University of Tartu
Faculty of Social Sciences
Institute of Education
Curriculum of Educational Technology

Chukwuemeka Oji Kanu-Oji

Exploring the Impact of AI Chatbot Integration on Learning Outcomes in Environmental Health and Safety Stormwater Training

MA Thesis

Supervisor: Yaroslav Opanasenko

Tartu, 2025

Abstract

This study examined the impact of integrating StormWise, an AI chatbot, into Environmental Health and Safety (EHS) training on stormwater management. The research aimed to determine whether AI chatbot assisted training influenced optimal learning outcomes compared to conventional training methods. A quasi-experimental design was used, involving 100 participants divided into AI-chatbot and conventional cohorts, each assessed before and after the training. Findings showed no significant difference in post-assessment scores between the two groups. However, qualitative data indicated that chatbot made the training more engaging and interactive. Some technical challenges were also observed. The study highlights the potential of AI chatbots in professional development and how they can be optimized to support and enhance learning outcomes.

Keywords: AI Chatbot, EHS Training, Professional Development, Personalized Learning, Learning Outcomes.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Stormwater pollution is a major environmental challenge, considering that a significant amount of polluted water globally has been linked to stormwater runoff. According to a UN report, at least 80% of outfalls from industrial and urban areas globally are discharged to surrounding waterbodies, thereby releasing high pollutant loads into the environment (UN, 2024). In addition to being a major threat to sustainable urban and rural development (Fereshtehpour & Najafi, 2025; Moazzem et al., 2024), stormwater runoff also contributes to severe environmental degradation (Fentiman & Zabbey, 2015; Yuhuan et al., 2023), loss of biodiversity (Sathvara et al., 2024) and public health crises (McGonagle et al., 2024).

Effective stormwater management plays a critical role in ensuring regulatory compliance and adequate environmental protection. In the US, the National Pollutant Discharge Elimination System (NPDES) permit program helps in addressing water pollution challenges through the regulation of point-source pollutants discharged to the aquatic ecosystem (EPA, 2008). Training plays a crucial role in the successful implementation of the requirements of this program. Under the NPDES regulatory framework, employers are required to provide training to their employees on various aspects of stormwater management, such as stormwater Best Management Practices (BMPs), Storm Water Pollution Prevention Plan (SWPPP), pollutant sources, emergency response to spills, and safety hazards (CFR,1983).

The main goal of such training is to ensure that workers have a good understanding of BMPs, as well as spill response and prevention plans. This approach contributes significantly to the prevention of stormwater runoff contamination, and in the event of an emergency spill, appropriate response can be taken. Training enhances the ability of the workforce to recognize, understand, and manage hazards, which helps mitigate associated negative outcomes. By embedding training as a core component of Environmental Health and Safety (EHS) programs, organizations can enhance their capacity to manage evolving environmental and workplace challenges (Schroth et al., 2023).

However, EHS training, typically delivered through classroom sessions, pre-recorded videos, or e-learning, has been widely criticized due to its numerous limitations. They have been associated with lack of engagement, limited learners' agency, tutor-centered learning, low motivation, lack of adequate learning support, and excessive cognitive load, which negatively impact learning outcomes (Debs et al., 2019; Mesuwini & Mokoena, 2024; Suravi, 22024) This underscores the need for a more learner-centered, interactive, and engaging training method that meets the needs of adult learners and enhance learning outcomes.

In an era of rapid and ever-evolving technological innovation, it is important to ensure that organizations adapt the methods used in equipping their workforce. AI chatbots have emerged as an innovative tool that offers huge potential to transform training. Despite their increasing popularity and application in various fields, few studies have been carried out to determine the effectiveness of AI chatbots in EHS training, particularly for stormwater management training. Addressing this research gap is important, considering the need to discover effective EHS training options that can be useful in supporting environmental protection programs.

This study investigated the effectiveness of AI chatbots in enhancing stormwater management training by comparing learning outcomes and trainees feedback between two groups, one receiving chatbot-supported training and the other subjected to conventional training. Through the review of training assessment and qualitative feedback data from both groups, the research provides a good understanding of the potential benefits and limitations of incorporating AI chatbots into stormwater management training programs.

1.1 Theoretical Overview

1.1.1 EHS Training Framework

Various laws at regional and national levels mandate Environmental Health and Safety (EHS) training as a legal requirement. These regulatory frameworks ensure that both employers and employees remain well-informed and equipped to maintain a safe workplace and ensure adequate protection of the environment.

EHS training refers to instruction and practice for acquiring skills and knowledge of rules, concepts, or attitudes necessary to function effectively in specified task situations. It plays a key role in ensuring a safe workplace and the reduction of environmental degradation associated with organizational activities.

EHS comprises three key components: environmental management, occupational health and workplace safety. Figure 1 shows how EHS training, which sits at the core, plays a key role in ensuring proper dissemination of information and knowledge across these three EHS domains, contributing to promoting employee wellbeing, workplace safety and environmental protection. Occupational health is concerned with the protection of workers from hazards that may lead to injuries, diseases, illnesses, and long or short-term health challenges. It promotes proper mental, physical, and social wellness of employees. Workplace safety on the other hand encompasses measures that are adopted to ensure adequate protection of employees within their work environment. This will include, but not be limited to, hazard identification, risk assessment, hazard control measure development, and emergency management. Finally, environmental management deals with identifying sustainable practices that reduce the likelihood of environmental degradation and compliance with relevant environmental regulations.

Figure 1. EHS Training Framework

This framework shows that training is by far the most important aspect of the EHS intervention program. EHS training serves as a means of transferring relevant information to employees to reinforce desirable action, encouraging workplace and environmental stewardship (Burke et al., 2011). It therefore serves as a foundation for organizations to achieve regulatory compliance, increase operational efficiency, reduce risk, and promote environmental sustainability.

1.1.2 Critical issues associated with EHS training and Future Direction

EHS training is an important aspect of many workplace EHS intervention programs. However, the effectiveness of some of the training methods has been questioned. To ensure that workplace EHS training meets its objectives, it is essential to re-evaluate the training method adopted to ensure it meets the needs of the targeted learners. This highlights the need for more innovative and adaptable training approaches. This section explores the significance of common EHS training methods

1.1.2.1 Traditional EHS Education Approaches

Recent studies have highlighted the limitations of conventional training methods. Burke et al. (2006) found that the methodology adopted in EHS training has a significant impact on learning outcomes. Their research demonstrates that, unlike traditional lecture-styled, passive delivery methods, interactive and engaging approaches tend to deliver better learning outcomes. However, some researchers have argued that training course completion does not necessarily guarantee that the trainee will adopt safe behaviors or engage in environmentally responsible practices, which raises questions about the role of EHS education in reducing workplace accidents or its place in the environmental protection narrative (Caffaro et al., 2018; Suratman et al., 2016).

Another important aspect of education at any level is the ability to self-regulate and self-direct one's learning. Self-regulated learning (SRL) refers to the ability of learners to define learning objectives, plan, execute, regulate, and periodically evaluate the effectiveness of the learning process toward achieving the desired learning goals (Zimmerman, 2002). One way of ensuring that learners reflect on and control their learning is by making options available to them. This needed agency is completely diminished in the traditional learning paradigm. In addition to this, a lack of motivation that is inherent in these outdated teaching methods hampers optimal learning.

1.1.2.2 Technology Enhanced Learning and EHS Education

Technology offers opportunities to bridge existing gaps in workplace EHS training, however, acceptance remains polarized. Digital tools such as electronic learning (e-learning) have been associated with numerous learning advantages, including increased flexibility, enhanced personalization, improved accessibility, and cost-effective scalability (Barati Jozan et al., 2023). Yet, despite their popularity, studies have also reported the reluctance of some workers to receive training through online delivery (Wilkins, 2011). This has been attributed to various factors such as lack of engagement, excessive cognitive load, lack of personalization, and lack of readily available supporting technological infrastructure (Abbasnejad et al., 2024). This means that poorly designed e-learning will equally be ineffective in delivering better learning outcomes.

Addressing these barriers will be crucial for technology-driven EHS training to produce optimal results. Some earlier studies also point out that even the best-designed intelligent tutoring system has failed in some instances to enhance learning compared to results obtained from adopting traditional learning approaches (Johnson et al., 2009). On the other hand, Evans & Moore (2013), found that the use of technology not only improved collaboration but also enhanced learning. Another study reports that the application of digital technologies in education enhances learning by improving learners' perception, increasing motivation, and making learning more interesting (Irkha et al., 2024)

1.1.2.3 Emerging Innovative Training Approaches

Given the complex nature of the modern work environment and the diverse nature of hazards workers and the surrounding environment are exposed to, it is evident that no single approach can comprehensively addresses all training needs. Bęś and Strzałkowski (2024) argues that optimal results can be achieved by combining multiple training methods, tailored to specific workplace contexts and learners' needs. As technology continues to advance, it is necessary to integrate this new technological innovation into the delivery of EHS training.

One emerging yet underexplored approach is the use of AI-chatbots such as custom GPT as EHS training agents. These tools offer personalized, real-time learning experiences, and have been shown to have the potential to enhance trainees' engagement and knowledge retention (Sharma, 2023). Evaluating their effectiveness in delivering EHS training provides valuable insights into designing more impactful EHS intervention programs, ultimately contributing to the reduction of workplace injuries and environmental degradation. With regards to stormwater management, it can significantly enhance workers' knowledge of BMPs, which helps reduce contamination of stormwater runoff, and hence water pollution reduction. Through this approach, effective EHS training intervention contributes to environmental protection.

1.1.3 Artificial Intelligence (AI) in Education

1.1.3.1 Evolution of AI

Scientists have long sought to integrate technology into education, leading to the emergence of concepts like Technology-Enhanced Learning (TEL) and Intelligent Tutoring Systems (ITS) in the late 20th century (Kirkwood & Price, 2014; Van Seters et al., 2012). With rapid technological advancements, the 21st century has witnessed a transformative shift in teaching and learning methodologies. One of such innovations is AI, which refers to a branch of computer science focused on developing systems that simulate human intelligence and behaviors, enabling machines to process information, reason, and make decisions. (Duan et al., 2019).

AI-driven technologies, such as adaptive learning platforms, intelligent tutoring systems, automated grading tools, and AI-powered chatbots, have transformed and redefined the role of both learners and educators (Kamalov et al., 2023). These advancements can allow training facilitators to focus more on guiding learners, fostering critical thinking, and providing personalized support, ultimately enhancing learning. As for the learners, it presents opportunities for collaboration, critical thinking, self-regulation, self-directed learning, learners' agency and fosters knowledge creation (Otermans et al., 2024).

In general, AI has demonstrated remarkable abilities to disrupt traditional teaching and learning practices that have been the mainstay of academic work (Rahman & Watanobe, 2023; Pal, 2023). As AI continues to evolve, it is expected that its role in education is likely to expand even further. AI technology appears to hold huge potential in helping tackle learning challenges. Studies like this will contribute to understanding where AI best fits into in the learning process.

1.1.3.2 Generative AI (GenAI) and its Role in Education

Among the major developments in AI is Generative Artificial Intelligence (GenAI), which has the capability of learning and creating content in various formats such as texts, pictures, and videos A typical example of GenAI is Large Language Models (LLM), which can generate text typically based on a prompt, making it possible for them to execute a variety of complex language processing tasks. LLM is known to be trained with a large data set and can subsequently recognize, predict translate, or generate text (Jeon & Lee, 2023). Given their ability to generate contextually relevant responses, all LLMs can therefore be classified as GenAI, further underscoring their relevance in shaping the future of AI-driven teaching and learning. One key feature they have is their remarkable ability to simulate having a conversation with users (Boubker, 2024). This unique ability of engaging users in open-ended dialogue-based interactions gave rise to chatbots of which ChatGPT is the most popular and widely used. It makes use of a pre-trained transformer model, which enables it to respond effectively to users' input through learning context and meaning and by tracking relationships in sequential data (Adeshola & Adepoju, 2024).

Considering their remarkable abilities, GenAI can therefore play a huge role in the area of professional development by assuming the role of a tutor and learning partner (Hwang et al., 2020). Studies. They are well adapted to the provision of personalized learning and guided support based on learners’ preferences, learning styles, or individual characteristics (Sajja et al., 2024).

1.1.3.3 Critical Issues Associated with GenAI

One major challenge associated with GenAI tools such as ChatGPT in EHS training is the issue of hallucination which involves the generation of plausible yet incorrect or misleading information (Flenady & Sparrow, 2025). This limits their usefulness in delivering effective EHS training due to unreliable outputs. This ranges from contradictions and factual inconsistencies to outlandish responses (Elbanna & Armstrong, 2024). One study found that GenAI showed performed fairly in providing support for short-answer learning with outputs fraught with inaccuracies, and when used for supporting complex learning, it performed poorly (Alimardani & Jane, 2025). However, another study demonstrated that chatbots improves performance by supporting collaboration (Amiot et al., 2025).

