International Journal of Doctoral Studies

An Official Publication of the Informing Science Institute
InformingScience.org

IJDS.org

Volume 20, 2025

CHATGPT IN DOCTORAL SUPERVISION: PROPOSING A TRIPARTITE MENTORING MODEL FOR AI-ASSISTED ACADEMIC GUIDANCE

Omiros Iatrellis*	University of Thessaly, Larissa, Greece	iatrellis@uth.gr
Areti Bania	University of Thessaly, Larissa, Greece	aretibania@uth.gr
Nicholas Samaras	University of Thessaly, Larissa, Greece	nsamaras@uth.gr
Ioanna Kosmopoulou	University of Thessaly, Larissa, Greece	iokosmop@uth.gr
Theodor Panagiotakopoulos	University of Patras, Patras, Greece University of Nicosia, Nicosia, Cyprus	tpanagiotakop@upatras.gr

* Corresponding author

ABSTRACT

Aim/Purpose	The potential of Generative AI in education is expanding, yet its role in PhD mentoring and academic guidance remains underexplored. This study evaluates how ChatGPT-generated recommendations can support PhD research, particularly in fostering sustainable and resource-efficient doctoral education.
Background	This study examines the ability of ChatGPT to provide structured guidance and actionable insights in PhD supervision. Using a real-world case study on disaster risk management, the research evaluates AI-generated recommendations across different prompt structures to determine their relevance, depth, and applicability to doctoral research.
Methodology	A structured evaluation was conducted using input prompts with varying contextual details, including naive prompts, supervisor-selected keywords, ChatGPT-generated keywords, and topic-specific concepts. Five external academic experts assessed the AI-generated outputs for appropriateness, interrater agreement, and research alignment.
Contribution	The study demonstrates that ChatGPT can enhance PhD supervision by providing structured academic recommendations, reducing administrative bur-

Accepting Editor Erin Breitenbach | Received: March 5, 2025 | Revised: June 17, June 23, 2025 | Accepted: June 24, 2025.
Cite as: Iatrellis, O., Bania, A., Samaras, N., Kosmopoulou, I., & Panagiotakopoulos, T. (2025). ChatGPT in doctoral supervision: Proposing a tripartite mentoring model for AI-assisted academic guidance. *International Journal of Doctoral Studies, 20,* Article 9. https://doi.org/10.28945/5579

	dens on supervisors, and contributing to the evolution of a "tripartite mentoring model" where AI, supervisors, and students collaborate to tackle complex research challenges.
Findings	AI-generated recommendations were most effective when structured around topic-specific concepts. Naive prompts also yielded relevant outputs, whereas keyword-based prompts resulted in less cohesive recommendations. Tailored prompts aligning with specific research pathways were rated as highly actionable and contextually grounded. ChatGPT demonstrated the ability to refine research methodologies and improve knowledge discovery.
Recommendations for Practitioners	Universities may consider incorporating AI tools such as ChatGPT to support PhD supervision, particularly to provide structured feedback and guidance. Supervisors should explore AI-assisted mentoring to optimize time-intensive advisory tasks and enhance research productivity.
Recommendation for Researchers	Researchers should explore the effectiveness of AI-driven academic guidance across various disciplines, assessing its impact on research quality, methodological rigor, and doctoral student development. Future studies may also investigate the ethical considerations of AI in PhD supervision, including potential biases in AI-generated recommendations and risks related to over-reliance on automated feedback.
Impact on Society	By reducing supervisory workload and enhancing research efficiency, AI-driven academic guidance can promote equitable access to high-quality doctoral education, fostering innovation and sustainable educational practices globally.
Future Research	Future research should evaluate AI-driven mentoring across multiple academic disciplines and multilingual contexts to better assess generalizability. Additionally, studies should explore ethical implications, including how disciplinary norms, cultural expectations, and linguistic diversity influence the effectiveness and appropriateness of AI in doctoral supervision.

| Keywords | generative AI, PhD mentoring, educational technology, academic guidance |

INTRODUCTION

Pursuing a PhD represents one of the most challenging academic journeys, requiring deep subject-matter expertise alongside advanced skills in research design, critical thinking, data analysis, and academic writing (Higgins & Thomas, 2016; Mian et al., 2020). This process is also highly resource-intensive, demanding significant time and effort from both students, who must navigate complex academic expectations, and supervisors, who provide continuous guidance and support (Su & Yang, 2023). These challenges underscore the relevance of innovative, technology-driven approaches that may optimize supervision, enhance efficiency, and support PhD students in overcoming research challenges (Kouatli, 2020).

