers on usability and satisfaction, typically assessed through interviews, surveys, and usability tests, with an emphasis on the average end-user (Nielsen, 2024; Norman, 2013; Kujala, 2003). UGC analytics broadens this focus to capture input from a wide spectrum of everyday users via online platforms, employing both quantitative (e.g., engagement metrics) and qualitative (e.g., topic or sentiment analysis) techniques to reveal consumer attitudes, pain points, and shifting expectations (Wu & Huberman, 2007; Goes, 2014). The comparative matrix, presented below as Table 1, demonstrates how Lead UX uniquely centers advanced, future-oriented users and hybrid analytic workflows, filling significant gaps left by traditional PD and UCD approaches in the context of large-scale digital ecosystems.

Factor	**Participatory Design (PD)**	**User-Centered Design (UCD)**	**User-Generated Contents (UGC) Analytics**	**Lead User Experience (Lead UX)**
Involvement	Involves all stakeholders directly in the design process.	Focuses on understanding and meeting the needs of end-users.	Everyday users are central including anyone who voluntarily shares their experiences online	Involves advanced users who are ahead of market trends. Investigates user leads in data.
Users	All stakeholders, including all end-users, are involved.	End-users are involved through interviews, surveys, and testing.	Broad range of users, including customers and fans, who share authentic experiences or	Advanced users with strong innovation needs. Users who express their ideas in online contexts.

			feedback in digital public forums or branded spaces	
Focus	Collaboration and democratization of the design process.	Usability and user satisfaction.	Mining insights about attitudes, pain points, satisfaction, and emerging needs	Future-oriented innovation. Emergent needs in online forums.
Execution Method	Ongoing collaboration with a broad range of stakeholders using tools such as workshops and brainstorming sessions	Iterative testing and refinement based on user feedback. User interviews, surveys, personas, and usability testing	Both quantitative (such as engagement, reach) and qualitative (such as theme, sentiment) approaches	Engagement with a small group of lead users and identifying leads, e.g. common user issues.
Tools to Identify User	Workshops, surveys, interviews, observations	User studies, surveys, interviews, observations	Social media listening platforms, AI-driven analytics, hashtag tracking, content management systems	Data analytic tools to identify lead users and common user issues to innovate
Related References	Muller & Kuhn (1993); Bratteteig & Wagner (2014)	Nielsen (2024); Norman (2013); Kujala (2003)	Wu & Huberman (2007); Goes (2014)	Baldwin & Von Hippel (2011); Schreier & Prügl (2008)

Table 1. Comparison of PD, UCD, UGC Analytics, and Lead UX: Key Factors, User Roles, and Distinctive Features.

Lead UX integrates the advantages of PD, UCD, and UGC analytics, as summarized in Figure 1. This framework advances beyond these established practices in two fundamental ways. First, it explicitly seeks out lead users, those whose needs, capabilities, and motivations are ahead of contemporary market trends. Second, it leverages scalable analytics to identify not only individual insights but also emergent collective concerns (lead topics) that signal future innovation opportunities. Figure 1 illustrates how Lead UX synthesizes and builds on the strengths of PD, UCD, and UGC analytics to offer a more comprehensive and forward-looking design framework.