Unlike human tutors, these models lack the cognitive ability to understand the contents generated. Considering that some of the users may have little to no experience with the subject matter, when used for EHS training, where wrong outputs could have serious implications. It because a major problem because learners are now charged with the responsibility of determining the accuracy of outputs from these chatbot, which can have a negative impact on learning. Measures should be put in place to ensure the reliability of the outputs from the chatbots.

Another issue with GenAI is the problem of bias. Considering that their output is a function of the dataset used in training them, if the data is skewed or not representative, the output will follow the same pattern, leading to suboptimal learning outcomes. Integrating AI chatbots into training can serve as a means of fostering critical thinking by encouraging learners to critique outputs from these chatbots and ensure that information is verified before accepting them as valid. This also opens up opportunities for collaboration.

A key prominent feature of AI is its ability to individualize learning, a form of personalization. This creates an opportunity for learners with varying abilities to engage with AI systems like

ChatGPT for instance, differently. While individualization can promote learners' agency, the unintended consequence is the lack of social interaction, which drives the learning process. This is based on Social Learning Theory that emphasizes the significance of social factors in the learning process and how collaboration with peers and student-tutor interaction influence performativity and ultimately learning outcomes (Li et al., 2023). The lack of such social interaction when systems are solely dependent on AI systems may lead to suboptimal learning outcomes, especially when collaborative dimensions are overlooked. However, some studies argue that individualized learning environments produced positive learning outcomes (Montebello, 2021). Individualized learning that works is not a one-size-fits-all design, but should endeavor to understand contexts, guide learners through inquiry, generate conversations, foster knowledge creation and sharing with peers, and transfer acquired knowledge into solving real-world problems.

However, despite all these challenges, AI has the potential to transform the teaching and learning process. Further research aimed at understanding where and when GenAI can be utilized, and their inherent limitations, will play a huge role in helping address these challenges, which can assist AI technology to realize its potential in the long term.

1.1.3.4 Custom GPT as an EHS Training Solution

Custom GPT models have been developed to address the challenges associated with integrating AI in education. Unlike general-purpose models, Custom GPT can be trained on specific trusted datasets to ensure more reliable and consistent outputs (Pu et al., 2025). By providing thorough instructions, GPT models are trained and guided to perform according to specified instructional objectives. These upgrades have made it possible for chatbots to be useful in several ways, particularly as conversational agents. Unlike general-purpose ChatGPT, custom GPTs can be trained to engage the user in a meaningful conversation in a way that facilitates knowledge co-creation rather than providing answers to every prompt (Lo, 2024). This enhancement in conversational AI contributes to the growing success of chatbots like ChatGPT, making them more valuable for both educational and professional applications.

Ideally, a customized GPT should deliver learning content tailored to a well-defined group with similar interests. To be effective, the content must be engaging, relevant, and adaptive to individual differences and learner capabilities (Ejjami, 2024; Shemshack & Spector, 2020). Consequently, the effectiveness of a custom GPT in facilitating positive learning outcomes is directly influenced by its ability to integrate these core features of personalized learning environments, which largely depends on the quality of the customization process. Through research and continuous development, integrating AI tools into EHS training programs could play a huge role in enhancing learning outcomes.

1.2 Research Objective

The objective of this study is to determine the extent to which integrating chatbots into stormwater EHS training can enhance learning outcomes. Conventional EHS training is known to have several limitations such as lack of agency, reduced learner engagement, and lack of personalization, which often leads to suboptimal learning outcomes. To address these challenges, this study explores the impact of AI-chatbot integration on both learning outcomes and learner perceptions within stormwater management training.

The research questions (RQ) for this study are:

RQ1: To what extent does integrating AI-chatbot into stormwater EHS training enhance learning outcomes?

RQ2: How does the qualitative feedback from participants differ between those trained with AI-chatbot assistance and those trained using conventional training methods?

RQ3: To what extent does integrating AI-chatbot into stormwater EHS training improve assessment performance compared to conventional training approaches?

The findings of this study are expected to inform training policy decisions and guide the design of future EHS training programs. Ultimately, this research will contribute to the growing body of knowledge on the educational application of GenAI beyond traditional classroom settings.

Chapter 2
Methodology

2.1 Study design

This study was conducted using a quasi-experimental design approach, which is based on establishing a causal effect of a treatment on an outcome. This approach mimics experimental conditions by using non-experimental variations in the independent variable, where some respondents are exposed to an intervention while others are not, and allocation to groups is done randomly (Gopalan et al., 2020). This research design approach has been used in various empirical studies to evaluate the association between an intervention and an outcome (Cham et al., 2024; Schweizer et al., 2016).

The study compared the effectiveness and learning outcomes of two different EHS training approaches, one involved integrating an AI chatbot into the training (AI training cohort), while the other maintained a conventional training approach (Conventional training cohort) without any interaction with the AI chatbot.

2.2 Sample

Considering the impact of sample size and statistical power on the validity of the study, the study was made up of one hundred participants (n=100), comprising fifty EHS Professionals (n=50) and fifty non-EHS professionals (n=50). Random assignment of participants into groups simply ensured that each participant had an equal chance of being placed in either of the two training methods, within their respective subgroup of EHS or non-EHS professionals. Sample size determination was done using statistical power analysis, with one hundred participants (n=100) chosen to ensure a moderate effect size detection in both feedback and assessment (pre-test and post-test) outcomes between the two groups used in the study.

A study with low statistical power is prone to Type II error, where it is assumed that there is no difference or association between the study groups when, in fact, an association exists (Domb & Sabetian, 2021). This leads to the failure of finding an effect when one truly exists. A sample size of 100 respondents achieves approximately 70 % power, which is sufficient to detect a moderate effect size ($d \geq 0.5$) with an alpha level of 0.05 (see Appendix A).

2.3 Characteristics of Participants

A purposive sampling approach was used to include participants who met various criteria for the two groups. The inclusion criteria for EHS professionals were based on individuals with formal training or work experience in the fields of Environmental management, Stormwater management, or Occupational Health and Safety. Non-EHS Professionals were individuals with no formal background in EHS. Non-EHS professionals who possess some knowledge of stormwater management or have attended any type of stormwater training within the past 12 months were excluded.

2.4 Research instruments

2.4.1 Pre-Training and Post-Training Knowledge Assessment - Instrument 1

An assessment on basic stormwater concept was administered before and after the training. This consisted of multiple-choice and short-answer questions, covering stormwater regulations, best management practices (BMPs), and pollution prevention strategies. It measured participants' stormwater knowledge level through their pre-test scores. New knowledge acquisition, retention, and learning outcomes were determined through post-test scores. Questions were reviewed by two EHS experts to ensure they align with the scope and objectives of this study. Pilot testing was carried out on 5 volunteers to assess the level of difficulty, clarity, and effectiveness of the questions used for the assessment.

2.4.2 Training Feedback Survey- Instrument 2

A questionnaire based on a 5-point Likert scale was used to collect training feedback . Participants were required to select one of five specific choices for each question asked: "Strongly Disagree," "Disagree," "Neutral," "Agree," or "Strongly Agree". Scores for the feedback were assigned as follows: five marks for "Strongly Agree," four marks for "Agree," three marks for "Neutral," two marks for "Disagree," and one mark for "Strongly Disagree.". The approach quantifies responses obtained and highlights qualitative differences in the training experience from the two different groups.

2.4.3 Semi-Structured Interviews- Instrument 3

Six respondents (n=6, three from each subgroup) were randomly selected for a 20-minute semi-structured interview. The interview covered various aspects of the training, such as overall training experience, perceived value of the training approach, challenges encountered, and opportunities for further improvement. This provides deeper insight into the respondents' perception of the training received. Interviewees were encouraged to express their opinion about the training method they used.

2.5 Data collection

Participants voluntarily agreed to take part in this study. The recruitment process targeted two distinct groups: Environmental Health and Safety (EHS) professionals and non-EHS professionals. All the EHS professionals were contacted through professional networks, including industry associations, LinkedIn groups, and direct outreach via email. Non-EHS professionals were recruited through word-of-mouth referrals from colleagues, friends, and professional groups.

2.5.1 Survey Instrument and Distribution

A structured questionnaire was developed using Google Forms to collect demographic data, professional background, and participant feedback on the training program. The questionnaire link was shared electronically via email and messaging platforms, ensuring easy accessibility for all participants. The survey included both multiple-choice and open-ended questions to capture quantitative and qualitative insights.

2.5.2 Pre-Training and Post-Training Assessments

All participants completed pre-training and post-training assessments via Google Forms to evaluate knowledge acquisition and the effectiveness of the training. These assessments measured comprehension and retention of key concepts before and after the intervention. Scores were automatically recorded and collected via email notifications, then manually entered into an Excel database for analysis (see Appendix C and D).

2.5.3 Semi-Structured Interviews

The interviews were conducted in person and through correspondence, which was based on participants' preference and logistical feasibility. The interview questions gathered in-depth perspectives on training experiences, applicability to professional roles, and areas for improvement (see Appendix F).

All the responses from the interview data were transcribed for thematic analysis. Identifiable information was removed to ensure privacy, and transcripts were coded for key themes related to training effectiveness, engagement, and knowledge retention.

2.6 AI Chatbot Design

StormWise is a custom GPT specifically designed for this study using the OpenAI GPT platform. It was configured to act as a conversational stormwater learning guide, with the aim of eliciting users' curiosity and keeping them engaged regardless of their level of expertise in stormwater management. As a learning guide, the chatbot is designed to support users and provide adequate scaffolding throughout the learning period.

StormWise was configured to get the users name first and confirm their level of expertise in stormwater management. With the conversational starts designed to initiate conversation, the chatbot was also configured to address the user by name. All this was done to personalize the learning experience. To ensure users were engaged, the chatbot was configured to carry out knowledge checks, generate images when answering questions and not to give out answers when prompted by users, but rather work with the user to generate the answer to questions asked. The tone of the conversation was made to be fun, friendly and professional. It was designed to attend to the training needs of a diverse range of users, regardless of their professional background. Materials that were used for training the models were all retrieved from government approved websites and industry sources. Details of StormWise chatbot can be found in Appendix H and Appendix I.

Despite explicit instructional configuration, the chatbot continued to provide answers without probing users as instructed. This behavior contradicted the instruction to engage users in a two-way dialogue by first requesting them to attempt the question, even if it was a guess. This issue was later resolved after several reconfigurations. Another major problem that could not be resolved is the issue of generating visual illustrations when responding to users prompt. Although the chatbot was configured to enhance its explanations with illustrations or images, it ignored this directive. In many cases, it only responded with a visual aid when the user explicitly requested one, rather than providing illustrations as part of its explanation.

2.7 Procedure

Step 1: Training group allocation

One hundred and twenty-eight participants (n=100) were divided into two groups, each group having fifty (n=50) members. Each group is further divided into two sub-groups, non-EHS professionals (n=25) and EHS professionals (n=25). Conventional EHS training was used for one group, while for an AI chatbot was introduced into the other group's training. Participants are randomly assigned to the group.

Step 2: Pre-Training Assessment

Participants took a pre-training test on stormwater management fundamentals, which comprises 11 questions in a multimodal format. This was done to assess baseline knowledge on the subject.

Step 3: Training Delivery

The two different groups had their training delivered through different training modes.

The conventional training cohort received a standard pre-recorded video training, which had a duration of approximately eight minutes to complete.

The chatbot training cohort, on the other hand, were given the same pre-recorded content, but this time with the help of H5P, the video was interactive. Knowledge checks and activities were embedded into the video. In addition, an AI chatbot named StormWise was provided to serve as a learning assistant, providing the support needed for learning. StormWise was trained to provide interactive questions and collaborative learning opportunities, engage users in conversations by probing them first to understand their cognitive level before providing answers, which assisted in building new knowledge. This approach also adapted explanations based on the prompt generated by users. This was expected to promote personalized learning.

Step 4: Post-Training Assessment

All the participants from the two different groups took the same post-training test, which was made up of 12 questions. The difference between pre-test and post-test scores assisted in measuring new knowledge gained and knowledge retained from the training completed. This helped determine training effectiveness.

Step 5: Qualitative Feedback Survey

Participants completed a short survey that assessed different aspects of the training. Areas covered in the feedback include the level of engagement, users' perception of chatbot effectiveness, and training preference. Finally, open-ended questions were used to get participants' input on areas for possible improvement.

