

Guess, Learn, Repeat: Intelligent Learning System with Synthetic and Counterfactual Training in a GeoGuessr-Inspired Classification Task

Marc Goutier
TU Darmstadt
goutier@ise.tu-darmstadt.de

Philipp Spitzer
Karlsruhe Institute of Technology
philipp.spitzer@kit.edu

Domenique Zipperling
University of Bayreuth & Fraunhofer FIT
domenique.zipperling@uni-bayreuth.de

Abstract

Training novices by experts is often costly and time-consuming. Alternatively, learning systems offer a scalable and automated alternative. However, learning systems offer another, yet underexplored advantage, over training with experts: Analyzing novices and providing personalized training. This study explores the use of synthetically generated images to improve novice image classification skills in a GeoGuessr-inspired classification task. By leveraging a counterfactual-based approach and synthetically generated personalized training data, we aim to enhance individual learning. In a controlled experiment where participants classify Google Street View images from four different cities, we compare the impact of personalized synthetic images against randomly assigned ones. Our findings indicate that personalized training improves classification accuracy, underscoring the potential of intelligent learning. These results highlight a promising direction for integrating synthetic data into adaptive training environments in game-like settings, paving the way for effective and personalized intelligent learning systems.

Keywords: Personalized Learning, Synthetic Data, Counterfactuals, GeoGuessr, Intelligent Learning System

1. Introduction

Artificial intelligence (AI) is transforming various domains, from medical diagnostics to financial forecasting and has been frequently mentioned as solution to battle shortage of skilled workers (Carioli et al., 2024; Madgavkar and White, 2025). One promising application are learning systems (Argote, 2015; Kochmar et al., 2022; Wong and Huang, 2011), which support training in complex tasks such as image classification. With skilled worker shortages, scalable personalized learning systems are vital, particularly in industries reliant on subject matter experts (SMEs). To stay competitive, organizations must upskill employees

(Clarizia et al., 2021), especially during times of demographic change (Engbom, 2018) resulting in competition for talent (Kwon and Jang, 2022). This intensifies the demand for more autonomous and effective teaching systems (Kang and Jackson, 2021). Traditionally, training novices requires expert guidance, which is both costly and time-consuming. As prior research highlights, conventional learning systems struggle to transfer tacit knowledge effectively, often requiring extensive hands-on expert involvement (Serban et al., 2020). To address this, learning systems provide automated and scalable alternatives, leveraging AI to facilitate learning (Gross et al., 2015; Spitzer et al., 2023). However, while AI-driven training can replace some aspects of expert instruction, its full potential remains untapped. For instance, research on how AI can be used to teach novices a new task but focus on a limited task domain with real data (Spitzer et al., 2024). In particular, the ability of AI to analyze individual learning patterns on the one hand and offer personalized training on the other has not yet been fully explored. Current learning systems often rely on standardized training programs that do not adapt to the specific needs of individual learners. Existing approaches, such as explainable AI (XAI)-based learning (Goyal et al., 2019; Vandenhende et al., 2022), aim to improve novices' understanding by providing explanations alongside generated examples (Spitzer et al., 2022). However, these methods primarily focus on making AI predictions interpretable rather than optimizing learning pathways. This one-size-fits-all approach limits the effectiveness of training. Furthermore, training datasets are often limited, making it difficult to expose learners to diverse and challenging examples (Baier et al., 2019). One potential solution is the use of synthetically generated training data, which can provide an unlimited supply of tailored learning material (Allmendinger et al., 2024), for example in the form of counterfactual examples (Goyal et al., 2019). However, the effectiveness of synthetic data in improving human learning outcomes remains an open question. Only few works have investigated how synthetically generated data can be used for training AI itself (Abufadda and

Mansour, 2021) or using it for educational purposes (Bulmer et al., 2022). For instance, while investigations on how AI-based training can improve novices' learning performance exist, they do not investigate the potential on utilizing synthetic data to further improve the system (Spitzer et al., 2024).

If training systems fail to provide personalized and effective learning experiences, novice learners may struggle to develop accurate classification skills, reducing overall classification accuracy. Prior research suggests that cognitive styles influence learning outcomes, meaning that learning is a very individual process and personalization might have a big influence on learning effectiveness (Spitzer et al., 2024). Moreover, a lack of personalized feedback can lead to overconfidence in incorrect classifications or excessive reliance on AI assistance without genuine skill acquisition (Diebel et al., 2025). Addressing this issue is essential for ensuring that learning systems enhance rather than hinder human learning.

Therefore, we propose the following research question: *How does personalized training using synthetically generated and counterfactual images, tailored to learners' previous errors, affect novices' classification accuracy?*

To investigate the impact of personalized synthetically generated training data, we conducted a controlled experiment in which participants classified Google Street View images in a GeoGuessr-inspired classification task (Walter, 2024). By adopting a counterfactual approach (Dai et al., 2022; Goyal et al., 2019), we generated synthetic images tailored to participants' prior mistakes, providing personalized learning experiences. This approach differs from prior XAI-based systems (Spitzer et al., 2024), as it does not merely explain classification decisions but actively adapts the training process to reinforce correct classification patterns. We conducted a between-subject experiment where we compared the classification accuracy of individuals receiving personalized synthetic images against those receiving random synthetic images to assess the impact of personalized training.