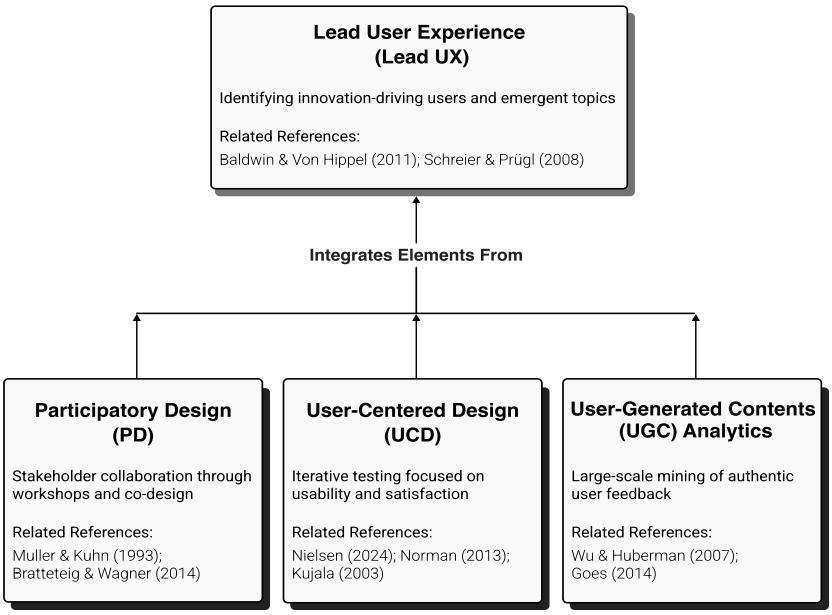


Figure 1. Relationship of PD, UCD, UGC Analytics, and Lead UX: Overlapping roles and strengths in IS design

3.1.1 Theoretical Propositions for Lead UX

While this viewpoint article presents Lead UX as a conceptual framework, the following propositions outline testable claims that future empirical research should investigate. These propositions clarify the theoretical contributions of Lead UX and set benchmarks for future validation studies. Since the main goal of developing the Lead UX framework is to create accurate, timely, relevant, and scalable interfaces unified in function and appearance, our propositions focus on these three aspects. As we conduct future research, validating these propositions against existing approaches will help develop cost-effective, scalable interfaces and market-aligned user experiences.

Proposition 1 (Early Detection): Lead UX will identify innovation opportunities earlier in the product development lifecycle than traditional PD or UCD methods, enabling proactive design decisions rather than reactive adaptations to already-established user demands.

Proposition 2 (Market Alignment): Design decisions informed by lead-user insights identified through Lead UX will demonstrate higher market acceptance and commercial relevance than those based solely on average-user feedback, because lead users' needs anticipate mainstream market trends.

Proposition 3 (Scalable Rigor): The hybrid automated-human workflow of Lead UX will enable systematic analysis of user-generated content at scale without sacrificing contextual validity, by leveraging computational efficiency for initial clustering while preserving human interpretive rigor for validation and labeling.

These propositions are not tested in this article but serve to guide future empirical efforts and to position Lead UX within a broader research agenda for IS design innovation.

3.2 Characteristics of Lead Users

Lead UX centers on advanced users whose contributions anticipate future mainstream demand. Drawing on Lüthje and Herstatt (2004) and Schreier and Prügl (2008), lead users are characterized by:

1. High product knowledge: deep familiarity with system features and technical nuances, enabling specific and technically informed feedback.

2. Strong locus of control: a proactive stance toward shaping outcomes, reflected in proposing remedies or alternatives rather than only reporting problems.

3. Innovativeness: willingness to experiment with, repurpose, or extend existing features, often revealing novel use cases.

4. Superior UX awareness: heightened sensitivity to usability and workflow issues and the ability to clearly articulate these concerns for designers and peers. These traits make lead users especially valuable in digital environments in which innovation-relevant content is concentrated among a small subset of highly engaged contributors.

3.3 Strategic Importance of Lead Users

Lead users are more than early adopters; they detect needs, performance gaps, and opportunities well before they are visible to the broader market (von Hippel, 2005). Their intensive, advanced use allows them to identify subtle friction points and envision breakthrough features.

Empirical studies demonstrate the commercial value of systematically engaging lead users. Urban and von Hippel (1988) found that innovations guided by lead-user insights were significantly more likely to succeed commercially. Lüthje and Herstatt (2004) showed that 38% of mountain biking users had developed product modifications, illustrating the prevalence of user-generated innovations that traditional methods often miss. Enthusiasts in Marin County, California, for example, created off-road "clunker" bicycles years before manufacturers recognized and served this market (von Hippel, 2006).