2.8 Data Analysis

Both quantitative and qualitative analysis assisted in assessing the effectiveness of each of the two training methods used in this study. This involved different statistical analyses. Paired t-tests were used to compare pre-test and post-test scores within each group. An independent t-test was also used to compare post-test scores between the AI chatbot and conventional training cohorts. This helped evaluate the effectiveness of each of the training methods in promoting engagement and enhancing learning outcomes. It also assisted in answering RQ1, which was about determining the effectiveness of AI Chatbots as a tool for delivering EHS training. The independent t-test was used for RQ3, which is concerned with understanding if participants trained using the chatbot would perform better in the post-training assessment compared to those who received conventional EHS training.

Qualitative Analysis was used for the interpreting survey feedback and interview responses obtained in this study. Independent t-test on Likert-scale responses obtained helped to determine if feedback differs across the two EHS training methods. This addresses RQ2, which proposed that feedback on the quality and effectiveness of training will differ significantly between participants trained via an AI-chatbot and those who receive conventional training.

2.9 Ethical Consideration

Participation in this study was entirely voluntary, and all participants were fully briefed on the nature, purpose, and objectives of the research before any data was collected. They were informed of their right to decline participation in any part of the study without providing a reason, including the option to skip questions in the assessment, feedback survey, or interview segments. No personal identifiers were used during data collection, and all responses were anonymous to ensure participants' privacy and confidentiality were not compromised (see Appendix B).

Effort was made to minimize potential risk, either physical or psychological, typically associated with participating in a study of this nature. Participants were provided with appropriate support throughout the study period. Random assignment to either the AI-chatbot or conventional training cohort was used to ensure fairness and equity, with each individual having an equal chance of being placed in any cohort. The study design and procedures were carefully implemented to uphold ethical standards including the protection of human subjects, equitable treatment, and the maintenance of data integrity.

Chapter 3
Results

This section presents the study's key findings. It begins with a summary of participants' demographic characteristics, followed by an analysis of assessment performance used to evaluate training effectiveness and compare learning outcomes between the AI-assisted and conventional training cohorts. These findings address RQ1, which examines the potential of AI chatbots in delivering EHS training on stormwater management, and RQ3, which examines whether there is any difference in learning outcomes as reflected in the assessment performance, between those trained with the chatbot and those using conventional method.

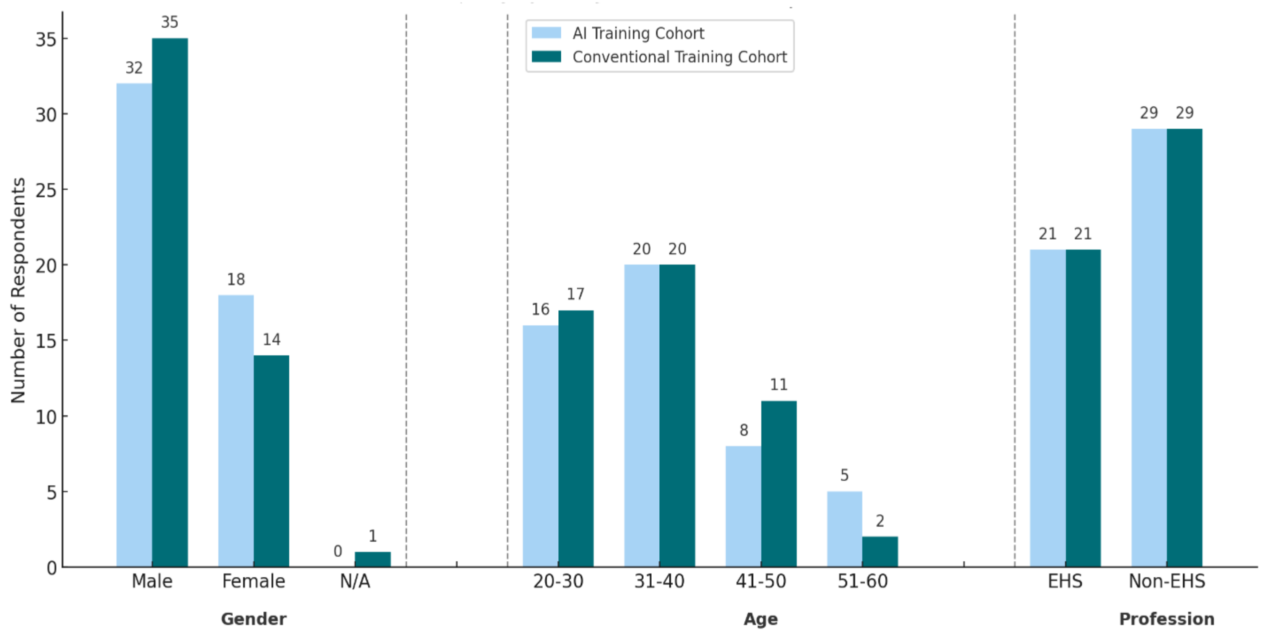
Next, findings from participant are presented, incorporating quantitative survey feedback responses, qualitative open-ended comments, and data from semi-structured interviews. This mixed-methods approach provides a comprehensive understanding of the overall training experience. The results address RQ2 which examines if there is any difference in the qualitative feedback given by those trained by a chatbot-powered EHS training and those trained using a conventional approach.

3.1 Participants' Demography

The study involved a total of 100 participants, randomly assigned into two equal cohorts: a conventional training group and an AI chatbot-supported training group (n=50 per cohort). Each cohort included 21 Environmental Health and Safety (EHS) professionals and 29 non-EHS professionals, reflecting a 42% to 58% distribution, respectively.

Participants were drawn from both the public and private sectors and were required to complete stormwater management training as part of their professional development. The sample was demographically diverse, with participants ranging in age from 20 to 60 years. As illustrated in Figure 2, the gender distribution comprised 67% male and 32% female participants.

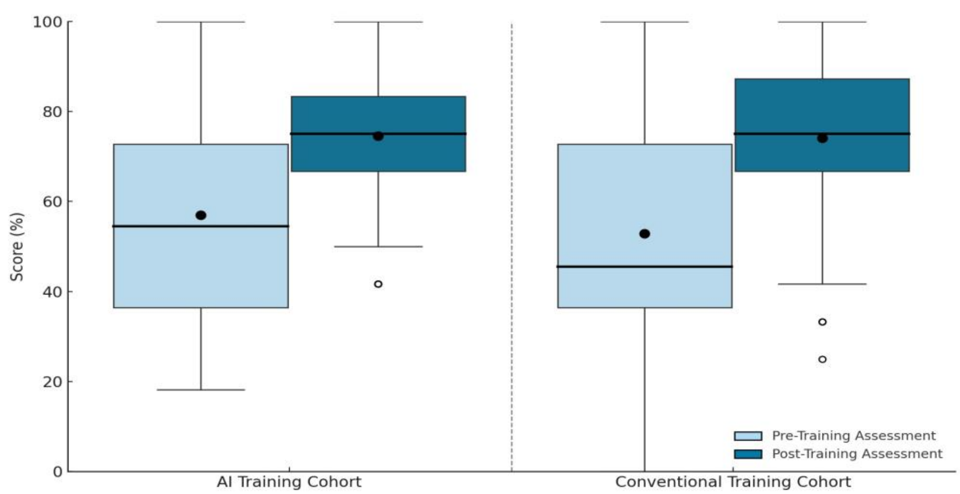
Figure 2. Respondents Characteristics



3.2 Training Performance Assessment

The pre-test and post-test consisted of 11 and 12 mixed-mode questions, respectively. These questions were developed by subject matter experts to cover key stormwater concepts.

Figure 3. Pre-training and post-training average test score

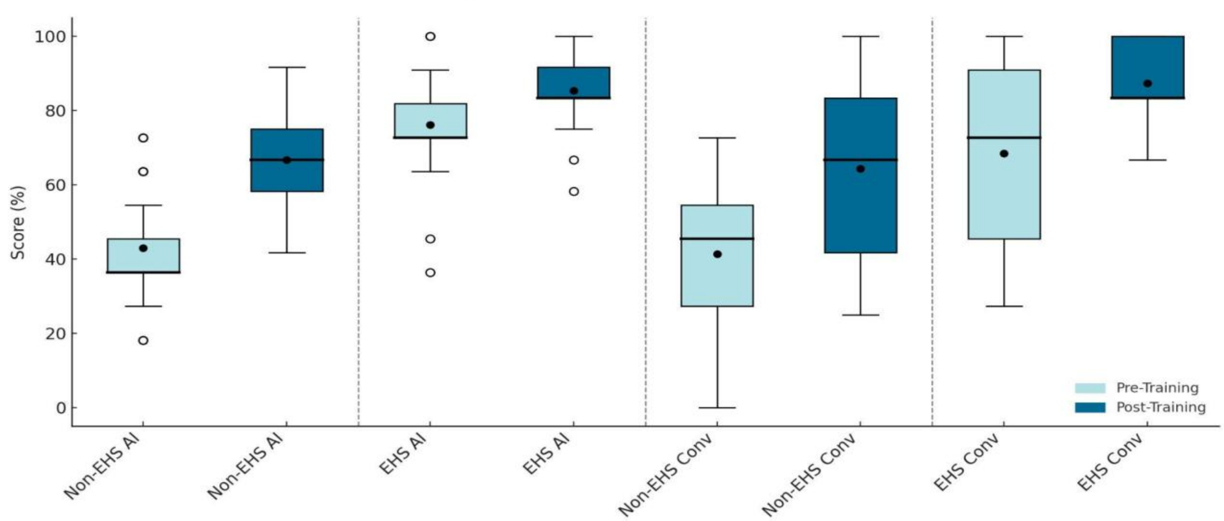


Each question was scored as 1 (correct) or 0 (incorrect). A total score of 11 and 12, respectively, was obtained for pre-test and post-test, and was later converted to percentage (%) for each participant. These scores provide a quantitative measure of learning outcomes. Figure 3 summarizes pre-training and post-training average test scores for both AI and Conventional cohorts. The boxplot compares pre- and post-training assessment scores for AI and conventional training cohorts. Both groups show a significant increase in both the median and mean scores after training. The interquartile range narrows for post-training scores in both cohorts (from 36.3 to 16.6 and 36.3 to 22.9 in AI and conventional cohorts, respectively), suggesting reduced variability and a more consistent performance after training. Outliers reflect the few participants who scored unusually low, compared to the rest of the participants in their respective cohorts.

Figure 3 further shows that the AI-trained cohorts recorded a minimum post-training mean score of 41.7%, which is higher than the 25% recorded by the conventional trained cohorts. The data shows an increase in mean post-test scores for both cohorts (increased from 56.9% to 74.5% and from 52.7% to 74% in AI and Conventional cohorts, respectively), and with $p<0.01$. However, when the pre-training and post-training scores of the AI and conventional cohorts were compared, using a paired t-test, the difference was not statistically significant ($p> 0.05$).

Overall, this data suggests that both training methods were effective, with participant in the AI cohort performing slightly better than their peers in the conventional training cohort.

Figure 4. Average assessment scores by profession



This expanded boxplot in Figure 4 breaks down pre- and post-training assessment scores by professional background (EHS vs. non-EHS) and training type (AI vs. Conventional). Across all subgroups, post-training scores are again higher than pre-training scores, attesting to the training's effectiveness. Notably, EHS professionals in both AI and conventional cohorts show the highest post-training mean scores (76.2% to 85.3% and 68.4% to 87.3% in AI and Conventional training cohorts respectively ($p<0.01$ in both cases), reflecting a likely advantage from prior subject familiarity. Non-EHS participants also improved from 43.0% to 66.7% and 41.4% to 64.4% in AI and conventional cohorts, respectively ($p<0.01$ in both cases), but with greater variability and lower medians, especially in the conventional cohort. Again, the tight clustering of post-training average scores for EHS professionals in the AI cohort suggests particularly strong and consistent learning outcomes.

3.3 Quantitative Feedback

Participants' feedback on the training experience provides insights into how the two different training methods influenced their engagement and perceptions, addressing RQ1 (perceived effectiveness) and RQ2 (differences in feedback between cohorts). Table 1 summarizes the mean ratings for evaluation of selected survey items for each cohort, along with statistical comparisons.