With the emergence of generative AI (GenAI) models such as ChatGPT, new possibilities have opened for supporting academic tasks across various domains (Chan et al., 2019). In education, GenAI has shown promise in areas like personalized learning, content creation, and skill development (Cardona & Cudney, 2019). Additionally, GenAI's ability to revolutionize educational methodologies by fostering sustainable practices and enhancing accessibility makes it a valuable tool in addressing the environmental and social dimensions of sustainability in education. However, its role in PhD supervision and research mentoring remains relatively underexplored, particularly its ability to enhance knowledge discovery, refine research methodologies, and provide structured academic guidance (Iatrellis et al., 2020). The key challenge lies in determining how contemporary AI models can

be effectively leveraged to understand nuanced academic contexts, generate original insights, and uphold the rigorous standards of scholarly work. This raises important questions about whether AI-generated recommendations can align with PhD students' research trajectories and how they compare to traditional mentoring approaches.

In this paper, we explore the application of ChatGPT in the domain of PhD academic counseling and support. We propose a structured evaluation to analyze how AI can assist PhD students and supervisors in navigating the multifaceted demands of academic research. By leveraging the strengths of ChatGPT, we aim to identify key areas where AI can provide value to PhD studies, such as offering recommendations on advanced methodological frameworks, in addition to the support ChatGPT has already demonstrated in tasks such as data analysis, academic writing, and literature review (Miller et al., 2015). By integrating expert evaluations, iterative prompt design, and systematic validation methods, this study offers a structured approach to assessing the feasibility and reliability of ChatGPT-generated academic guidance. This exploration seeks to highlight both the opportunities and challenges of integrating AI into the PhD research process, offering a roadmap for its effective utilization in third-cycle programs. To achieve this, we evaluated the relevance and accuracy of ChatGPT's outputs in providing academic guidance for a PhD proposal focused on disaster risk management. The results highlight the potential of ChatGPT to support complex academic endeavors by addressing key dimensions of disaster management research, offering insights into its broader applicability in higher education and research domains.

BACKGROUND

DOCTORAL SUPERVISION MODELS AND CHALLENGES

Global perspectives on doctoral supervision highlight a range of conceptual frameworks and shared challenges. Classic models often follow a master-apprentice or cognitive apprenticeship approach, wherein an expert supervisor gradually reduces guidance as the student gains skills and independence (Polkinghorne et al., 2023). Complementing this is a dialogic view of supervision that treats research training as a collaborative, two-way process, with knowledge and feedback co-constructed through open dialogue between student and supervisor. Despite these pedagogical approaches, empirical studies indicate persistent issues across contexts. Supervisors face heavy workloads and time constraints (spending on average only ~52 hours per student annually) (Polkinghorne et al., 2023), a demanding load alongside other academic duties, which in turn can limit the frequency and depth of feedback students receive. Indeed, research has documented supervision characterized by ad-hoc support and limited feedback often disconnected from students' actual research needs (Maher et al., 2013).

Additional pressures arise from the growing interdisciplinarity of doctoral research, which requires supervisors to navigate multiple domains of expertise and manage co-supervision arrangements, further complicating coordination and consistency (McCallin & Nayar, 2012). Furthermore, doctoral students often experience extended periods of independent work with limited guidance, contributing to feelings of isolation, delayed progress, and reduced confidence in research decision-making (Melián et al., 2023). As doctoral education expands globally, institutions face growing challenges in ensuring timely, consistent, and high-quality supervisory support, particularly as supervisor workloads continue to intensify and student cohorts become increasingly diverse in terms of disciplinary backgrounds, methodological needs, and expectations.

THE EMERGING ROLE OF GENERATIVE AI IN DOCTORAL RESEARCH SUPPORT

The advent of GenAI has catalyzed transformative changes across multiple domains, including education and academic research (Han, 2024). These tools have demonstrated significant potential in

streamlining time-intensive processes, enhancing productivity, and facilitating innovation. In academia, GenAI supports a variety of tasks, such as drafting and organizing research content, data analysis, and hypothesis generation, making it a valuable assistant for both students and educators (Maphosa & Maphosa, 2023; Moundridou et al., 2024).