Digital platforms offer parallel cases. Chris Messina's (2007) proposal of the hashtag (#) on Twitter, later documented by Huddleston (2020), illustrates how a lead user introduced a practice that became central to platform functionality. Such contributions show how lead users can shape core system features and social conventions.

Organizational examples underscore similar dynamics. At 3M, products developed through systematic lead-user engagement achieved annual sales of approximately $146 million after five years, about eight times those of products developed through traditional approaches (Lilien et al., 2002). LEGO Ideas enables advanced fans to submit and refine new set concepts; selected fan-designed sets have become bestsellers, demonstrating how structured lead user engagement can reduce market risk and strengthen community ties (King & Lakhani, 2013). Similar effects can be seen in the information domain when we observe "power editors" set the agenda on user-generated content pages (Panciera et al., 2009)

Lead users also serve as valuable partners for iterative prototyping and co-design. Their willingness to experiment, test new workflows, and provide detailed feedback creates a living laboratory effect, enabling organizations to validate hypotheses, refine features, and develop solutions with those most invested in system advancement. By systematically surfacing and analyzing lead user contributions, organizations tap into innovation signals that inform both immediate feature improvements and broader strategic directions. Recognizing and leveraging lead user insights is thus central to the Lead UX approach, transforming overlooked feedback into sustained innovation.

3.4 Identifying Lead Users: From Manual Coding to Human-Guided Automated Analytics

Traditionally, lead users have been identified through manual coding of qualitative data, an approach suitable for small-scale PD or UCD but not for the volume and diversity of contemporary UGC (Krippendorff, 2018). Lead UX instead adopts a hybrid strategy that combines automated analysis with expert human interpretation.

Computational tools such as topic modeling and other natural language processing (NLP) techniques first cluster large text corpora around salient themes (Wu & Huberman, 2007; Goes, 2014). These clusters may involve repeated complaints, detailed feature suggestions, or advanced workarounds. Human analysts then review and label the clusters, applying the four lead-user characteristics—high product knowledge, strong locus of control, innovativeness, and superior UX

awareness—to identify which contributors are likely lead users.

This workflow enables IS professionals to efficiently identify both pervasive concerns and high-potential lead users, without manually inspecting all UGC. Future research can refine this process by developing automated lead-user scoring models and comparing alternative NLP and classification techniques across platforms and domains.

3.5 Lead Topics: Signal Amidst the Noise

Lead UX detects innovation signals at both the individual (lead user) and community (lead topic) levels. This builds on work on collective intelligence, which shows that large-scale user input can reveal opportunities and risks not visible from isolated observations (Surowiecki, 2004; Ickler & Baumöl, 2012).

Lead topics are recurring, persistent themes in UGC that point to high-impact opportunities for system improvement. They are typically discovered through automated clustering of user comments and posts into thematic groups. Human analysts then assess which clusters represent genuine innovation signals, drawing on lead-user traits (e.g., technical depth, solution focus) and considering the novelty and practical relevance of the ideas expressed.