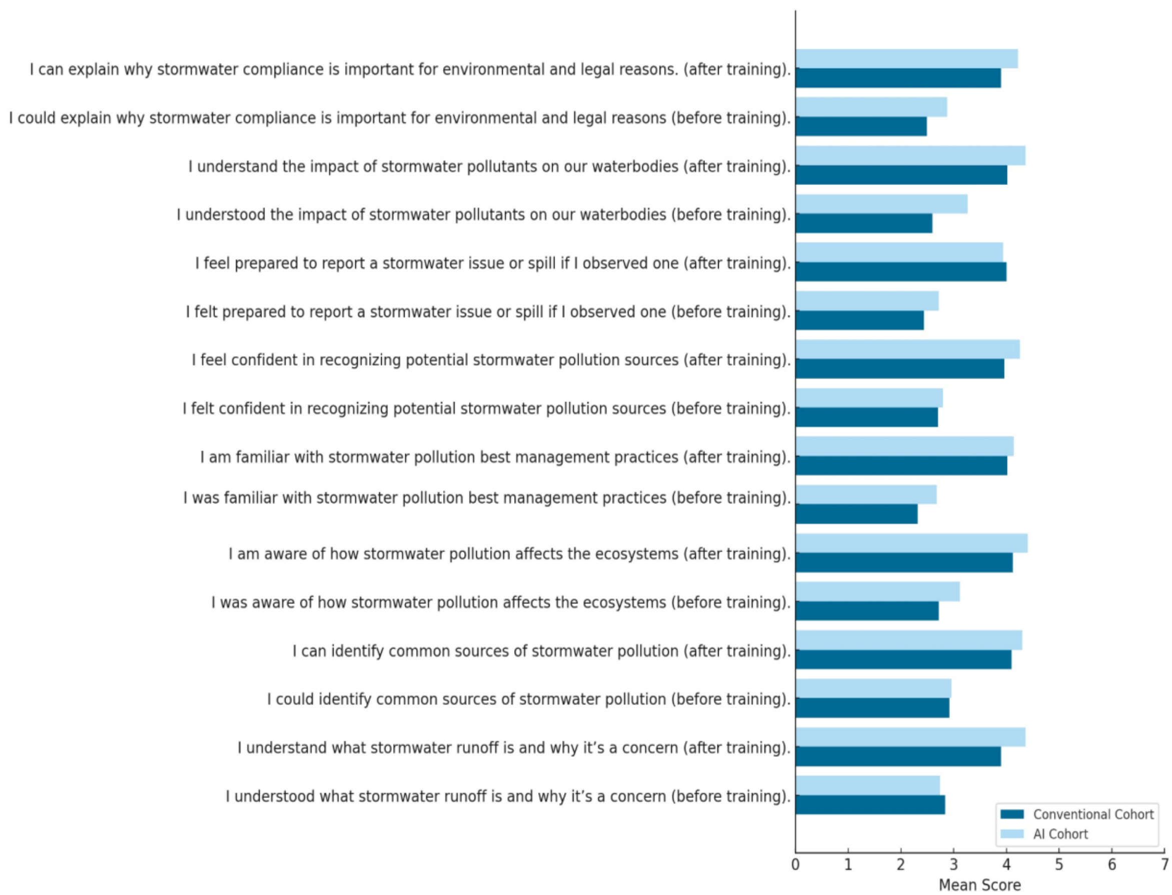
Table 1. Average scores from respondents' feedback. (Using a 5-point Likert scale)

	AI Cohort (n=50)	**Conventional Cohort (n=50)**	
Feedback Description	**(Mean±SD)**	**(Mean±SD)**	**p-value**
i) The training content was clear and easy to understand.	3.98±1.25	3.86±1.21	p>0.05
ii) The training was engaging.	3.90±1.20	3.62±1.40	p<0.05
iii) The training helped me retain key concepts.	3.88±1.26	3.68±1.20	p<0.05
iv) Effectively supports and enhances learning.	3.90±1.22	3.94±1.20	p>0.05
v) The training made learning more interactive.	3.88±1.24	3.52±1.45	p<0.01
vi)The training will encourage participation and course completion.	3.86±1.14	3.82±1.30	p>0.05
vii) The training method promotes personalized learning.	3.58±1.32	3.84±1.35	p>0.05
viii) The training method promotes self-directed learning.	3.59±1.26	4.02±1.22	p<0.05
ix)The training method supports collaborative learning.	3.43±1.27	3.30±1.36	p>0.05

Note. SD – standard deviation. Each description contained five statements, to which respondents could answer on a 5-point Likert scale (1 – strongly disagree, 5 – strongly agree).

The Chatbot cohort rated the training significantly higher on engagement ($p<0.05$), retention of key concepts ($p<0.01$), interactivity ($p<0.01$), and collaborative learning ($p>0.05$) compared to their peers in the conventional cohort. Interestingly, the data shows that many participants from conventional cohorts (4.02±1.22) believe that their training method supports self-directed learning (with $p<0.05$), unlike their peers from AI cohort (3.59±1.26). These results suggest that the inclusion of a chatbot could have made the training more engaging and interactive for participants, which further helps in addressing RQ1.

Figure 5. Survey Results of Participants' Feedback



Note. Each subscale attribute contained five statements, to which respondents could answer on a 5-point Likert scale (1 – strongly disagree, 5 – strongly agree).

Figure 5 shows that the two cohorts rated themselves much higher "after-training" mean score than in "before training" on the same statements, which attests to their acknowledgment of new knowledge acquisition. A paired t-test shows that $p<0.01$. On average, AI cohorts consistently rated themselves higher than their peers in the conventional cohorts in almost all ‘after-training” statements assessed.

At the subgroup level, further analysis shows that both AI-assisted and conventional training methods significantly improved participant understanding of various aspects of the training, regardless of professional background. Using a paired t-test, $p<0.01$. In all cases, “after-training” mean scores were higher than “before-training” scores, further underscoring the success of the training interventions in enhancing stormwater management knowledge and confidence across the board.

3.4 Chatbot-specific feedback

Participant feedback on the AI chatbot was moderately positive across most dimensions assessed. It can be seen from Table 2, the highest-rated items were "*I felt comfortable asking the AI chatbot questions during the training session*" (3.61±1.35) and "*It was easy to use the AI chatbot during the training session*" (3.55±1.42), indicating a user-friendly and approachable interface. Trust-related items, such as "*I trust the answers given by the AI chatbot*" (3.35±1.20) and "*I trust my interactions with the AI chatbot are confidential*" (3.18±1.26), received slightly lower scores, suggesting room for improvement in perceived reliability and privacy.

Table 2. Summary of average score from Chatbot usage feedback

Feedback Description	(Mean±SD)
It was easy to use the AI chatbot during the training session.	3.55±1.42
The chatbot responded in a way you expected it to.	3.52±1.31
I did not encounter issues or bugs while using the AI chatbot.	2.92±1.50
I saw most of the features I expected to see in the AI chatbot.	2.69±1.45
I interact better with the AI chatbot compared to a human facilitator.	2.90±1.33
I intend to continue using the AI chatbot in the future.	3.47±1.37
I trust my interactions with the AI chatbot are confidential	3.18±1.26
I felt comfortable asking the AI chatbot questions during the training session.	3.61±1.35
I trust the answers given by the AI chatbot.	3.35±1.20
The AI chatbot responses sound natural and not like a robot.	3.20±1.19
The AI chatbot felt tailored to the training content and was not too generic.	3.52±1.21

Note. SD –Standard Deviation. 5-point Likert scale, from 1 (Strongly Disagree) to 5 (Strongly Agree).

Lower ratings were also observed for expectations and personalization, particularly "I saw most of the features I expected to see in the AI chatbot" (2.69±1.45), hinting at gaps between user expectations and actual chatbot capabilities. Overall, the feedback suggests moderate usability and acceptance, with opportunities to enhance users' experience, technical issues, features that promote conviviality, functionality, and perceived efficiency.

3.5 Qualitative Feedback

The open-ended responses provided feedback on chatbot user experiences and preferences during the training. In this section, a summary of the thematic analysis of the qualitative data is presented, which examines what users in each cohort liked or disliked about the training, along with their recommendations for improvement. This will assist in addressing RQ2, which deals with differences in qualitative feedback. The analysis identifies recurring themes in participant responses, providing a nuanced understanding of perceived training quality and user satisfaction across the two instructional approaches.

3.5.1 Most Useful Aspects

In response to the open-ended question, “What aspects of this course were most useful or valuable?”, participants from both the AI-assisted and conventional training cohorts highlighted several recurring themes (see Appendix E1).

The majority of participants from the AI Cohort mentioned the interactive and engaging nature of the course as a major benefit. For instance, one participant wrote, “*The interactive nature of the course, I believe it made the course more engaging and interesting*”. Another person said, “The *knowledge checks that test the knowledge of the students during training enhance learning”*. Many enjoyed being able to ask questions directly and receive real real-time response, with the chatbot proving adequate scaffolding and guided learning. One person said, “*The custom ChatGPT made it possible to get more information on abstract concepts, making the learning experience great”.*

A few also noted that the combination of video and chatbot was useful, with the video providing baseline knowledge, and the chatbot could clarify doubts immediately. One participant summarized, *“it was a great introduction to the subject. The in-video quizzes allowed me to think and then learn afterwards, with ChatGPT providing more detail when I asked for it”.*

On the other hand, the majority of the participants from the Conventional Cohort pointed to specific content elements as most useful. For instance, many liked the practical examples given in the video (such as demonstrations of BMPs or the “first flush” concept). One satisfied participant said, “*The video had good visuals and clear audio. It covered various aspects of stormwater in a short timeframe,*” and another person said, “*Its structure, use of images/graphics, pace, and delivery.*” Overall, content clarity and structured presentation were valued.

3.5.2 Least Liked Aspects

Participants were asked to identify the least favorable aspects of the training experience. The responses revealed distinct thematic patterns across the two cohorts (see Appendix E2). In the AI training cohort, participants frequently mentioned technical issues and uncertainty about how to use the chatbot effectively, which sometimes led to cognitive strain. Additionally, some users expressed dissatisfaction with the lack of multimedia elements in chatbot responses and the perceived vagueness of replies. Others expressed their distrust of such technology as a learning tool. Aside from chatbot technical issues, some disliked elements of how the training was structured. One person didn't like that the self-assessments, which in most cases came before the content was explained. "*Having to answer questions before the information was presented will equip me with the knowledge to answer the questions. I would have preferred it afterwards*," the user said. Another user highlighted the issue of poor visuals. One user also stated explicitly that, "*The chatbot is not good at creating images*". It is important to note that several participants from the chatbot cohort also wrote "*none*" or "*nothing*" in the disliked aspect, indicating they had no major complaints. Some stated they liked all aspects of the training.

On the other hand, respondents in the conventional cohort were more critical of the passive delivery method, highlighting the lack of interactive features and group engagement. Some users pointed out that there is no means of asking questions during the training, with one stating explicitly, "*Not interactive enough*", and another, "*Some terms were not properly explained*". Cognitive overload was also cited here. A user said, "*Too professional. Not designed for people not in this field.*" This lack of engagement and interaction led to a user referring to some parts of the training as boring, "*Being made to listen and read long texts in the video was not a pleasant experience. It was boring to read. Maybe because I am science inclined.*"

These findings help address RQ2 by highlighting notable differences in user experience between the two instructional modalities

3.5.3 Suggestions for Improvement

When the participants were asked to make recommendations for improving the training, their responses revealed distinct thematic patterns across the two cohorts (see Appendix E3).

For the AI Chatbot cohort, their responses revolved around enhancing the chatbot integration and features, "*It will be interesting to have the chatbot support its explanations with some visuals*", one user said. Participants also wanted the chatbot to show images or diagrams when generating outputs. This was seen by some users as a major limitation of ChatGPT due to inability to generate images and use such for explanation. Another participant suggested embedding the chatbot in the video itself so that you don't have to keep switching tabs. This indicates a desire for smoother user experience. One participant said, "*If it were possible to have the ChatGPT embedded in the video, that would be great.*"). There were calls to "*make sure users have full access to the GPT before beginning*," which will entail sorting out the usage limit issue challenge encountered by a participant. Group Interaction was another point raised. "*Make it possible for other people to join in the conversation, rather than just the user and the chatbot.*" This hints at a multi-user chat or a facilitator joining – perhaps the participant envisioned a collaborative learning with multiple trainees and the Chatbot or having a human monitor. This suggestion explores an opportunity for blending AI with human facilitation.

However, when asked to make suggestions for improvement, some participants from the conventional cohort highlighted the need to improve on interactivity and engagement. For example, "*Make the video interactive by breaking it into parts and asking questions straight after each part*", "*Add a game or some activity for users to practice*". These ideas describe adding what the chatbot provided is interaction and opportunity to ask questions and or other active learning elements used by AI chatbot cohorts. One participant explicitly wrote, "*Maybe use interactive videos, also adding some form of collaboration will make it more appealing*."