This study demonstrates that personalized synthetic training images improve classification accuracy, highlighting the potential of synthetically generated learning environments. Our findings suggest that counterfactual training data can enhance human learning by providing personalized feedback and tailored training data. By extending prior work on AI-based learning, we show that synthetic data can serve not only as an alternative to real-world training datasets but also as a tool for dynamic, personalized instruction. This research contributes to the growing field of learning systems by offering empirical evidence on the effectiveness of synthetic data in adaptive training.

Furthermore, our work lays the foundation for future research into scalable, AI-driven personalized learning by intelligent learning systems across various domains.

2. Theoretical Background

2.1. Learning Systems

Learning systems play a vital role in supporting the acquisition of new skills (Argote, 2015). Traditional systems such as e-learning platforms and tutoring systems have shown success in organizational learning contexts (Kochmar et al., 2022; Wong and Huang, 2011), but they are often context-specific and costly to develop and maintain (Derouin et al., 2005; Serban et al., 2020). More recently, advances in AI and XAI have introduced more scalable and flexible alternatives to human teaching. Machine teaching, for example, uses AI to convey domain knowledge through curated examples (Zhu et al., 2018), and recent research has explored how XAI can enhance these systems by offering explanatory feedback, including counterfactual explanations (Goyal et al., 2019; Spitzer et al., 2022; Vandenhende et al., 2022).

Unlike traditional automation, (X)AI-based systems aim not just to perform tasks but to transfer expert knowledge—both explicit and tacit—by modeling and explaining expert decision-making behavior (Sanzogni et al., 2017). Because learning is embedded in a socio-technical system involving both humans and AI, the effectiveness of (X)AI-based instruction may hinge on how well it aligns with learners' preferences (Szymanski et al., 2021).

Recent studies have begun to demonstrate the potential of such systems for supporting novice skill acquisition, particularly when training is personalized and aligned with learners' cognitive needs (Spitzer et al., 2024). Yet, with the rise of generative AI, little is known about how these systems can be tailored to adapt to users' needs and fully leverage the generative capabilities of AI, such as the creation of synthetic training data (Feuerriegel et al., 2024).

Designing intelligent learning systems, such as those generating supplementary examples, requires a human-centered approach that prioritizes user values, ethics, and usability by asking “what is needed?” rather than “what is possible.” Human-centered AI emphasizes augmenting human capabilities while ensuring transparency, fairness, and trust (Alfredo et al., 2024). In education, this approach fosters inclusive, ethical, and responsive tools (Luo et al., 2025), including game-based environments that adapt challenges and feedback to sustain engagement while respecting learner autonomy and well-being (Miles et al., 2024).

2.2. Google Street View Images and GeoGuessr

Google Street View offers georeferenced, 360-degree panoramic imagery of real-world environments, enabling detailed visual analysis of diverse global locations (Ijaz et al., 2019). GeoGuessr builds on this by transforming Street View into an interactive spatial reasoning task, where users infer geographic locations based solely on visual cues (Farmer and Ho, 2025). Both platforms have been employed in studies on human–AI collaboration and human–computer interaction. Google Street View, in particular, has been used to investigate the impact of XAI on user reliance (Walter, 2024) and performance (Walter et al., 2023), as well as to support user studies exploring perception, navigation, and decision-making. Additionally, it has been utilized in virtual reality contexts (Ijaz et al., 2019) to examine spatial cognition and immersive interaction.

3. Hypothesis Development

Educational psychology theories like mastery learning (Bloom, 1984) and scaffolding (Fan et al., 2025) emphasize the importance of presenting challenges that are slightly beyond a learner’s current competence level (Bloom, 1984). Furthermore, learning from one’s errors is a learning strategy that, while often avoided, has been shown to yield promising learning results, especially when coupled with corrective feedback on mistakes, as it not only fosters deeper understanding but also improves long-term performance (Metcalfe, 2017). Therefore, personalized training or learning represents a pedagogically valid and evidence-based strategy, when being grounded in the principle that learning is most effective when instructional content (e.g., corrective feedback) aligns with an individual’s prior performance and specific learning needs (Kan, 2024). With personalized feedback, learners are able to focus their attention on task-relevant features, thereby facilitating more efficient encoding, retrieval, and refinement of mental models. Empirical studies on adaptive learning systems (e.g., tutoring systems, personalized recommender engines, and e-learning platforms) show that personalized interventions outperform standardized instruction in terms of learning outcomes (Grenander et al., 2021; Kochmar et al., 2020; Scarlatos et al., 2024; Troussas et al., 2023).

In learning systems, personalized feedback has been realized through adaptive content delivery that dynamically responds to learner input, often guided by XAI techniques such as counterfactual explanations (Goyal et al., 2019), which can support understanding by illustrating alternative outcomes (e.g., underlying decision boundaries). Similarly, large language models

like GPT have been employed to generate natural language justifications that resemble tutor-like feedback (Scarlatos et al., 2024). While these approaches improve the interpretability of synthetically generated content, they often fall short in delivering instruction that is both targeted and tailored to the learner’s individual errors.