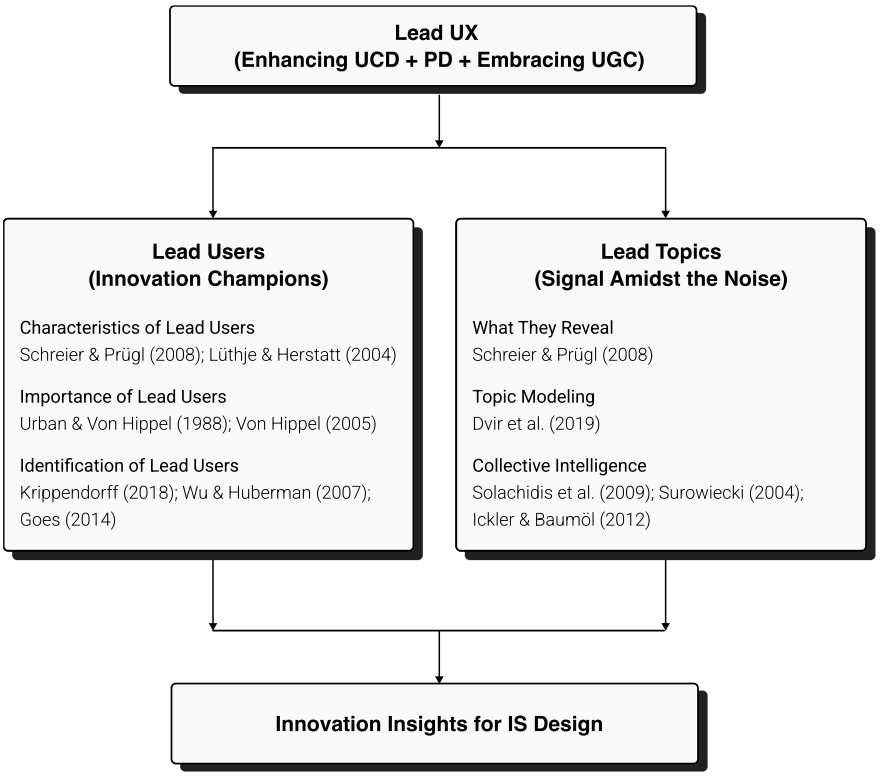


Figure 2. Identifying Lead Users and Topics: Workflow for extracting innovation from user-generated content

By combining lead user identification with lead topic analysis, as illustrated in Figure 2, Lead UX helps IS professionals shift from reactive adaptation, responding to well-established demands, to proactive innovation guided by early signals from user communities. The following section explains how these concepts can be operationalized in practice, translating the framework into concrete workflows for research and development.

4. Operationalizing Lead UX in Information Systems

To move Lead UX from concept to practice, information systems researchers and practitioners need a hybrid workflow that combines computational analytics with expert human interpretation. The process starts with the systematic collection and cleaning of user-generated content from relevant digital platforms, ensuring data quality and integrity. Prior work

outlines protocols for managing unstructured user comments and extracting actionable insights, including best practices for preprocessing and analytic rigor (Mitchell, 2018; Lebens, 2021). Natural language processing techniques, such as topic modeling, then organize this content into thematic clusters that reflect recurring concerns, emerging needs, or innovative suggestions. Human analysts review these clusters and apply established lead user criteria—high product knowledge, strong locus of control, innovativeness, and superior UX awareness—to identify those most likely to yield forward-looking insights. These high-potential clusters inform system enhancements, help prioritize development efforts, and guide ongoing user engagement strategies.

Lead UX improves scalability and supports early identification of innovation opportunities, but it also faces limitations in data quality and analytic rigor. Large volumes of user-generated content increase the risk of including fake or bot-generated reviews, which can distort findings if not properly filtered. Robust data quality procedures, including automated detection and removal of suspicious or non-human entries, are therefore essential preprocessing steps.

Analytic rigor presents a second challenge. Any workflow involving human coding or interpretation, manual or hybrid, is vulnerable to bias. In manual coding, different raters may classify the same reviews differently; accordingly, inter-rater reliability measures, such as the intraclass correlation coefficient (ICC), are used to establish consensus and build confidence in the coded results. In topic modeling and other semi-automated approaches, algorithms first cluster themes from large text corpora and humans then label the output. Because grouping is performed automatically, this interpretive stage does not readily lend itself to ICC-style metrics. Instead, analysts should emphasize clear documentation of labeling criteria and periodic validation exercises to minimize interpretive bias. In practice, both human coding and automated clustering contribute to data integrity and analytical rigor. Used together, they complement each other in extracting information and reducing noise in user-generated comments, thereby advancing the development of the Lead UX paradigm. Continued improvements in data quality management and analytic transparency are essential to strengthen the rigor and trustworthiness of Lead UX implementations, regardless of the specific workflow adopted.