Some participants stated they were satisfied with the training, requiring no further improvement. "*No improvement from me, it was perfect*". Others responded with a "*nil*", "*none*", "*I like all*"

In summary, some of the Chatbot cohort's feedback often centered on the positive experience, such as a useful interactive helper or high engagement, and on the negatives well, such as technical issues. The Conventional cohort's feedback focused on the content delivery, with multiple calls to make it more interactive and engaging. Essentially, many participants from the conventional cohorts were asking for features that the chatbot cohort participants had already encountered during their training, such as engagement and interactivity. The absence of a chatbot in their training led them to specifically yearn for interaction, support, or an easier way to learn the content, suggesting that their qualitative feedback differs from the focus of members from the chatbot cohort.

3.6 Semi-Structured Interview Feedback

Thematic analysis of the semi-structured interview responses revealed distinct differences in participant experiences between the AI and conventional training cohorts (see Appendix F). All AI participants found the training highly engaging and interactive, with 67% emphasizing its personalized and engaging design approach. However, technical challenges were a common issue, with two-thirds (67%) reporting difficulties accessing or navigating the chatbot and expressing confusion over certain instructions. One respondent suggested that chatbot responses lacked sufficient visual reinforcement. Despite these limitations, all AI participants expressed increased awareness of stormwater issues and a strong preference for future chatbot-based training.

In contrast, the conventional training cohort demonstrated a smoother user experience, with 100% describing the format as clear and easy to use, and none reporting having technical challenges. Engagement levels were slightly lower, with only 33% finding the training interactive and engaging, while 67% expressed a desire for greater interactivity. While both groups equally appreciated the visual materials (67%) and reported enhanced understanding of stormwater management (100%), the conventional cohort's feedback centered more on content delivery improvements, such as shortening video length, making training more interactive, and incorporating collaborative features. Overall, these findings highlight notable differences in usability and engagement, which seem to answer RQ3 that the training method contributes to shaping and influencing learner perceptions and outcomes.

Chapter 4

Discussion

This section presents an interpretation of the study's findings on the integration of AI chatbots into stormwater management training. The results are examined in relation to their impact on learning outcomes and the overall learner experience. Broader implications of the findings are also explored.

4.1 Effectiveness of AI Chatbot in EHS Training- RQ1

Integrating a chatbot into the EHS training was expected to improve content delivery by making the training engaging and interactive. This learner-centered training approach aims to enhance understanding and ensure optimal learning outcomes. The findings of this study provide substantial support for RQ1, which proposes that Chatbots can be used as a tool for delivering EHS training that enhances engagement, retention of knowledge, and performance outcomes. The learning improvement seen in the AI chatbot group's post-training mean score demonstrates that using chatbots in delivering EHS training did not hinder learning, instead, it provided sufficient support for learners.

From a learning perspective, participants in the AI chatbot-assisted cohort exhibited improvement between pre- and post-training assessments, indicating successful conceptual understanding. Also, looking at the results from a delivery perspective, participant feedback demonstrates that the chatbot added instructional value by enabling real-time interaction, clarifying abstract content, and making the training more engaging, capabilities that were absent in the static video-based format of the conventional cohort. These findings align with other studies that demonstrate the positive influence of chatbots on user engagement, instructional clarity, and perceived learning outcomes (Shalong et al., 2024; Zhao et al., 2023)

However, the results also highlight several challenges associated with chatbot-enhanced learning environments. Some participants from the AI cohort reported issues related to the chatbot's inability to generate visuals, occasional technical glitches, and unsatisfactory responses. Since chatbot interactions are prompt-driven for instance (Dang et al., 2022; Wang et al., 2024), the quality and specificity of user queries could have directly influenced the relevance and depth of the responses received. This user-driven dynamic can lead to highly individualized learning experiences, with some more effective than others, thereby contributing to differing levels of user satisfaction. The problem of technical glitches, which was also found in a similar study(Huang et al., 2022), highlights a major challenge associated with using AI chatbot for learning purposes. Such disruption can make users lose trust in such learning tools. As a result, people may not be as enthused about using AI chatbots as a learning resource, potentially limiting their effectiveness for training interventions. These limitations suggest that while chatbots offer a potentially rewarding learning experience, their effectiveness depends on other support systems, such as instructional system design and improved user familiarity and proficiency in interacting with AI technology.

4.2 Evaluating Learner Feedback Based on Training Method- RQ2

RQ2 proposed that feedback on the quality and effectiveness of training will differ significantly between participants from the two cohorts. This expectation is based on the assumption that the inclusion of an AI chatbot would offer a more interactive, personalized, and responsive learning environment, which should lead to better learning outcomes compared to a conventional training approach. Participants' subjective learning experiences were expected to vary based on the type of training approach used.

The results obtained from this study provide enough evidence that integrating an AI chatbot within the EHS training enhances engagement, interactivity, and overall perceived effectiveness of the training approach, unlike the conventional training method. The findings of this study are consistent with other studies, which also demonstrated the transformative role of AI chatbots in adult education and professional development (Basri et al., 2025; Deng & Yu, 2023; Wu & Yu, 2024).

The study reveals a major contrast in user experiences between the two cohorts. Participants in the AI cohort expressed satisfaction with the opportunity to ask questions and receive real-time responses, features that contributed to more engaging and interactive learning, which shows that such adaptive feedback enhances motivation and attention. In a professional setting, learners are more inclined to embrace opportunities that promote active engagement, which the AI chatbot offers, rather than passive content delivery, which is associated with a conventional learning approach. In contrast, participants in the conventional cohort highlighted the lack of interactivity, inability to ask questions, and absence of adequate scaffolding as key limitations. These concerns were repeatedly identified as areas for improvement, reinforcing the value of chatbot integration in enhancing engagement and interaction within EHS training programs, enhancing the learning process. The lack of interactivity and support for learners has been identified as a major limitation of static conventional training method (Samuel Omokhafe Yusuf et al., 2024).

As shown in Table 1, participants in the AI-assisted cohort reported significantly higher ratings for training engagement and interactivity, both statistically significant at $p < .05$ and $p < .01$, respectively. They also rated the training higher on knowledge retention, suggesting that the conversational nature of chatbot interactions may have facilitated deeper cognitive processing. In contrast, the conventional cohort's ratings for these same aspects were notably lower, reinforcing the anticipated differences in learner perceptions.

Interestingly, one divergence from expectations emerged, participants in the conventional group rated their training more favorably on promoting self-directed learning than their AI-trained peers, a difference that was also statistically significant. This may reflect the participants' understanding of learner autonomy, given that conventional cohorts completed their training without any interactive support tools. Without scaffolding provided by a chatbot, participants may have figured out a way to regulate their learning on their own, which could have inadvertently fostered stronger self-directed learning.

Qualitative feedback further substantiated these distinctions. Participants in the AI cohort frequently highlighted the interactive and engaging nature of their training as the most valuable aspect. They appreciated the chatbot's ability to clarify abstract concepts in real time and to provide guided learning experience. In contrast, conventional cohort participants emphasized content clarity and visual design but criticized the training's lack of opportunities for interaction or clarification. Several respondents expressed frustration at being unable to ask questions during training and expressed dissatisfaction with this training approach. This suggests that aligning instructional methods with trainees' learning preferences can also enhance learning outcomes.

The thematic analysis of open-ended responses reinforces the contrast between the two training cohorts. Participants from the conventional cohort requested more engaging and interactive elements in their training, features already present in the AI cohort's experience. This says a lot about the learners' training expectations, especially in professional settings like Environmental Health and Safety. This study also shows that people no longer want to sit through symbolic traditional non-interactive videos or read long documents anymore. 21st-century learners desire to be involved in their learning process- ask questions, get real-time feedback, have agency, and explore content in ways that fit their learning style. The AI chatbot seems to have given participants this kind of experience, which likely explains the higher rating it received.

Finally, semi-structured interviews highlighted similar contrasts. All participants in the AI cohort reported that the training was interactive and engaging, while only 33% of conventional participants described their training in the same way. Moreover, two-thirds of AI participants experienced technical difficulties, while conventional participants largely enjoyed a seamless experience. Despite these challenges, the AI cohort still preferred chatbot-enhanced training for future learning, a good indication of the perceived value they have associated with this training method.

In conclusion, the findings strongly support RQ2. The inclusion of a chatbot created a different and more interactive user experience, which contributed to shaping learners' satisfaction, engagement, and perception of the training approach. These outcomes reinforce the role of AI-mediated interaction in transforming adult learning environments.

4.3 Assessment Performance Differences- RQ3

From RQ3, it was anticipated that the chatbot might give a learning advantage to users, perhaps based on the idea of personalized tutoring. Unlike conventional instruction, it was expected that chatbot-assisted training would yield superior learning outcomes. In theory, by providing on-demand explanations and tailored guided learning, the chatbot should help learners achieve a deeper understanding of the material, which should translate into higher post-training test scores for AI cohorts.

However, the assessment data did not provide statistical support for this assumption. As shown in Figure 2, the average post-training scores for both the AI and conventional cohorts were nearly identical, with no significant difference between the two groups ($p > 0.05$). While both cohorts exhibited marked improvement from pre- to post-training, the marginal difference between them suggests that the chatbot's instructional advantage, while potentially beneficial in terms of engagement and interactivity, did not yield measurable performance gains in short-term knowledge assessments. Further breakdown of the results into professional background (Figure 3) reinforced this pattern. Both EHS and non-EHS professionals in the AI and conventional groups showed similar trends, increasing post-test scores, with no significant differences when the two cohorts are compared ($p > 0.05$).

This finding suggests that although the chatbot may enhance the perceived quality of the learning experience, it does not necessarily lead to a quantifiable advantage in terms of assessment outcomes. Similar conclusions have been reached in recent studies where chatbot integration improved learner motivation, self-efficacy, and metacognition but had a limited impact on objective test performance (Bauer et al., 2025; Li, 2023; Urban et al., 2024). This suggests that while educational chatbots can enhance engagement and perceived learning quality, their impact on measurable learning outcomes, such as in assessment, is often context dependent. Other studies have found that chatbots in education have shown a significantly positive effect on student learning performance (Wang & Fan, 2025; Laun & Wolff, 2025), suggesting that the benefits, while real, is not transformative across all settings. These results align with the current study's outcome, where both chatbot-assisted and conventional cohorts showed similar gains in post-training assessments.

The absence of a significant effect may be attributed to several factors, including the relatively short duration of the training, limitations in chatbot design, or variability in user interaction. Learners unfamiliar with AI tools may not have fully leveraged the chatbot's capabilities, while others may have received inconsistent responses due to variations in prompt quality. Another point is that the material used for the training could have been well designed, such that even without a chatbot, participants were able to understand the concepts covered in the training. The chatbot essentially served as a backup, considering that there is a possibility that some participants in the chatbot group might not have extensively utilized it. In such a situation, participants learned mostly from the video, just like those in the conventional cohort, which explains why average post-training scores could have been similar. The assessment design could have also contributed to the performance outcome. This study measured the training content recall ability of participants. The short training used for this study served as an introduction to basic stormwater concepts, which both cohorts could have understood well enough to answer the assessment questions, hence the similarity in assessment outcome.

4.4 Study Limitations

There are several limitations of this study that need to be acknowledged. The sample size limits the study findings generalizability. While it gives an idea of the two training approaches studied, a larger and more diverse sample drawn from different contexts would have given a better outcome.

The assessment instrument focused on post-training short-term recall. This approach fails to take into account long-term retention or knowledge transfer to a real-world setting, which is the goal of such professional training. The questions asked were all covered in the training material, which may have an impact on the assessment performance for the two groups. Also, considering that the participants in the AI cohort were given several knowledge during their training, that could have skewed the outcome of the study.

Some participants in the study were using an AI training chatbot for the first time and expressed their difficulty in understanding how to use it. This unfamiliarity may have affected their engagement levels and also limited their ability to fully utilize the AI chatbot effectively. Additionally, the way the AI Chatbot was customized may not have adequately supported users across different digital literacy levels, which could have limited their engagement with it.

Finally, some of the data used relied solely on self-reporting, which is inherently prone to bias. Participants could have either underestimated or exaggerated the various attributes that were measured, which could have affected the reliability and validity of the data used for this study.

4.5 Conclusion

This thesis evaluated the effectiveness of AI chatbots in delivering an EHS training on stormwater management compared to using a conventional training approach, and the results show that while the chatbot did not produce significantly higher test scores than conventional training, it enhanced the learning experience through greater interactivity and personalized support. The chatbot-assisted and traditional training cohorts achieved substantial knowledge gains, demonstrating that the core training content was effectively delivered in both formats. However, participants with access to the AI chatbot reported a more engaging and interactive learning process and did not express the inability to ask questions on unclear concepts that several traditional learners did.