Frameworks for integrating GenAI into educational settings emphasize iterative approaches, stakeholder engagement, and continuous improvement to balance innovation with ethical considerations. For instance, the 4E framework – embrace, enable, experiment, and exploit – offers a systematic model for aligning policy, curriculum, and technological adoption in academic institutions (Kuhail et al., 2023). Such structured approaches can help harness the benefits of GenAI while mitigating its limitations, ensuring that it complements rather than replaces human expertise.

One of the most prominent benefits of GenAI lies in its capacity to improve academic processes. (For more information see OpenAI at https://openai.com/) Studies highlight its role in language polishing, error identification, summarization, and even coding support, particularly benefiting researchers who deal with multidisciplinary or technical topics (Bahrini et al., 2023). GenAI tools also enable personalized learning and adaptive teaching by acting as learning partners or teaching assistants, optimizing resources and enhancing accessibility. However, these transformative capabilities come with significant challenges. Scholars have increasingly voiced concerns regarding AI hallucination instances where AI confidently provides incorrect or misleading information, as well as inherent biases in AI outputs that can reinforce stereotypes or skew perspectives (Bang et al., 2023; Huang et al., 2025). Moreover, the potential erosion of critical thinking skills among users, who might overly rely on automated recommendations without adequate skepticism, presents another critical issue that demands attention in the responsible integration of these technologies (Zhou et al., 2024).

In the context of higher education, GenAI has been recognized for its potential to address specific needs in PhD-level research and supervision. It can support tasks such as literature summarization, academic drafting, and initial ideation, and it can also enhance conceptual clarity, framework development, and methodological rigor, all of which are crucial for advancing academic research (Bahrini et al., 2023). Such contributions may be particularly valuable given the resource-intensive nature of PhD studies, which demand substantial commitments of time and expertise from both students and advisors.

Despite these advancements, the adoption of GenAI in higher education raises critical concerns. Issues including academic integrity, plagiarism, data privacy, and trustworthiness of AI-generated content remain prominent, requiring institutions to establish clear guidelines and responsible usage policies aligned with regulations such as GDPR (Anwar et al., 2024; Weidinger et al., 2022). Additionally, significant variability exists in scholars' familiarity and expertise with prompt engineering skills, influencing the effectiveness and reliability of AI-generated outputs (White et al., 2023). Disparities in access to GenAI tools and varying degrees of user proficiency further pose equity challenges, which must be proactively addressed through institutional training and support programs (Kuhail et al., 2023).

Supervisors and institutions are encouraged to adapt to the evolving role of GenAI by promoting responsible use while fostering independent critical thinking and intellectual rigor among students (Bahrini et al., 2023). While the current state of research underscores the transformative role of GenAI in academic processes such as literature review, data analysis, and writing, its integration into the relational processes of PhD mentoring remains underexplored. Much of the existing literature focuses on task-specific applications of GenAI but provides limited insights into how these tools can complement and enhance the mentoring relationship between PhD advisors and students. This mentoring process involves not only technical guidance but also fostering intellectual development, critical thinking, and navigating complex academic challenges – areas where human advisors have traditionally played a central and irreplaceable role.

To address this gap, this paper investigates the potential of GenAI to complement traditional PhD mentoring by providing targeted support, aiding conceptual development, and relieving specific burdens within the PhD journey. By critically assessing both the advantages and potential risks, this study seeks to extend current understanding of GenAI's role in higher education, proposing pathways for its ethical and effective integration into PhD advising practices while explicitly acknowledging the importance of maintaining robust human oversight and scholarly integrity.

METHODS

CASE SELECTION

This study was conducted using a real-world PhD case from the University of Thessaly, focusing on decision-support systems for first responders in disaster management. This topic provided a complex, multi-dimensional research context highly suitable for evaluating ChatGPT's ability to generate academic guidance across various disciplinary pathways and research design challenges. The PhD was part of the RES-Q Living Lab, addressing challenges of data integration, real-time validation, and predictive analytics to enhance situational awareness for emergency response.