In summary, foundational work in educational psychology highlights that optimal learning occurs when challenges slightly exceed current ability and are supported by personalized feedback. While AI-based personalized feedback improves outcomes, it often overlooks specific errors. Tailoring corrective input to these individual mistakes would further increase personalization and, as supported by theories of mastery learning and scaffolding (Bloom, 1984; Fan et al., 2025), is expected to strengthen training effects. Our study tests this by delivering personalized corrective feedback through counterfactual and synthetic materials in a GeoGuessr-inspired image classification task. Therefore, we state the following hypothesis:

H1: *Participants receiving personalized counterfactual synthetic images will show greater improvements in classification accuracy than those receiving randomly assigned synthetic images.*

Beyond personalization, this study also investigates whether learning occurs using only synthetic generated training materials. When learning is measured by classification accuracy, theories of contextual learning and simulation fidelity stress that effective transfer requires generated images to reflect reality, sharing critical elements with the operational environment (Brown et al., 1989; Renganayagalu et al., 2019).

Emerging research on synthetic data suggests that generative models can augment or even replace real data for training purposes, provided the synthetic content captures task-relevant features (Liu et al., 2024). In our case, the classification task is visually grounded. While real-world images are commonly used in training tasks, recent advances in generative AI allow for the simple generation of high-quality synthetic images that closely resemble real environments (Adamkiewicz et al., 2025). If such images can effectively convey geographic visual cues (e.g., by controlling the training images for realism and accuracy by a pretrained classification model), then, based on e.g., contextual learning or fidelity in simulation theory, exposure to synthetically generated images, if personalized or not, should enable learning. Therefore, we hypothesize that participants will demonstrate learning gains over time with synthetically generated training data, regardless if the participants are exposed to personalized or unpersonalized, synthetically generated images.

H2: *Participants will improve their classification accuracy over time, demonstrating that learning is possible with synthetically generated images.*

4. Method

4.1. Experiment Design



Figure 1. Illustration of the experimental design.

We conducted a controlled between-subjects experiment to investigate the impact of personalized counterfactual synthetic images on novice learning in a GeoGuessr-inspired image classification task (Walter, 2024). We pre-registered our experiment with the Open Science Foundation (OSF) (Goutier, 2024) where we estimated to need at least 74 participants to achieve significant results. Data collection took place in June 2024 in an online experiment. Participants were randomly assigned to one of two conditions:

Personalized Training Condition: Participants received synthetic training images tailored to their previous misclassifications, using counterfactuals.

Random Training Condition: Participants received synthetic training images selected randomly from the overall dataset, independent of their mistakes. The experiment began with an illustration and explanation of the main question type, followed by an introductory page informing participants that the survey

questions would now begin. The core of the experiment consisted of a repeated cycle of test and training phases (Kühl et al., 2022). In the test phase, participants were presented with five real-world images to classify, with four possible answers: Berlin, Hamburg, Jerusalem, and Tel Aviv (Walter et al., 2023). Only one answer was correct. In the subsequent training phase, participants were given five synthetic images with labels, which they could study for as long as they needed. This cycle was repeated four times, with the final cycle consisting of a test phase only. The experiment ended with a demographic questionnaire. Figure 1 gives an overview of the experimental design.

The training images were real pictures of the four cities and similar in both conditions. While in the random training condition the selection of the training images from the pool of all available synthetic images was randomized, the logic for the personalized training condition was more advanced. We tracked twelve different counters for each possible error (e.g., selecting Hamburg for a Berlin image would increase the "BerlinHamburg" counter).

For each misclassification, the synthetic images shown during the training phase were specifically tailored to address that error. For example, if a participant misclassified a real image of Berlin as Hamburg, they would receive a synthetic image of Berlin that had been generated to resemble Hamburg, serving as a counterfactual visual contrast. This helped the learner recognize distinctive features of Berlin that differentiate it from Hamburg, and vice versa.

If no mistakes were made in the testing phase, five random images were shown in the training phase. If a participant made only one type of mistake when testing (e.g., the participant made the mistake of selecting Hamburg for a Berlin image, the BerlinHamburg was increased), they received three images for that mistake and two for the opposite mistake in the following training phase (e.g., the BerlinHamburg counter was the only registered value, therefore the participant receives three images for the counterfactual BerlinHamburg and two images for the counterfactual HamburgBerlin). The two highest counters determined the images shown in the training phase, with three images corresponding to the highest counter and two to the second highest. If several error counters were equally high, one of them was chosen at random. Figure 2 represents an exemplary first cycle of testing and training.