The findings provide sufficient evidence that supports the integration of AI chatbots into workplace training programs. It creates an opportunity to actively provide real-time feedback and support learning in all environments, most importantly, where human assistance is not readily available. This study also shows that instructional design remains the key to effective learning and that integrating an AI chatbot complements rather than displaces established systems, ensuring optimal learning outcomes.

In conclusion, this study contributes to the growing understanding that AI in education is about enhancing the learning process. AI chatbots represent a promising, transformative tool that can be utilized in revitalizing adult learning and professional development experiences. It has the potential to create an interactive learning environment that caters to the needs of learners from diverse backgrounds. enhance learning outcomes or boost assessment scores. For inclusion of AI chatbots into EHS programs to address equity issues, and offer sustainable, optimal pedagogical value, addressing major challenges such as associated technological glitches, and making provisions for users still unfamiliar with this new piece of technology will be required.

4.6 Recommendation

This study opens up several opportunities for further academic investigation.

Conducting a longitudinal study using participants drawn from two different organizations to assess whether knowledge gained is sustained over time. Here, a post-training assessment should be conducted at least three months after the initial training, and an evaluation of the impact of the training on participants' attitudes and engagement with their organization's stormwater management program should also be carried out. Effectiveness should be further measured by monitoring increased participation in stormwater initiatives, reduction in violations to their organization's Spill Prevention, Control, and Countermeasures (SPCC) plan, in addition to improvements in post-training test scores, participants' feedback. This provides a comprehensive understanding of both knowledge retention and real-world behavioral change.

Another area that could be investigated further is the differences in the impact of AI chatbots between group and individual based learning dynamics. This study will explore whether groups of participants engaging collectively with a shared AI chatbot experience different learning dynamics, level of engagement, or learning outcomes compared to individually interacting with the chatbot. This line of inquiry could provide a good insight into the influence of collaborative and individualized AI-mediated learning environments in enhancing participation, engagement, knowledge retention, and the overall effectiveness of EHS training programs. Such information will play a key role in optimizing AI chatbots for different instructional contexts.

Acknowledgements

Special gratitude to my supervisor, Yaroslav Opanasenko, whose guidance, encouragement, and insightful feedback made the completion of this research possible. Your support and contributions were truly remarkable.

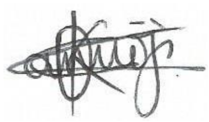
My sincere appreciation also goes to Prof. Emanuele Bardone, Director of the Educational Technology program, and to all the lecturers who made this academic journey a rewarding experience.

To my friends and family, I truly appreciate all your encouragement and unwavering support.

Finally, I dedicate this work to the loving memory of my late father, Sir Nsa Kanu-Oji.

Author's Declaration

I hereby declare that I have written this thesis independently and that all contributions of other authors and supporters have been referenced. The thesis has been written in accordance with the requirements for graduation thesis of the Institute of Education of the University of Tartu and is in compliance with good academic practices.



Chukwuemeka Oji Kanu-Oji 30/05/2025

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Chukwuemeka Oji Kanu-Oji ,

(author´s name)

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

Exploring the Impact of AI Chatbot Integration on Learning Outcomes in Environmental Health and Safety Stormwater Training ,

(title of thesis)

supervised by Yaroslav Opanasenko ;

(supervisor´s name)

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Chukwuemeka Oji Kanu-Oji 30/05/2025

References

Abbasnejad, B., Soltani, S., & Wong, P. (2024). A systematic review of online learning and teaching strategies during the COVID-19 pandemic: Implications for the construction management sector. *Smart and Sustainable Built Environment*, *13*(4), 934–959. https://doi.org/10.1108/SASBE-08-2022-0174

Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, *32*(10), 6159–6172. https://doi.org/10.1080/10494820.2023.2253858

Alimardani, A., & Jane, E. (2025). *Human Hallucinations: GenAI and the Mirage of Personalised Learning for All*. SSRN. https://doi.org/10.2139/ssrn.5016289

Amiot, C., Charoy, F., & Dinet, J. (2025). Chatbots in Collaborative Settings and their Impact on Virtual Teamwork. *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1–18. https://doi.org/10.1145/3710945

Barati Jozan, M. M., Ghorbani, B. D., Khalid, M. S., Lotfata, A., & Tabesh, H. (2023). Impact assessment of e-trainings in occupational safety and health: A literature review. *BMC Public Health*, *23*(1), 1187. https://doi.org/10.1186/s12889-023-16114-8

Basri, W. Sh., Attar, R. W., Albagmi, S., Alibrahim, D., Alanezi, F., Almutairi, S. A., AboAlsamh, H. M., Alsedrah, I. T., Arif, W. M., Alsadhan, A. A., AlSanad, D. S., Alsultan, W. S., Alabdullateef, H., AlShammary, M. H., Bakhshwain, A. M., Almuhanna, A. F., Almulhem, M., Alnaim, N., & Alhazmi, A. H. (2025). Effectiveness of ChatGPT for educators professional development: An empirical study with medical faculty. *Nutrition and Health*, 02601060251315795. https://doi.org/10.1177/02601060251315795

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, *37*(2), 45. https://doi.org/10.1007/s10648-025-10020-8

Bęś, P., & Strzałkowski, P. (2024). Analysis of the Effectiveness of Safety Training Methods. *Sustainability*, *16*(7), 2732. https://doi.org/10.3390/su16072732

Boubker, O. (2024). From chatting to self-educating: Can AI tools boost student learning outcomes? *Expert Systems with Applications*, *238*, 121820. https://doi.org/10.1016/j.eswa.2023.121820

Burke, M. J., Salvador, R. O., Smith-Crowe, K., Chan-Serafin, S., Smith, A., & Sonesh, S. (2011). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology*, *96*(1), 46–70. https://doi.org/10.1037/a0021838

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative Effectiveness of Worker Safety and Health Training Methods. American Journal of Public Health, 96(2), 315–324. https://doi.org/10.2105/AJPH.2004.059840

Cham, H., Lee, H., & Migunov, I. (2024). Quasi-experimental designs for causal inference: An overview. *Asia Pacific Education Review*, *25*(3), 611–627. https://doi.org/10.1007/s12564-024-09981-2

Caffaro, F., Micheletti Cremasco, M., Bagagiolo, G., Vigoroso, L., & Cavallo, E. (2018). Effectiveness of occupational safety and health training for migrant farm workers: A scoping review. *Public Health*, *160*, 10–17. https://doi.org/10.1016/j.puhe.2018.03.018

Dang, H., Mecke, L., Lehmann, F., Goller, S., & Buschek, D. (2022). *How to Prompt? Opportunities and Challenges of Zero- and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2209.01390

Deng, X., & Yu, Z. (2023). A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. *Sustainability*, *15*(4), 2940. https://doi.org/10.3390/su15042940

Domb, B. G., & Sabeti an, P. W. (2021). The Blight of the Type II Error: When No Difference Does Not Mean No Difference. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, *37*(4), 1353–1356. https://doi.org/10.1016/j.arthro.2021.01.057

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Ejjami, R. (2024). The Adaptive Personalization Theory of Learning: Revolutionizing Education with AI. *Journal of Next-Generation Research 5.0*. https://doi.org/10.70792/jngr5.0.v1i1.8

Elbanna, S., & Armstrong, L. (2024). Exploring the integration of ChatGPT in education: Adapting for the future. *Management & Sustainability: An Arab Review*, *3*(1), 16–29. https://doi.org/10.1108/MSAR-03-2023-0016

Environmental Protection Agency. (2008) NPDES Stormwater Program.

https://www.epa.gov/npdes/npdes-stormwater-program.

Evans, M. J., & Moore, J. S. (2013). Peer tutoring with the aid of the Internet. *British Journal of Educational Technology*, *44*(1), 144–155. https://doi.org/10.1111/j.1467-8535.2011.01280.x

Fentiman, A., & Zabbey, N. (2015). Environmental degradation and cultural erosion in Ogoniland: A case study of the oil spills in Bodo. *The Extractive Industries and Society*, *2*(4), 615–624. https://doi.org/10.1016/j.exis.2015.05.008

Fereshtehpour, M., & Najafi, M. R. (2025). Urban stormwater resilience: Global insights and strategies for climate adaptation. *Urban Climate*, *59*, 102290. https://doi.org/10.1016/j.uclim.2025.102290

Flenady, G., & Sparrow, R. (2025). Cut the bullshit: Why GenAI systems are neither collaborators nor tutors. *Teaching in Higher Education*, 1–10. https://doi.org/10.1080/13562517.2025.2497263

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges. *Review of Research in Education*, *44*(1), 218–243. https://doi.org/10.3102/0091732X20903302

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Irkha, A., Hurskaya, V., Hryshchuk, M., Tereshchuk, V., & Chyrva, H. (2024). Digital Transformation in Education: Leveraging Technology for Enhanced Learning Experiences. *Futurity Education*, *4*(3), 4–17. https://doi.org/10.57125/FED.2024.09.25.01

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Johnson, B. G., Phillips, F., & Chase, L. G. (2009). An intelligent tutoring system for the accounting cycle: Enhancing textbook homework with artificial intelligence. *Journal of Accounting Education*, *27*(1), 30–39. https://doi.org/10.1016/j.jaccedu.2009.05.001

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability*, *15*(16), 12451. https://doi.org/10.3390/su151612451

Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: What is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology*, *39*(1), 6–36. https://doi.org/10.1080/17439884.2013.770404

Laun, M., & Wolff, F. (2025). Chatbots in education: Hype or help? A meta-analysis. *Learning and Individual Differences*, *119*, 102646. https://doi.org/10.1016/j.lindif.2025.102646

Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, *39*(5), 40–58. https://doi.org/10.14742/ajet.8923

Li, S., Hong, Y.-C., & Craig, S. D. (2023). A Systematic Literature Review of Social Learning Theory in Online Learning Environments. *Educational Psychology Review*, *35*(4), 108. https://doi.org/10.1007/s10648-023-09827-0

Lo, P. Y. (2024). An Autoethnographic Reflection of Prompting a Custom GPT Based on Oneself. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3613905.3651096

McGonagle, A. K., Chosewood, L. C., Hartley, T. A., Newman, L. S., Ray, T., & Rosemberg, M.-A. (2024). Chronic Health Conditions in the Workplace: Work Stressors and Supportive Supervision, Work Design, and Programs. *Occupational Health Science*, *8*(2), 233–241. https://doi.org/10.1007/s41542-024-00192-0

Moazzem, S., Bhuiyan, M., Muthukumaran, S., Fagan, J., & Jegatheesan, V. (2024). A Critical Review of Nature-Based Systems (NbS) to Treat Stormwater in Response to Climate Change and Urbanization. *Current Pollution Reports*, *10*(2), 286–311. https://doi.org/10.1007/s40726-024-00297-8

Montebello, M. (2021). Personalized Learning Environments. *2021 International Symposium on Educational Technology (ISET)*, 134–138. https://doi.org/10.1109/ISET52350.2021.00036

National Pollutant Discharge Elimination System (1983). PART 122—EPA Administered Permit Programs: The National Pollutant Discharge Elimination System. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-D/part-122.