4.1 Future Research Directions and Proposed Evaluation Approaches

As a viewpoint article, this paper offers a theoretical and practical foundation for Lead UX rather than empirically testing the framework. Although the core methodology and conceptual claims are grounded in prior literature, empirical research is needed to assess the effectiveness of Lead UX across organizational contexts, platforms, and user populations.

Future work should pursue three complementary validation approaches. First, comparative case studies can examine how Lead UX identifies innovation opportunities relative to traditional design methods (Yin, 2009; Eisenhardt & Graebner, 2007), documenting both timing and commercial relevance. Second, expert practitioner assessments can evaluate feasibility, scalability, and implementation requirements across organizational settings (Wandersman et al., 2012; Slaghuis et al., 2011). Third, algorithmic benchmarking can compare automated lead user–topic identification against manual expert coding (Hripcsak & Rothschild, 2005), using metrics such as precision, recall, and F1-score across platforms. Prior work suggests that hybrid automated–manual approaches often outperform single methods in classification accuracy and interpretability, a claim that future Lead UX studies can test explicitly. Together, these investigations will clarify the conditions under which Lead UX delivers the most significant strategic value and inform the implementation of best practices.

5. Call to Action: Transforming IS Practice Through Lead UX

The volume and velocity of user-generated content (UGC) are reshaping IS design. Slow responses to emerging user signals can lead to costly redesigns, reduced satisfaction, and missed opportunities for innovation. Lead UX offers a disciplined, scalable path for shifting from reactive adaptation to proactive innovation.

Evidence from established lead user practices underscores this potential. At 3M, products developed through systematic lead-user engagement achieved annual sales of about $146 million after five years, eight times those of products developed through traditional methods (Lilien et al., 2002). LEGO's crowdsourced innovation platform, LEGO Ideas, likewise transforms fan submissions into best-selling products by systematically identifying and amplifying advanced users' ideas, demonstrating that organizations can scale lead user engagement while maintaining community connection (King &

Lakhani, 2013).

IS professionals and educators can operationalize Lead UX through the following steps:

- **Integrate into curricula:** Use community-based service-learning to have students apply Lead UX analytics to real small-business challenges, enhancing both learning and local innovation capacity (Mitchell, 2018).
- **Design interactive classwork:** Employ prototyping assignments in asynchronous online courses, so students iterate on system designs informed by Lead UX signals (Mitchell, 2018).
- **Blend computational and interpretive analysis:** Combine topic modeling and other NLP tools with expert review to identify lead users and emerging topics at scale, augmenting rather than replacing qualitative methods.
- **Embed lead-topic monitoring in development:** Build regular cadences (e.g., within Agile sprints) for reviewing and prioritizing emergent user signals, turning high-potential findings into hypotheses for rapid prototyping and testing.
- **Cultivate lead user engagement channels:** Establish sustained mechanisms, such as advisory panels, beta communities, or dedicated feedback spaces, that systematically involve advanced users in design decisions and treat them as partners in innovation.

By adopting Lead UX, IS professionals can redefine user feedback from a retrospective evaluation tool into a real-time driver of innovation. Automated topic detection and human interpretation become integral to iterative development, enabling teams to detect friction points and opportunities earlier than competitors. Treating users as co-designers promotes continuous, participatory innovation, systematically surfacing breakthrough ideas from user communities and reducing risk in development cycles.

6. Conclusion

This viewpoint article has addressed two central research questions: (1) How can IS professionals systematically identify and amplify the perspectives of lead users whose insights may signal emergent needs but are often overlooked in traditional large-scale analyses? (2) What analytic strategies enable the scalable extraction of actionable findings from large volumes of unsolicited user content?