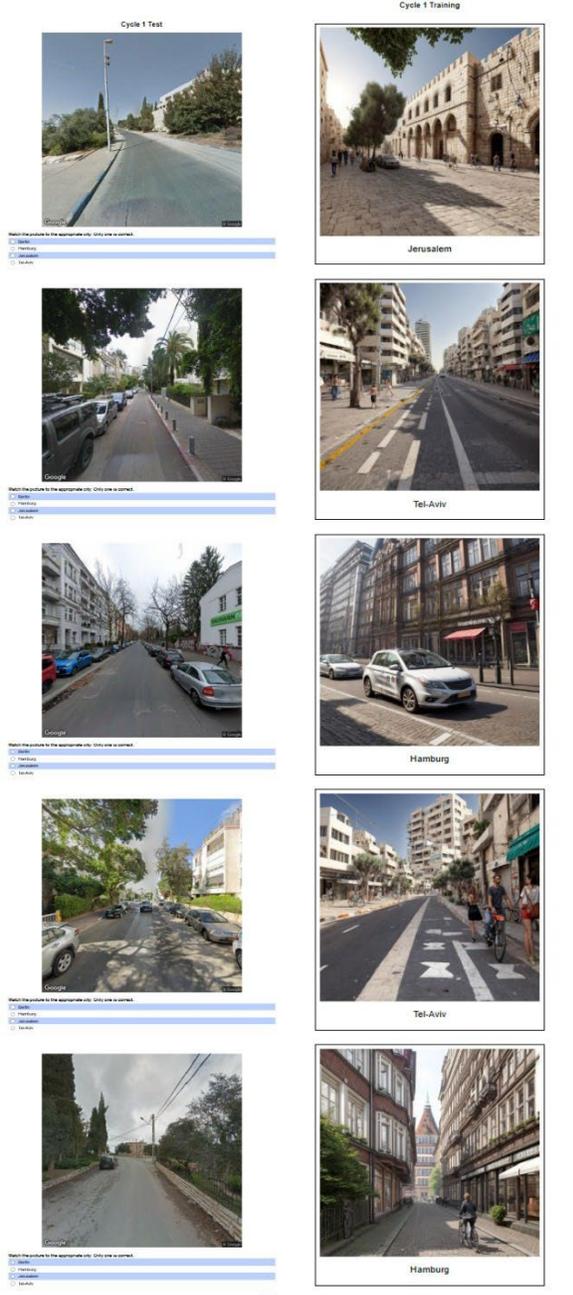


Figure 2. First cycle of testing (left) and training (right) in the experiment.

To evaluate our hypotheses, we analyzed participants' classification accuracy across the four cycles of the experiment. For each condition (personalized vs. random training), we calculated the mean accuracy achieved in every cycle. Cycle 1 reflects baseline performance before any training intervention, while cycles 2, 3, and 4 represent performance after the respective training phases. By plotting the mean accuracies in each round, we were able to show the learning curve of each group. To compare accuracy

between groups, we applied one-tailed and two-tailed independent (H1) and paired (H2) t-tests.

4.2. Image Selection and Synthetic Image Generation

For our questionnaire we selected two kinds of images: actual images of a city to test the participant and synthetic images to train the participant. For comparison with previous research in this field, we restricted ourselves also to the cities of Hamburg, Berlin, Jerusalem and Tel-Aviv (Walter, 2024).

To collect the real-world images for the classification task, we used the Google Street View Static API (Google, 2024). We generated a pool of randomized GPS coordinates within the urban boundaries of each of the four target cities. These coordinates were passed to the API to retrieve street-level images from a driver's perspective (Figure 3). The retrieved images were then subjected to a multi-stage validation process. First, we manually reviewed each image to ensure it depicted a plausible urban scene representative of the intended city, excluding images that were either of poor quality, located indoors, or not clearly identifiable as urban environments. Next, we validated the selected images using a pretrained geolocation estimation model (Müller-Budack et al., 2018) to verify that the visual content corresponded to the correct city. Only images correctly classified by the model were retained for use in the experiment. In total, we used 20 images in our experiment, 5 images for each of the four cycles of classification.



Figure 3. Google Street View image from Berlin.

For the training phase, we had to generate synthetic images designed to resemble realistic urban scenes. These were created using Leonardo.Ai (Jie et al., 2023; Leonardo AI, 2024) in Python. As a base, we provided real Street View images from the respective cities and

augmented them with a prompt formulated as follows: “Generate a realistic image on the streets of {City Name} as it would look like on Google Street View, driving along the street. Use the given image as reference.” To avoid unintended and distracting results, particularly close-up portraits of people, we included the negative prompt: “Persons or humans or people looking into the camera or being in focus!”. This adjustment was introduced based on iterative observations during development. In addition, the image prompt was weighted at 1.3, using photo realism settings.

To create the final counterfactual training images, we repurposed the initially generated synthetic images by instructing Leonardo.Ai to regenerate them to resemble one of the three alternative cities. This ensured visual continuity while offering contrasting geographic context (Figure 4). The label under each training image showed only the correct city name, not the type of error (e.g., "Berlin" instead of "Berlin-Hamburg"). These newly generated counterfactuals were once again validated by the geolocation estimation model (Müller-Budack et al., 2018), which confirmed whether the intended target city was visually represented in the image.



Hamburg

Figure 4. Synthetic image of Hamburg as counterfactual for Berlin.

Finally, all synthetic images were manually reviewed by a team of 4 researchers to filter out visual artifacts and unrealistic renderings that would compromise the photo-realistic quality of the training material. In total, we used 181 different synthetic images to reflect all possible combinations of counterfactuals in the final experiment.

5. Results

Our survey included 144 participants, divided into two groups: 71 in the random training condition and 73

in the personalized training condition. The majority of participants were from Germany (92). In addition, two participants were from Brazil and the USA, and one from Hungary, India and Indonesia. 45 participants choose not to disclose their location. The average time to complete the survey was 320 seconds (5 minutes and 20 seconds), with a maximum of 627 seconds and a minimum of 221 seconds. There were 30 unfinished runs, representing 20.83% of the total, with 15 (21.13%) in the random training condition and 15 (20.55%) in the personalized training condition.