During the initial phase of this research initiative, a PhD student and member of the RES-Q Living Lab led the effort by undertaking a doctoral project titled "Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration." We chose this study because different research pathways were proposed, making this case ideal for evaluating ChatGPT's ability to provide tailored academic guidance across multiple, complex scenarios. The diversity of research directions allowed for a comprehensive assessment of the model's capacity to adapt to varied academic challenges and offer relevant, domain-specific recommendations. As part of these efforts, the PhD student conducted a preliminary literature review to identify key gaps and challenges in the current disaster management systems. The review revealed the following critical issues that inhibit first responders' ability to operate effectively during emergencies:

- *Fragmentation of Data Sources:* Emergency response systems rely on both traditional data streams, such as environmental sensors and infrastructure feeds, and nontraditional sources, including social media and citizen-generated reports. However, these heterogeneous data streams often lack interoperability and integration, leading to incomplete situational awareness.
- *Reliability of Nontraditional Data:* Nontraditional data sources are frequently invalidated, resulting in potential misinformation and reduced trust in decision-making tools.
- *Inconsistent Data Formats and Lack of Adaptable Interfaces:* First responders face difficulties in processing and prioritizing information due to incompatible data formats and a lack of customizable, user-friendly interfaces.

These findings are corroborated by the International Forum to Advance First Responder Innovation (2018), which highlights these issues as Capability Gap 4, emphasizing the challenge of integrating diverse and nontraditional data sources into incident command operations. This alignment reinforces the significance of addressing these challenges to improve the speed and accuracy of decision-making during emergencies. Recognizing these challenges, the PhD student, together with the advisory committee, explored potential approaches to address the identified gaps. After thorough deliberation, three pathways were proposed as the next viable research direction:

- *Geospatial Intelligence and Remote Sensing:* This was proposed to enhance situational awareness and decision-making by integrating real-time spatial data from sources such as satellite imagery, drones, and Geographic Information Systems (GIS) (Ghosh, 2023; Hasanuzzaman et al., 2023). These technologies enable first responders to monitor disaster-affected areas, track resources, and visualize critical information through user-friendly interfaces. By addressing the fragmentation of data and providing actionable insights, this pathway aligns

with the PhD's goal of developing an advanced decision-support framework for emergency scenarios.

- *Digital Twins:* This technology offers a novel approach to real-time disaster simulation and training (Yu & He, 2022; Yun et al., 2022). By integrating live data streams into virtual replicas of real-world systems, Digital Twins enable first responders to visualize and interact with dynamic scenarios. This pathway focuses on improving situational awareness and decision-making through interactive, user-centric interfaces that simulate disaster evolution in real-time.
- *Semantic Web Approaches:* These approaches were identified as a promising solution to achieve advanced data integration (Bania et al., 2022, 2024). By leveraging ontologies, Resource Description Framework (RDF), and SPARQL Protocol and RDF Query Language, this pathway aims to harmonize heterogeneous data streams and enable seamless interoperability. These technologies would provide first responders with consistent, machine-readable datasets, facilitating real-time access to actionable insights.

Each of these pathways provides a unique perspective on addressing the abovementioned research challenges. Semantic web approaches focus on the integration of diverse data sources, geospatial intelligence, and remote sensing to enhance situational awareness through real-time spatial data analysis, and digital twins provide real-time simulation and visualization capabilities. These represent well-established approaches that, in the context of the PhD, must be examined independently to determine their distinct benefits, challenges, and feasibility.

RESEARCH DESIGN

Building on these findings, we designed a study, the overview of which is presented in Figure 1. The case study used for ChatGPT prompts was crafted as a composite representation of the key challenges in developing the appropriate recommendations for a specific PhD. The process follows a structured evaluation flow: PhD proposal summary → ChatGPT → Proposed Solutions/Support → Evaluation by External Academic Experts. This design ensures a systematic exploration of how generative AI can contribute to addressing these challenges within the scope of the PhD.

Practical limitations, such as the maximum input length and other constraints like ambiguity in instructions, formatting challenges, or context retention issues, guided the structuring of the case. The design ensures that the AI prompts are comprehensive yet concise, facilitating the generation of meaningful outputs while encouraging an academic tone aligned with scholarly discourse. This structured evaluation flow allows academic experts to rigorously assess the relevance, depth, and practical value of ChatGPT's outputs in addressing critical challenges in PhD-level DRM research.