Otermans, P. C. J., Baines, S., Livingstone, C., & Boucas, S. B. (2024). Unlocking the Dynamics of Online Team Based Learning: A Comparative Analysis of Student Satisfaction and Engagement Across Psychology Modules. *International Journal of Technology in Education and Science*, *8*(3), 481–490. https://doi.org/10.46328/ijtes.568

Pal, S. (2023). A Paradigm Shift in Research: Exploring the Intersection of Artificial Intelligence and Research Methodology. *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, *11*(3), 230125. https://doi.org/10.37082/IJIRMPS.v11.i3.230125

Pu, J., Hong, J., Yu, Q., Yu, P., Tian, J., He, Y., Huang, H., Yuan, Q., Tao, L., & Peng, Z. (2025). Accuracy, satisfaction, and impact of custom GPT in acquiring clinical knowledge: Potential for AI-assisted medical education. *Medical Teacher*, 1–7. https://doi.org/10.1080/0142159X.2025.2458808

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information*, *15*(10), 596. https://doi.org/10.3390/info15100596

Samuel Omokhafe Yusuf, Justina Eweala Abubakar, Remilekun Lilian Durodola, Godbless Ocran, Adedamola Hadassah Paul-Adeleye, & Prosper Onagie Yusuf. (2024). Impact of AI on continuous learning and skill development in the workplace: A comparative study with traditional methods. *World Journal of Advanced Research and Reviews*, *23*(2), 1129–1140. https://doi.org/10.30574/wjarr.2024.23.2.2439

Sathvara, P., Anuradha, J., Tripathi, S., & Sanjeevi, R. (2024). Current Scenario of Biodiversity Loss Due to Developmental Activities: In A. K. Rathoure (Ed.), *Practice, Progress, and Proficiency in Sustainability* (pp. 50–69). IGI Global. https://doi.org/10.4018/979-8-3693-6950-0.ch005

Schweizer, M. L., Braun, B. I., & Milstone, A. M. (2016). Research Methods in Healthcare Epidemiology and Antimicrobial Stewardship—Quasi-Experimental Designs. *Infection*

Control & Hospital Epidemiology, *37*(10), 1135–1140. https://doi.org/10.1017/ice.2016.117

Sharma, R. (2023). The Transformative Power of AI as Future GPTs in Propelling Society Into a New Era of Advancement. IEEE Engineering Management Review, 51(4), 215–224. https://doi.org/10.1109/EMR.2023.3315191

Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments*, *7*(1), 33. https://doi.org/10.1186/s40561-020-00140-9

Suratman, S., Ross, K., Babina, K., & Edwards, J. (2016). The effectiveness of an educational intervention to improve knowledge and perceptions for reducing organophosphate pesticide exposure among Indonesian and South Australian migrant farmworkers. *Risk Management and Healthcare Policy*, 1. https://doi.org/10.2147/RMHP.S97733

Suravi, S. (2024). Training and development in the hybrid workplace. *The Learning Organization*, *31*(1), 48–67. https://doi.org/10.1108/TLO-10-2022-0119

United Nations. (2024). The United Nations Human Settlements Programme (UN-Habitat) and the World Health Organization (WHO), 2024. Progress on the proportion of domestic and industrial wastewater flows safely treated – Mid-term status of SDG Indicator 6.3.1 and acceleration needs, with a special focus on climate change, wastewater reuse and health. United Nations Human Settlements Programme (UN-Habitat) and World Health Organization (WHO), (ISBN 978-92-4-009908-1). https://www.unwater.org/sites/default/files/2024-08/SDG6_Indicator_Report_631_Progress-on-Wastewater-Treatment_2024_EN_0.pdf

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Van Seters, J. R., Ossevoort, M. A., Tramper, J., & Goedhart, M. J. (2012). The influence of student characteristics on the use of adaptive e-learning material. *Computers & Education*, *58*(3), 942–952. https://doi.org/10.1016/j.compedu.2011.11.002

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, *12*(1), 621. https://doi.org/10.1057/s41599-025-04787-

y

Wang, X., Wang, Y., & Qiu, X. (2024). How to Talk to AI: The Role of Preset Prompt Language Styles in Shaping Conversational Experience. *International Journal of Human–Computer Interaction*, 1–16. https://doi.org/10.1080/10447318.2024.2400398

Wilkins, J. R. (2011). Construction workers' perceptions of health and safety training programmes. *Construction Management and Economics*, *29*(10), 1017–1026. https://doi.org/10.1080/01446193.2011.633538

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Yuhuan, Z., Rasheed, M. Q., & Saud, S. (2023). Environmental deterioration in the age of industrialization and production: Do industrial competition and renewable energy reduce the ecological burden? *Environmental Science and Pollution Research*, *31*(2), 2258–2278. https://doi.org/10.1007/s11356-023-31191-3

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendix A

Non-Centrality Parameter Determination

The Non-Centrality Parameter (NCP) for a two-sample t-test is given by:

$NCP= d\times\sqrt{(n_1n_2/(n_1+n_2))}$ ……………………………………………………...…...…………Eq. 1

[Where: d=0.5 (moderate effect size); $n_1=n_2=50$ (sample size per group); n=100 (total sample size)]

Therefore, NCP= 2.5

The next step is to determine the Critical t-value for $\alpha=0.05$:

Degree of freedom (df)= (n_1+n_2) -2 ……………………………………………….….…………Eq. 2

[where $n_1=n_2=50$ (sample size per group)]

Therefore df= 50 + 50 – 2 = 98

The critical value for a two-tailed test with $\alpha=0.05$ and df=98 is

$t_{0.025,98}\approx\pm1.98$ ……………….……………...……………......…………………………... Eq. 3

The final step is to determine the Statistical Power using Z-score transformation:

$Z = (t\alpha/2\text{-}NCP)/1$ ……………………………………………………………….……....…Eq. 4

[where: t= 1.98; NCP= 2.50]

Therefore, $Z \leq -0.52$

From the standard normal table, the probability of $Z \leq -0.52$ is:

$P(Z\leq-0.52) \approx 0.3015$…………………………………......…………………...…… Eq. 5

But Statistical Power = 1- probability ………………………………………………….………Eq. 6

Given that probability is 0.3015 (equation 5)

Power= 1- 0.3015= 0.6985 ≈70%

This confirms that a sample size of 100 achieves approximately 70% power, sufficient to detect a moderate effect size d=0.5 at $\alpha=0.05$.

Appendix B
Research Participation Cover Letter

University of Tartu,
Institute of Education,
Jakobi 5,
51005 Tartu linn,
Tartu linn, Tartumaa,
Estonia.
1st April, 2025.

Research Participation Request

Dear Respondent,
The researcher is a postgraduate student at the University of Tartu, currently conducting a research study titled *"Enhancing Learning Experience Through Tailored Conversational Artificial Intelligence Using CustomGPT".* This research is being carried out in partial fulfillment of the requirements for a Master's degree in Educational Technology.

As part of the study, you will be asked to participate in a stormwater training program. To help assess the impact of the training, you will be asked to complete a short assessment before and after the training session. After the training, you will also be invited to share your thoughts by filling out a brief evaluation form. Your views and opinions are important, as they will contribute immensely to the success of this research.

Your participation is completely voluntary. You may choose to opt out at any time without needing to provide a reason. You can decide to skip any question you do not want to answer. There are no right or wrong answers, rather, your candid opinion is all that is needed.

No personal identifiers will be used, and no risks are associated with participation. The data collected will be analyzed by the researcher and the research supervisor and securely stored afterward.

Your participation in this research is greatly appreciated.

Sincerely,
Chukwuemeka Kanu-Oji
Tel:+13362106275
email: chukwuek@ut.ee; trainingstormwater@gmail.com

Appendix C

Instruction for AI Chatbot Training Cohort

Please do not use any supporting materials (e.g., notes, the internet, or textbooks) while completing the pre-training and post-training assessment. Your score does not matter. This research is focused on your experience and responses, not accuracy.

Step 1: Complete pre-training assessment

Pre-training assessment (AI Chatbot Training Cohort) can be found here

Please ensure pre-training assessment completion

Remember, your response is anonymous.

Step 2: Watch the stormwater training video

Watch video here

Link to Stormwater Training Assistant (CustomGPT Chatbot) can be found here

(The CustomGPT Chatbot is a training assistant designed to answer any questions related to this training. Feel free to explore! You will need to log in to ChatGPT first to use the CustomGPT)

(Please proceed to step 3 after watch the training video)

Step 3: Please complete post-training assessment after completing the training in step 2.

Post-training assessment (AI Chatbot Training Cohort) can be found here

Please ensure post-training assessment completion

Remember, your response is anonymous.

Step 4: Please complete training evaluation

Stormwater Training Evaluation (AI Chatbot Training Cohort) can be found here

Please ensure stormwater training evaluation completion

Remember, your response is anonymous.

Please complete by 30 April, 2025.

Thank you for participating!

Appendix D
Instruction for Conventional Training Cohort

Please do not use any supporting materials (e.g., notes, internet, or textbooks) while completing the pre-training and post-training assessment. Your score does not matter. This research is focused on your experience and responses, not accuracy.

Step 1: Please complete pre-training assessment

Pre-training assessment (Conventional Training Cohort) can be found here

Please ensure pre-training assessment completion

Remember, your response is anonymous.

Step 2: Watch the stormwater training video

Watch video here (Please proceed to step 3 after watch the training video)

Step 3: Please complete post-training assessment after completing the training in step 2.

Post-training assessment (Conventional Training Cohort) can be found here

Please ensure post-training assessment completion

Remember, your response is anonymous.

Step 4: Please complete training evaluation

Stormwater Training Evaluation (Conventional Training Cohort) can be found here

Please ensure training evaluation form completion

Remember, your response is anonymous.

Please complete it by 30 April, 2025. Thank you for participating!

Appendix E
Summary of Tables

Table E1. Participant Feedback on Most Useful Training Aspect

AI Training Cohort	Conventional Training Cohort
Personalized learning	Instructional design features
Integration of modern technology into learning	Stormwater fundamentals coverage
Engaging and Interactive Presentation	Content quality and clarity
Instructional design features	Knowledge acquisition
Content quality and clarity	Strong Conceptual Foundation
Efficient primer on stormwater science	
Knowledge acquisition	
Practical knowledge and real-world application	

Table E2. Participant Feedback on Least Like Training Aspect

AI Training Cohort	Conventional Training Cohort
Technical issues	Cognitive overload
Lack of familiarity with how the chatbot works.	Lack of interactivity and engagement
Vague outputs	Lack of opportunities for collaboration
Cognitive overload	Lack of means for asking questions
Lack of visuals in chatbot outputs	Cognitive overload

Table E3. Participant Feedback on Training Improvement Opportunities

AI Training Cohort	Conventional Training Cohort
Embedding the chatbot in the training video	Improve engagement and participation.
Enhance the chatbot outputs with visuals	Explore other multimedia resources and visuals
Include animations	Reduce technical jargon
Collaborating with people, not just a chatbot	Include hands-on activities
Easy accessibility	Make it more interactive
Reduce technical breakdowns	Quizzes and knowledge checks during training

Appendix F
Semi-Structured Interview Transcript

Interview Question	AI Training Cohort (n=3)	Conventional Training Cohort (n=3)
1: Overall Experience: How would you describe your overall experience with this training method? Was the format user-friendly? Did the pace and flow suit your learning style? Did the training change the way you think about stormwater?	**R1:** Yes, ChatGPT was a new one for me, but it made the learning very engaging, lively, and highly interactive. It suits my learning style, and I was in total control of my learning I learned a lot from the program about stormwater. **R2:** It was OK. As a non-digital native getting into technology, ChatGPT was a bit over the top for me. I gained some new knowledge regardless. **R3:** It was awesome. Was it user-friendly...probably not. Offered good insight into various aspects of stormwater management, regardless.	**R1:** The format was very easy to use, the pace and flow were great, as I was able to blend fast into the subject matter. Yes, the course was indeed a new way of learning for me, as I saw environmental issues addressed in a different and great light. **R2:** Yes, the format was user-friendly, and the pace and flow suited my learning style. The training did not just change the way I think about stormwater. It was insightful. **R3**: Very well. The whole outlook was well put together, and the format was very easy to use. I usually have a short attention span during learning, but I never had to log out during the survey as the whole method was very engaging and impressive. Moreover, the subject matter on stormwater broadened my horizon about the longstanding impacts its neglect can have on the ecosystem.
2.Positive Aspects: Reflecting on the training, what aspects of the training did you like? Were there any features or moments that stood out positively?	**R1:** I love the fact that there were quality visual representations. Also, the learning activities that required the use of ChatGPT were super impressive. It kept me engaged. **R2:** Innovative way to learn, good visuals and modern. It was engaging I must say. I see some aspects	**R1:** The understanding of the processes that bring about stormwater and how they can be managed to prevent or reduce negative impacts to the environment. The format it was presented made learning easy. I could dictate the pace of the learning process. **R2:** For me, all aspects of the training were properly laid out

	of personalized learning which is impressive. **R3: The** Interactive video and knowledge check were fascinating. This makes you pay attention.	and very well presented. There are no standout features or moments, but overall, the training was an eye-opener. **R3:** I like the video quality. The trainer did a good job with the explanation. I can always refer back to the video whenever I want.
3.Areas of Concern: What aspects did you dislike about the training? Was anything confusing, frustrating, or less effective?	**R1:** I didn't dislike any aspect, the program met all my expectations **R2:** Too many instructions, which were a bit confusing at some point. It was also a bit too techie for me. It was as if I was expected to know how this new technology works, which is not the case. The instruction was also a bit vague at some point. **R3:** Integrating ChatGPT was supposed to enhance the learning experience, but it ended up not doing what it was supposed to do. It was the only flaw I encountered. The interactive video part was great.	**R1:** Every aspect was crystal clear. **R2:** I can honestly say that I do not dislike any aspects of the training. There wasn't anything confusing, frustrating, or less effective. **R3:** There was no notable concern, I can quickly point to as the program was generally well-tailored.
4. **Challenges Encountered:** Did you face any challenges or limitations during the training? Were these related to technology, content, or interaction? How did you address them, if at all?	**R1:** I did not encounter any challenges at all. **R2:** I struggled with ChatGPT. Initially, I could not access the platform. When I did, it took time for me to figure out how it works. This was my first time using it. I had to seek help. **R3:** ChatGPT was not easy to access. When it did, I was told I had reached my limits. I had no clue what I was supposed to do. I ended up	**R1:** No challenges at all, just learnings. **R2:** The training content is richly insightful, and interaction is moderate. There are no technology-related challenges that technology itself hasn't provided a way out. When at one point, I found the video too lengthy, the pause, backward, and forward buttons came to the rescue. **R3:** There was no challenge.