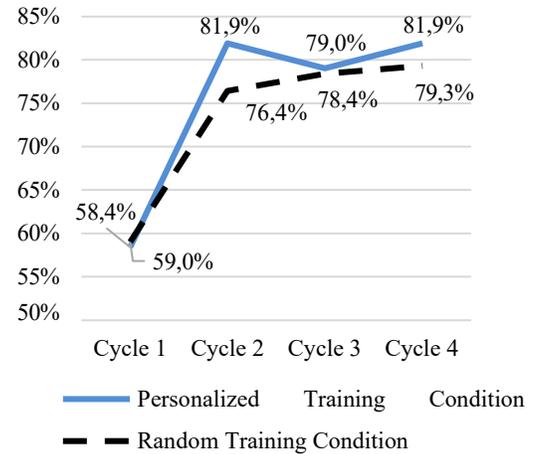


Figure 5. Plotted learning curves for both conditions in %.

Figure 5 displays the learning curves, where each point represents the mean accuracy achieved in the respective cycle. To test H1, we first compared the accuracy of both groups in cycle 1. A two-tailed independent t-test showed no statistically significant difference in initial accuracy between the personalized and random training conditions ($p = 0.87$, $t = -0.16$), confirming that the groups were comparable at the start. In cycles 2, 3, and 4, we observed that the learning curve for the personalized training condition consistently remained above the curve for the random training condition. While the absolute difference between the two conditions remained modest, the consistent directional improvement led us to conduct a one-tailed t-test on the mean of the accuracy of the cycles 2–4 combined. The result indicated a marginally significant advantage for the personalized training group ($p = 0.059$, $t = 1.56$). Given the directional nature of the hypothesis and the consistent trend across all post-training cycles, we interpret this as sufficient evidence to support H1: Participants receiving personalized counterfactual synthetic images showed greater improvements in classification accuracy than those receiving randomly assigned synthetic images.

To evaluate H2, we compared the accuracy between cycle 1 (pre-training) and cycle 4 (post-training) within

each condition separately. In both the personalized and random training groups, classification accuracy significantly improved over time. One-tailed paired t-tests confirmed that these improvements were statistically significant for both groups ($p < 0.001$, $t = -5.56$ for the random, $p < 0.001$, $t = -6.62$ for the personalized training condition). These findings support H2, indicating that learning is possible with synthetic training images—regardless of whether they were personalized.

6. Discussion

6.1. Theoretical Contributions

This research on our suggestion of an intelligent learning system offers several theoretical contributions to the literature on learning systems, personalized training, and synthetic data in human learning. Anchored in our GeoGuessr-inspired classification task, we extend existing knowledge by demonstrating how personalization and counterfactual training data can be productively applied to enhance classification skills.

First, we contribute to theories of adaptive learning and personalized training by providing empirical evidence that personalized feedback, tailored to a novice's previous errors, improves learning outcomes. While prior research has established the general value of adaptive systems, our study highlights the specific value of counterfactual training data in guiding novices toward more accurate classifications. This adds nuance to existing theories by illustrating how accuracy-based adaptation can be implemented through visual training. Second, we extend emerging discussions on counterfactual training data by operationalizing them as dynamic training in a real-time learning environment. Although counterfactuals have gained traction in the explainable AI literature, their role in improving human learning remains underexplored. Our study shows that presenting novices with counterfactual synthetic images, i.e., examples that contrast directly with their classification errors, can lead to measurable improvements in learning. This supports the idea that contrastive feedback strengthens classification skills and provides a more precise form of error correction.

Third, we contribute to the growing work on synthetic data in human learning systems. Much of the prior research has focused on synthetic data for algorithmic training or data privacy. In contrast, we shift attention toward its teaching potential. Our findings demonstrate that synthetically generated images are not only realistic and valid as training material, but also effective in improving human learners' accuracy in visual classification tasks. This opens new theoretical

avenues for integrating generative AI with personalized learning design.

Lastly, we position our study at the intersection of game-inspired learning environments and intelligent training systems, a space often separated in the IS literature. While game-based learning is typically associated with motivation and engagement, our findings suggest it can also be a platform for structured and adaptive skill development. By grounding our task in the familiar challenge of GeoGuessr and classifying city landscapes based on visual cues, we show how theories of adaptive feedback and personalization can extend into engaging, visually rich training contexts.

6.2 Practical Contributions

This study offers several practical contributions for the design and implementation of intelligent learning systems. First, our findings demonstrate that synthetic images can be effectively used to train human learners in complex visual classification tasks. This has important implications for training in resource-constrained environments where real-world data may be scarce, sensitive, or costly to curate. By showing that synthetic training material can lead to measurable learning gains, we offer a practical alternative for developers of learning platforms in education, geography, and visual reasoning.

Second, we present a concrete approach to real-time personalization based on learner behavior. The system dynamically tracks user errors and adapts training content accordingly by selecting counterfactual images that correct specific misclassifications. This fine-grained, error-based personalization strategy is implementable in practice and can be easily integrated into learning platforms, offering a scalable method for tailoring instruction without requiring human teachers.

Third, by embedding this training approach into a classification task, our study shows how educational tools can benefit from playful, engaging environments. Game-based settings not only make learning more accessible and enjoyable for novices but also enable natural data collection and feedback loops that support personalization. Developers of educational games and tutoring systems can build on our findings to enhance both learner engagement and outcome effectiveness through adaptive feedback.

Finally, our work informs the design of human-centered AI systems. It illustrates how synthetically generated content can serve not merely as a backend tool for automation, but as a direct and meaningful interface for human learning. This contributes to current practice in AI-supported education by offering a roadmap for how generative models can be harnessed to provide personalized, actionable training content—while still

keeping the human in control of the learning process. Hence, our work contributes to the development of scalable, personalized, and effective self-learning through intelligent learning systems, thereby potentially helping to mitigate the shortage of skilled workers in industry and manufacturing.