We introduced Lead UX, a framework that integrates von Hippel's lead-user innovation theory with scalable computational analytics and human interpretive rigor. Lead UX positions advanced, innovation-driving users and emergent collective concerns (lead topics) as focal points for user research, complementing rather than replacing participatory design and user-centered design. By combining automated clustering techniques with expert interpretation grounded in established lead-user characteristics, Lead UX enables organizations to convert big data challenges into strategic opportunities for sustained innovation.

The framework has implications for both research and practice. For researchers, Lead UX offers avenues for empirical validation through comparative case studies, expert practitioner assessments, and algorithmic benchmarking, which can clarify when and why Lead UX outperforms traditional design approaches. For practitioners, Lead UX provides implementable workflows for monitoring UGC, identifying high-potential innovations, and prioritizing design decisions, which can accelerate time-to-market and reduce innovation risk.

Limitations remain. Empirical effectiveness across different organizational contexts, platforms, and user populations is still to be tested. Optimal tuning of the hybrid workflow, balancing automated efficiency with human interpretive accuracy, requires further investigation, as do potential biases introduced by algorithmic clustering or analyst interpretation.

The next era of IS design depends less on isolated genius than on combining analytics with community-driven creativity. By systematically revealing, interpreting, and acting on latent innovation within user communities, IS professionals can transform design from a reactive, inward-facing discipline into a proactive, collaborative partnership with users. The call to

action is to activate Lead UX now: use it to unlock strategic foresight, refine analytic practices through rigorous testing, and position users as central co-creators in sustainable innovation.

7. References

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. In W. Bainbridge (Ed.), *Berkshire encyclopedia of human-computer interaction* (pp. 763-768). Berkshire Publishing Group.

Baldwin, C., & von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, 22(6), 1399-1417.

Bratteteig, T., & Wagner, I. (2014). *Disentangling participation: Power and decision-making in participatory design*. Springer.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.

Dvir, N., & Gafni, R. (2019). Systematic improvement of user engagement with academic titles using computational linguistics. *arXiv preprint arXiv:1906.09569*.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32.

Goes, P. B. (2014). Editor's comments: Big data and IS research. *MIS Quarterly*, 38(3), iii-viii.
Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the F-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, 12(3), 296–298.

Ickler, H., & Baumöl, U. (2012). Adding value with collective intelligence: A reference framework for business models for user-generated content. In J. Altmann, U. Baumöl, & B. Krämer (Eds.), *Advances in collective intelligence 2011* (pp. 41-50). Springer.

King, A., & Lakhani, K. R. (2013). Using open innovation to identify the best ideas. *MIT Sloan Management Review*, 55(1), 41.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage Publications.

Klein, B. D. (2017). On the development and application of a framework for understanding the properties and information quality of online reputation systems. *Journal of the Midwest Association for Information Systems*, 2017(1), Article 3.

Koskinen, I., Zimmerman, J., Binder, T., Redström, J., & Wensveen, S. (2011). *Design research through practice: From the lab, field, and showroom*. Morgan Kaufmann.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications.

Krumm, J. (2018). *Ubiquitous computing fundamentals*. CRC Press.

Kujala, S. (2003). User involvement: A review of the benefits and challenges. *Behaviour & Information Technology*, 22(1), 1-16.

Kuutti, K. (2009). User-centered design and beyond: Towards more human-centered design. *Design Issues*, 25(3), 29-41.

Lebens, M. (2021). Using prototyping to teach the design thinking process in an asynchronous online course. *Journal of the Midwest Association for Information Systems*, 2021(2), Article 3.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., & von Hippel, E. (2002). Performance assessment of the lead user idea-generation process for new product development. *Management Science*, 48(8), 1042-1059.

Lüthje, C., & Herstatt, C. (2004). The lead user method: An outline of empirical findings and issues for future research. *R&D Management*, 34(5), 553-568.

Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies*, 55(4), 587-634.