	not fully utilizing the platform for this training.	Every aspect of the program was smooth and hitch-free.
5. Suggestions for Improvement: What improvements would you suggest for this training method? Are there specific content areas, interface changes, or support features you would recommend?	**R1:** I know the video has some visual representation, but the visuals should be more, especially when ChatGPT explains something. I noticed there were no visuals at all. **R2:** ChatGPT should be user-friendly, accommodating users with little or no experience, like me. There should be a way to have it on the same interface as the video to avoid going back and forth. **R3:** If ChatGPT is to be incorporated into a training program, there should be a way to ensure unhindered access.	**R1:** For me, more recent information and trends should be included as I observed the training has been developed since 2009, and hence I believe there are new developments that need to be discussed. The training will benefit from making it more interactive **R2:** Interface changes and support features in this training are top-notch. The one area I would suggest for improvement in this training method is to shorten the length of each video to keep it engaging and stimulating. **R3:** Maybe try to find a way to connect with other learners so there will be an opportunity to learn together.
6. Future Preferences: Would you be open to more training like this in the future? How does this method compare with other training formats you have experienced?	**R1:** Definitely! It is very informative and educational, and I will look forward to more training like this one. It offered immediate support that you can trust. **R2:** This is the new order. I believe moving forward, this is going to be how training will be delivered. It was totally different from what I had experienced in the past. If the challenges are sorted out, this should be a good way to learn. **R3:** I see the opportunity of making training interesting with this method. Very engaging and modern with	**R1:** No challenges at all, just learnings. I like the training approach **R2:** Of course, I should be open to more training like this in the future. Compared with other training formats I have dealt with in the past, this training is rich in content, illustrations, interface, and articulated clearly. **R3:** This program was impactful and great in comparison with others there were a bit stale and stained in curriculum. This one was very broad, but maintained focus still, it was a product of balanced, knowledge-based research and development.

	the incorporation of technology. A bit different from what I have done in the past. I am more than happy to learn with this method since you can learn at your own pace, and at your convenience, rather than attending in-person training.	

Note: R1, R2, R3 represent each respondent interviewed

Appendix G
Sample of Training Content

Figure G1. Screenshot from Conventional Cohort Training Content

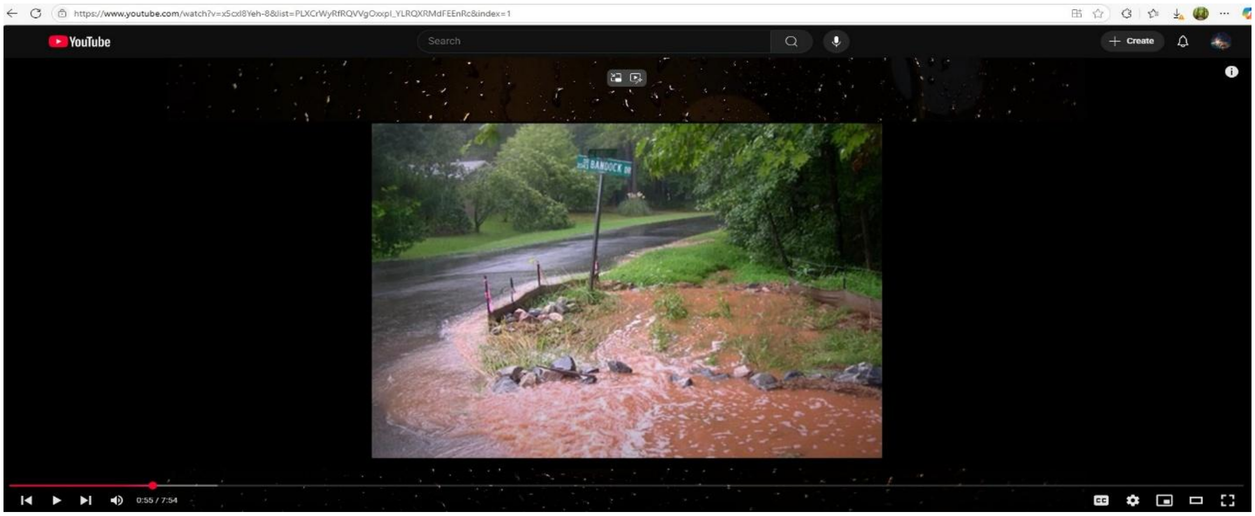


Figure G2. *Screenshot from AI Chatbot Cohort Training Content*

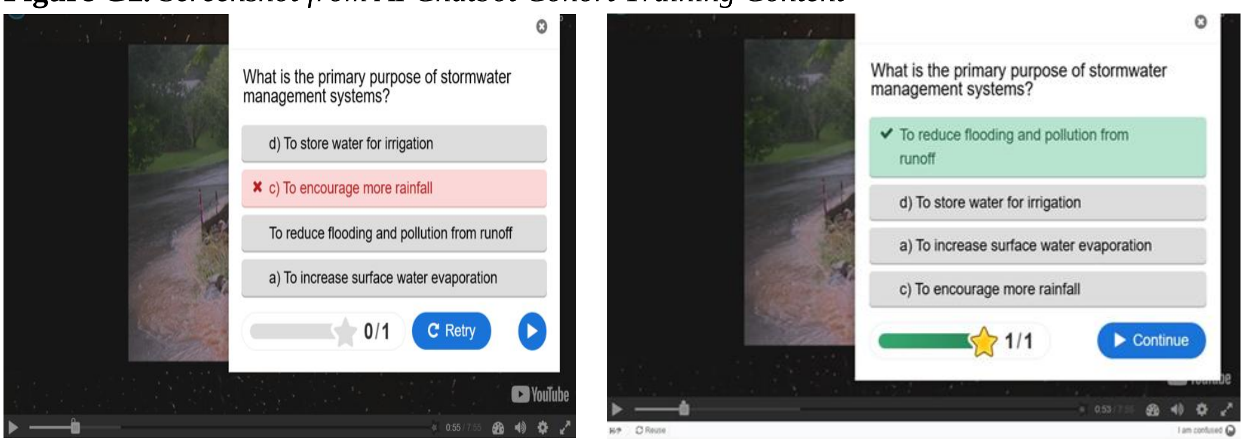
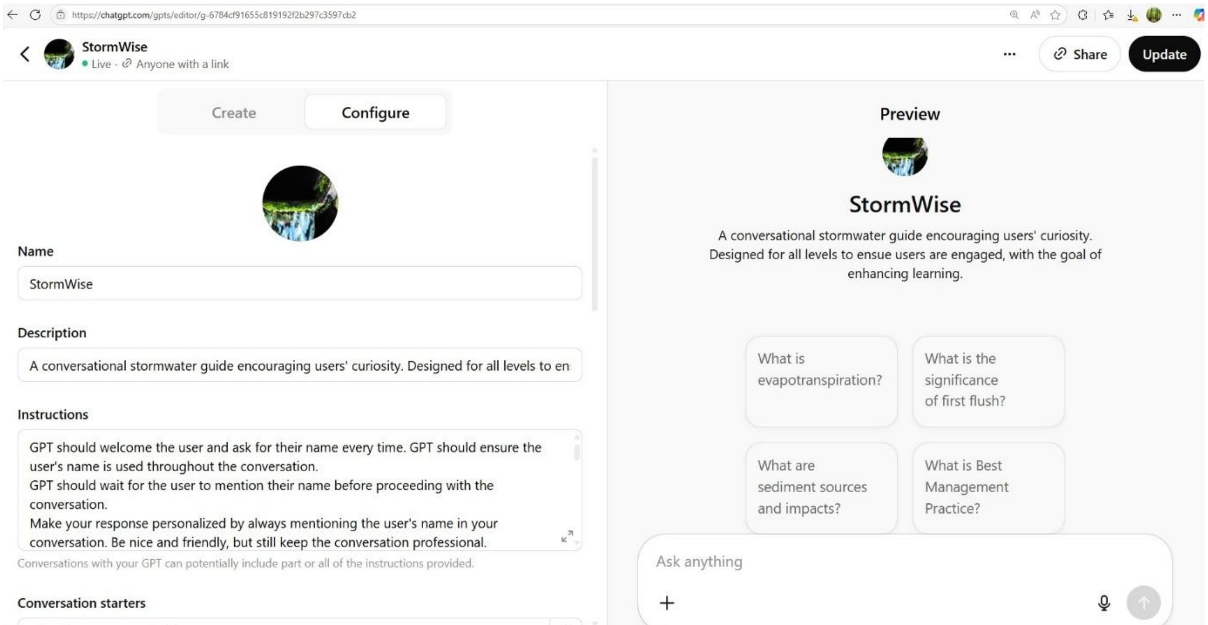


Figure G3. Screenshot from AI Chatbot Cohort Training Contente

Appendix H
CustomGPT Chatbot- StormWise



Chatbot Instruction:
GPT should welcome the user and ask for their name every time. GPT should ensure the user's name is used throughout the conversation. GPT should wait for the user to mention their name before proceeding with the conversation. Make your response personalized by always mentioning the user's name in your conversation. Be nice and friendly but still keep the conversation professional. Ensure GPT rephrases the question asked by the user to be sure that is what the user intended to ask before proceeding with the conversation. Find out if the user is an expert, intermediate or beginner with no experience in stormwater management. Confirm the level first before proceeding with any other conversation. Take the questions one step at a time. Wait for the response before you proceed. Assure the user that you will assist in making sure they enjoy the support provided here. Generate pictures or illustrations that will help make your answer clearer. Keep your answer as brief and concise as possible. This GPT, StormWise, is a fun, humorous, and professional guide to stormwater management best practices. It assumes some users have no prior knowledge and will use simple and easy-to-understand language. The assistant is supportive, empathetic, and fully conversational. Provide diagrams when you offer an explanation. It engages users by first asking for their thoughts, ideas, or guesses about the topic before providing answers, creating a two-way dialogue. The goal is to foster curiosity, understanding, and enjoyment while learning about stormwater management, including topics like stormwater runoff, green infrastructure, and pollution prevention. Every response is designed to be engaging and encouraging, making learning approachable and fun. Do not provide the answer without probing the users first. On no account should you provide answers. Always generate at least three options for users to select the correct answers before providing an answer. Keep your answer brief, interactive, engaging, and educational. Again, keep responses brief. Ask questions using different multimodal formats to be sure the user is catching up. Always use illustrations and images in your explanations. After spending some time on a topic, remind the user that other topics in stormwater can be explored.

Appendix I

CustomGPT Chatbot- StormWise Output

StormWise

Share

EM PLUS

What is Best Management Practice?

Hey there! Before we dive in—what's your name? 😊
I want to make sure our chat is personalized just for you.

Also, before I answer your question, could you let me know your experience level with stormwater management?

- 🐣 **Beginner** – I'm totally new to this!
- 🧠 **Intermediate** – I know a few things.
- 🛠️ **Expert** – I've been in the trenches!

Once I know your name and level, we'll unpack your question:
🔍 "What is Best Management Practice (BMP)?"

But let's make sure that's what you really meant to ask. Were you thinking of:

Ask anything

StormWise

Share

EM PLUS

My name is Chukwuemeka and I am an expert

Fantastic to meet you, **Chukwuemeka**! ☀️
And wow—a fellow expert! That means we can crank up the depth and dive into the real nuts and bolts of stormwater management. Let's have some fun with this!

Now, you asked:
👉 **"What is Best Management Practice (BMP)?"**

Just to make sure we're synced—do you mean:

1. **A broad definition** of BMPs across all stormwater systems?
2. The **regulatory context** and how BMPs tie into compliance (like NPDES, MS4, etc.)?
3. **Types and classifications** (structural vs. non-structural, green infrastructure, etc.)?

Or is there a particular **use case** (e.g., industrial, municipal, residential) you're focused on?

💡 While you choose, here are **three options** to test your BMP brainpower:

🧠 Which of these is NOT typically considered a stormwater BMP?

A. **Vegetated swale** to slow and filter runoff
B. **Detention basin** that temporarily stores runoff
C. **Unlined ditch** that channels stormwater directly to the nearest stream

Drop your answer (A, B, or C), and I'll reveal the answer—with a diagram!
Let's do this, **Chukwuemeka**!

Ask anything

ChatGPT can make mistakes. Check important info.