7. Conclusion, Limitations and Further Research

This study explored the impact of personalized counterfactual synthetic images of an intelligent learning system in a GeoGuessr-inspired classification task. By simulating a visual classification task with geographically grounded content, we demonstrated that personalized training significantly improves novices' classification accuracy compared to randomly assigned training. These results highlight the power of personalization and the effectiveness of synthetic data in game-like learning contexts.

Beyond confirming the utility of adaptive training strategies, our findings show that even without real-world images, synthetically generated content can support meaningful skill development. This opens up promising avenues for scalable, automated, and personalized learning environments that combine intelligent feedback with playful, engaging formats inspired by games.

While the results are encouraging, several limitations must be acknowledged. First, the study focused on a single type of classification task involving urban imagery from four cities, which may limit generalizability to other domains. Second, participants were primarily novices recruited from online platforms. It remains unclear how well the results would translate to more experienced users, if our suggested intelligent learning system can also improve the results of already knowledgeable users. Third, we have also not measured learning-specific factors of our participants. Participants with different styles of learning could react differently to our intelligent learning system and may require a different way of personalization. Fourth, although our synthetic images were carefully validated, we did not compare them directly with real training images. Future studies could benchmark synthetic and real data to better understand their relative training value.

Future work should explore the application of counterfactual training data in a broader range of domains, in visual contexts such as medical image interpretation, satellite imagery analysis, or emergency response planning, but also in other domains without visual training data. Moreover, integrating full-fledged game mechanics (e.g., such as points, levels, leaderboards, and time pressure) could further enhance engagement. Although the task was inspired by

GeoGuessr, the experiment was conducted in a survey format without competitive elements or gamified interaction to ensure experimental control. Therefore, we may have not fully captured the motivational dynamics present in GeoGuessr. In addition, future research should examine the cognitive mechanisms underlying counterfactual learning, including how visual attention and memory processes are shaped by exposure to corrective synthetic examples. Investigating how different user profiles respond to personalization, including also different learning styles of users, could help inform more inclusive and effective intelligent learning systems tailored to diverse learners.

Future research on generative AI in learning systems must address ethical and cognitive risks from synthetic data, as such systems often reproduce or amplify societal biases (Wei et al., 2025). These risks are especially concerning in early education, where children may internalize stereotypes (Retelsdorf et al., 2015). Research should examine how algorithmic bias in counterfactual training data transmits to learners, since systematic under- or misrepresentation can lead to biased mental models.

Moreover, recent findings show that people may dismiss AI-generated information when it conflicts with their own biases (e.g., gender bias), underscoring the complex interplay between human and algorithmic bias (Zipperling et al., 2025). Understanding this dynamic is key to building fair, transparent, and ethical learning systems. Addressing these risks becomes increasingly important as generative AI outputs grow more realistic, risks of bias transmission and altered user perceptions increase; for instance, seeing AI as conscious can influence both AI interactions and human-human behavior (Guingrich and Graziano, 2024).

As models for text and image generation become more lifelike, their outputs also risk being overly persuasive, raising concerns about undue influence on learners' reasoning (Matz et al., 2024). Further, empirical studies show overreliance on AI can impair critical thinking, performance, and motivation while weakening memory and creativity (Abbas et al., 2024; Fan et al., 2025).

Addressing these challenges will not only ensure that intelligent learning systems can be responsibly deployed, but also promise to enhance the theoretical and practical effectiveness of personalized training and further develop our proposed intelligent learning system into a tool that can be used across multiple domains.

References

Abbas, M., Jam, F. A., and Khan, T. I. (2024). "Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students."