Messina, C. [@chrismessina]. (2007, August 23). How do you feel about using # (pound) for groups. As in #barcamp [Tweet]. *Twitter*. https://twitter.com/chrismessina/status/236136209

Mitchell, A. J. D. (2018). Small business website development: Enhancing the student experience through community-based service learning. *Journal of the Midwest Association for Information Systems*, 2018(2), Article 4.

Muller, M. J. (2007). Participatory design: The third space in HCI. In A. Sears & J. A. Jacko (Eds.), *Human-computer interaction handbook* (pp. 1051-1068). Lawrence Erlbaum Associates.

Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM*, 36(6), 24-28.
Nielsen, J. (2024, January 30). 10 usability heuristics for user interface design. *Nielsen Norman Group*. https://www.nngroup.com/articles/ten-usability-heuristics/

Norman, D. A. (2013). *The design of everyday things* (Revised and expanded ed.). Basic Books.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly*, 36(3), 841-864.

Schreier, M., & Prügl, R. (2008). Extending lead-user theory: Antecedents and consequences of consumers' lead userness. *Journal of Product Innovation Management*, 25(4), 331-346.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices*. Lawrence Erlbaum Associates.

Slaghuis, B., Steding-Jessen, M., Carvalho, S. W., Evers, W., & Ouwens, M. (2011). Framework and tools for evaluating scalability of health interventions. *Health Policy*, 105(2–3), 154–162.

Solachidis, V., Dimitriadis, S., Drosou, A., Tzovaras, D., & Konstantinidis, K. (2009). Collective intelligence generation from user contributed content. In A. Fink, B. Lausen, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence* (pp. 769-778). Springer.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday.

Urban, G. L., & von Hippel, E. (1988). Lead user analyses for the development of new industrial products. *Management Science*, 34(5), 569-582.

von Hippel, E. (2005). *Democratizing innovation*. MIT Press.

Wandersman, A., Duffy, J., Firebaugh, P., Noonan, R., Lubell, K., Malloy, P., ... & Saul, J. (2012). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 50(3–4), 445–459.

Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45), 17599-17601.

Yin, R. K. (2009). Case study research: *Design and methods* (4th ed.). SAGE Publications.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), 274-291.

Author Biographies

Javad Norouzi Nia is an Assistant Professor of Business Information Systems at Central Michigan University. With expertise in user experience, human-computer interaction, and data analytics, his research focuses on understanding how people interact with digital systems. In healthcare, he uses predictive analytics and machine learning to develop innovative methods for analyzing eye-tracking data, with a particular focus on predicting health outcomes, such as chronic pain. His work has been published in peer-reviewed outlets. In addition to his research, Dr. Norouzi Nia teaches courses in Data Analytics, Database Management, and IS Capstone. His teaching philosophy emphasizes experiential learning and project-based instruction, equipping students with practical, industry-relevant skills. Through capstone projects involving interactive prototype development and real-world data management, he fosters a balance between theoretical understanding and hands-on application, preparing students to excel in the dynamic fields of business and information systems.

Vishal Shah is an Associate Professor of Information Systems at Central Michigan University. His research interests include IS pedagogy, neuro IS, UX design, social media, IT and learning, game-based learning, and telemedicine. He has published in journals such as AIS Transactions on Replication Research, International Journal of Telemedicine Applications, Journal of Education for Business, and Technical Communication Quarterly. He is also Director of University Alliance Program at Central Michigan University, aimed at delivering relevant and rigorous education in keeping with changing technology, and works with leading forms like SAP, Celonis, Dow, and many Fortune 500 companies, and is a recipient of the Edwin J Towle Professorship for research excellence.

Gustav Verhulsdonck is an associate professor in the Business Information Systems department at Central Michigan University. His research interests are UX design, artificial intelligence, and technical communication. He has published *UX Writing: Designing User-Centered Content* (Routledge, 2023) together with Jason Tham and Tharon Howard.