- International Journal of Educational Technology in Higher Education*, 21(1), 10.
- Abufadda, M., and Mansour, K. (2021). "A Survey of Synthetic Data Generation for Machine Learning." *Proceedings of the International Arab Conference on Information Technology (ACIT)*, 1–7.
- Adamkiewicz, K., Woźniak, P. W., Dominiak, J., Romanowski, A., Karolus, J., and Frolov, S. (2025). "PromptMap: An Alternative Interaction Style for AI-Based Image Generation." *Proceedings of the International Conference on Intelligent User Interfaces*, 1162–1176.
- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., and Martinez-Maldonado, R. (2024). "Human-centred learning analytics and AI in education: A systematic literature review." *Computers and Education: Artificial Intelligence*, 6, 100215.
- Allmendinger, S., Hemmer, P., Queisner, M., Sauer, I., Müller, L., Jakubik, J., Vössing, M., and Kühl, N. (2024). Navigating the Synthetic Realm: Harnessing Diffusion-Based Models for Laparoscopic Text-to-Image Generation. In A. Shaban-Nejad, M. Michalowski, and S. Bianco (Eds.), *AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health* (pp. 31–46). Springer Nature Switzerland.
- Argote, L. (2015). Knowledge transfer and organizational learning. In *The Wiley Blackwell handbook of the psychology of training, development, and performance improvement* (pp. 154–170). Wiley Blackwell.
- Baier, L., Jöhren, F., and Seebacher, S. (2019). "Challenges in the Deployment and Operation of Machine Learning in Practice." *Proceedings of the ECIS*, 163.
- Bloom, B. S. (1984). "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *Educational Researcher*, 13(6), 4.
- Brown, J. S., Collins, A., and Duguid, P. (1989). "Situated Cognition and the Culture of Learning." *Educational Researcher*, 18(1), 32–42.
- Bulmer, M., Coote, L., and University of Queensland. (2022). "The Role of Synthetic Data in Teaching and Learning Statistics." *Proceedings of the International Conference on Teaching Statistics (ICTS)*.
- Carioli, P., Czarnitzki, D., and Fernández, G. P. (2024). *Evidence on the Adoption of Artificial Intelligence: The Role of Skills Shortage* (SSRN Scholarly Paper No. 4774544). Social Science Research Network.
- Clarizia, F., De Santo, M., Lombardi, M., and Santaniello, D. (2021). "E-Learning and Industry 4.0: A Chatbot for Training Employees." In X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi (Eds.), *Proceedings of Fifth International Congress on Information and Communication Technology* (pp. 445–453). Springer.
- Dai, X., Keane, M. T., Shalloo, L., Ruelle, E., and Byrne, R. M. J. (2022). "Counterfactual Explanations for Prediction and Diagnosis in XAI." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 215–226.
- Derouin, R. E., Fritzsche, B. A., and Salas, E. (2005). "E-Learning in Organizations." *Journal of Management*, 31(6), 920–940.
- Diebel, C., Goutier, M., Adam, M., and Benlian, A. (2025). "The Price of AI Assistance: The Undermining Effect of AI-Generated Code on Developers' Procedural Knowledge." *Proceedings of the ECIS*.
- Engbom, N. (2018). "Firm and Worker Dynamics in an Aging Labor Market." 2018 Meeting Papers, Article 1009.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., and Gašević, D. (2025). "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance." *British Journal of Educational Technology*, 56(2), 489–530.
- Farmer, T. S., and Ho, C.-J. (2025). "Examining Interactions Between User Characteristics and Explanation Modalities on Inducing Complementarity." *Proceedings of the ECHI Conference on Human Factors in Computing Systems*, 1–7.
- Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2024). "Generative AI." *Business & Information Systems Engineering*, 66(1), 111–126.
- Google. (2024). *Street View Static API overview*. Google for Developers. Retrieved June 14, 2025, from <https://developers.google.com/maps/documentation/streetview/overview>
- Goutier, M. (2024). *Intelligent Learning System For Image Classification*. OSF Registries. Retrieved June 13, 2025, from <https://osf.io/tq9j4>
- Goyal, Y., Wu, Z., Ernst, J., Batra, D., Parikh, D., and Lee, S. (2019). "Counterfactual Visual Explanations." *Proceedings of the International Conference on Machine Learning*, 2376–2384.
- Grenander, M., Belfer, R., Kochmar, E., Serban, I. V., St-Hilaire, F., and Cheung, J. C. K. (2021). *Deep Discourse Analysis for Generating Personalized Feedback in Intelligent Tutor Systems* (arXiv:2103.07785). arXiv.
- Gross, S., Mokbel, B., Hammer, B., and Pinkwart, N. (2015). "Learning Feedback in Intelligent Tutoring Systems. Report of the FIT Project, Conducted from December 2011 to March 2015." *KI - Künstliche Intelligenz*, 29(4).
- Guingrich, R. E., and Graziano, M. S. A. (2024). "Ascribing consciousness to artificial intelligence: Human-AI interaction and its carry-over effects on human-human interaction." *Frontiers in Psychology*, 15.
- Ijaz, K., Wang, Y., Ahmadpour, N., and Calvo, R. A. (2019). "Immersive VR Exergames for Health and Wellbeing." *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–4.
- Jie, P., Shan, X., and Chung, J. (2023). "A Comparative Analysis Between Leonardo.Ai and Meshy as AI Texture Generation Tools." *International Journal of Advanced Culture Technology*, 11(4), 333–339.
- Kan, Z. (Ed.). (2024). *Gagne's Theory of Instruction*. Springer Nature Singapore.
- Kang, L., and Jackson, S. (2021). "Tech-Art-Theory: Improvisational Methods for HCI Learning and Teaching." *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), 82:1-82:25.
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., and Pineau, J. (2020). "Automated Personalized Feedback Improves Learning Gains in An Intelligent Tutoring System." *Artificial Intelligence in Education*, 12164, 140–146.

- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., and Pineau, J. (2022). "Automated Data-Driven Generation of Personalized Pedagogical Interventions in Intelligent Tutoring Systems." *International Journal of Artificial Intelligence in Education*, 32(2), 323–349.
- Kühl, N., Goutier, M., Baier, L., Wolff, C., and Martin, D. (2022). "Human vs. supervised machine learning: Who learns patterns faster?" *Cognitive Systems Research*, 76, 78–92.
- Kwon, K., and Jang, S. (2022). "There is no good war for talent: A critical review of the literature on talent management." *Employee Relations: The International Journal*, 44(1), 94–120.
- Leonardo AI. (2024). *AI Image Generator—Create Art, Images & Video | Leonardo AI. AI Image Generator - Create Art, Images & Video | Leonardo AI*. Retrieved June 14, 2025, from <https://leonardo.ai/>
- Liu, R., Wei, J., Liu, F., Si, C., Zhang, Y., Rao, J., Zheng, S., Peng, D., Yang, D., Zhou, D., and Dai, A. M. (2024). "Best Practices and Lessons Learned on Synthetic Data." *Proceedings of the Conference on Language Modeling*.
- Luo, J., Zheng, C., Yin, J., and Teo, H. H. (2025). "Design and assessment of AI-based learning tools in higher education: A systematic review." *International Journal of Educational Technology in Higher Education*, 22(1), 42.
- Madgavkar, A., and White, O. (2025). *How to address US labor shortages | McKinsey*. Retrieved August 30, 2025, from <https://www.mckinsey.com/industries/public-sector/our-insights/empowering-the-us-workforce>
- Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., and Cerf, M. (2024). "The potential of generative AI for personalized persuasion at scale." *Scientific Reports*, 14(1), 4692.
- Metcalfe, J. (2017). "Learning from Errors." *Annual Review of Psychology*, 68(1), 465–489.
- Miles, T., Aragon, C., and Evans, S. (2024). *Taking a Human-Centered Approach to Games-Based Learning to Explore Data Science Ethics* (SSRN Scholarly Paper No. 4947161). Social Science Research Network.
- Müller-Budack, E., Pustu-Iren, K., and Ewerth, R. (2018). "Geolocation Estimation of Photos Using a Hierarchical Model and Scene Classification." *Proceedings of the European Conference on Computer Vision*, 11216, 575–592.
- Renganayagalu, S. K., Mallam, S., Nazir, S., Ernstsén, J., and Haavardtun, P. (2019). "Impact of Simulation Fidelity on Student Self-efficacy and Perceived Skill Development in Maritime Training." *TransNav, International Journal on Marine Navigation and Safety Od Sea Transportation*, 13(3), 663–669.
- Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "'Michael can't read!' Teachers' gender stereotypes and boys' reading self-concept." *Journal of Educational Psychology*, 107(1), 186–194.
- Sanzogni, L., Guzman, G., and Busch, P. (2017). "Artificial intelligence and knowledge management: Questioning the tacit dimension." *Prometheus*, 35(1).
- Scarlatos, A., Smith, D., Woodhead, S., and Lan, A. (2024). "Improving the Validity of Automatically Generated Feedback via Reinforcement Learning." *Proceedings of the International Conference on Artificial Intelligence in Education*, 280–294.
- Serban, I. V., Gupta, V., Kochmar, E., Vu, D. D., Belfer, R., Pineau, J., Courville, A., Charlin, L., and Bengio, Y. (2020). "A Large-Scale, Open-Domain, Mixed-Interface Dialogue-Based ITS for STEM." *Artificial Intelligence in Education*, 12164, 387–392.
- Spitzer, P., Kühl, N., and Goutier, M. (2022). "Training Novices: The Role of Human-AI Collaboration and Knowledge Transfer." *Proceedings of the International Conference on Machine Learning, Workshop on Human-Machine Collaboration and Teaming*.
- Spitzer, P., Kühl, N., Goutier, M., Kaschura, M., and Satzger, G. (2024). "Transferring Domain Knowledge with (X)AI-Based Learning Systems." *Proceedings of the ECIS*.
- Spitzer, P., Kühl, N., Heinz, D., and Satzger, G. (2023). "ML-Based Teaching Systems: A Conceptual Framework." *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2), 348:1-348:25.
- Szymanski, M., Millecamp, M., and Verbert, K. (2021). "Visual, textual or hybrid: The effect of user expertise on different explanations." *Proceedings of the International Conference on Intelligent User Interfaces*, 109–119.
- Troussas, C., Papakostas, C., Krouska, A., Mylonas, P., and Sgouropoulou, C. (2023). "Personalized Feedback Enhanced by Natural Language Processing in Intelligent Tutoring Systems." *Proceedings of the International Conference on Intelligent Tutoring Systems, Augmented Intelligence and Intelligent Tutoring Systems*, 667–677.
- Vandenhende, S., Mahajan, D., Radenovic, F., and Ghadiyaram, D. (2022). "Making Heads or Tails: Towards Semantically Consistent Visual Counterfactuals." *Proceedings of the European Conference on Computer Vision*, 261–279.
- Walter, M. C. (2024). "Follow Me, Everything Is Alright (or Not): The Impact of Explanations on Appropriate Reliance on Artificial Intelligence." *Proceedings of the HICSS*.
- Walter, M. C., Broder, H. R., and Förster, M. (2023). "Boosting Benefits, Offsetting Obstacles – the Impact of Explanations on AI Users' Task Performance." *Proceedings of Wirtschaftsinformatik*.
- Wei, X., Kumar, N., and Zhang, H. (2025). "Addressing bias in generative AI: Challenges and research opportunities in information management." *Information & Management*, 62(2), 104103.
- Wong, W.-T., and Huang, N. N.-T. (2011). "The Effects of E-Learning System Service Quality and Users' Acceptance on Organizational Learning." *International Journal of Business and Information*, 6(2), 205–225.
- Zhu, X., Singla, A., Zilles, S., and Rafferty, A. N. (2018). *An Overview of Machine Teaching* (arXiv:1801.05927). arXiv.
- Zipperling, D., Deck, L., Lanzl, J., and Kühl, N. (2025). "It's only fair when I think it's fair: How Gender Bias Alignment Undermines Distributive Fairness in Human-AI Collaboration." *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*.