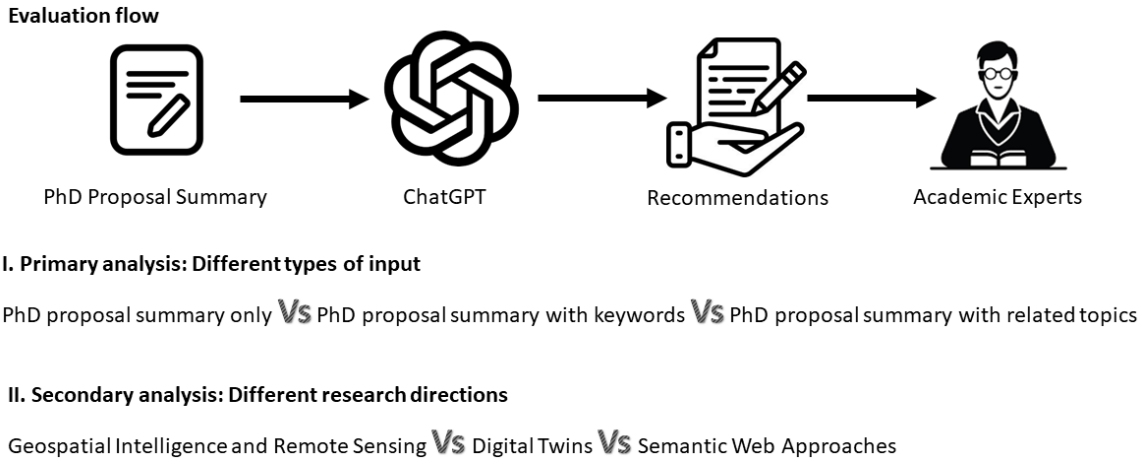

Figure 1. Study design overview

PRIMARY ANALYSIS

ChatGPT is designed to produce contextually relevant outputs based on input prompts. To evaluate its potential for providing academic counseling specifically for PhD studies, we composed input prompts reflecting varying levels of background detail as part of the primary analysis. Previous research indicates that the quality and specificity of AI outputs are heavily influenced by how prompts are structured (Cheng et al., 2023; Iatrellis et al., 2024). Therefore, our primary analysis assessed how ChatGPT's responses varied when different types of input, ranging from minimal to highly detailed, were provided.

To systematically evaluate ChatGPT's ability to generate relevant and actionable academic recommendations, we tested four variations of input prompts, each designed to explore how different levels of contextual detail influence the generated output:

- *Naive Prompt* - A basic case study description that included only a general summary of the PhD proposal.
- *Keyword-Based Prompts* - A PhD proposal summary supplemented with keywords, obtained in two ways:
  - *Supervisor-Selected Keywords:* Key terms manually chosen by PhD supervisors based on the proposal.
  - *ChatGPT-Generated Keywords:* Key terms suggested by ChatGPT based on the initial proposal.
- *Topic-Specific Prompt* - A PhD proposal summary incorporating research concepts derived from Structural Topic Modeling (STM), which analyzed peer-reviewed journal articles on digital technologies and disaster management (Kabra et al., 2024).

SECONDARY ANALYSIS

The outputs generated by ChatGPT for each prompt were evaluated by a panel of five external academic experts representing diverse disciplines. These evaluators assessed the responses for their appropriateness based on relevance, depth, and practical applicability. The percentage of responses rated as appropriate for each prompt type was calculated. Additionally, the evaluators ranked the outputs in order of quality, with criteria clearly outlined to ensure consistency. To minimize bias, evaluators were blinded to the type of prompt used to generate each output.

In addition to analyzing prompt details, the study examined ChatGPT's capability to provide support tailored to specific academic approaches. Building on the three key research directions identified in the Case Selection section - Geospatial Intelligence and Remote Sensing, Digital Twin, and Semantic Web Approaches - prompts were carefully designed to guide ChatGPT in offering targeted recommendations aligned with these pathways.

For each pathway, the prompts instructed ChatGPT to propose recommendations on how the student should approach the research, detailing suggested methods, key steps, and potential challenges. The outputs were then evaluated by academic experts, who assessed their alignment with best practices within each area.

EXPERT EVALUATION PROCESS

To ensure a rigorous and unbiased assessment of ChatGPT-generated recommendations, we engaged five external academic experts with specialized knowledge in AI, in education, PhD supervision, disaster risk management, and data-driven decision-making. The selection process prioritized diversity in expertise to capture a broad evaluation perspective, ensuring that the assessment reflected both AI-driven methodologies and domain-specific research considerations. The experts were selected based on their established academic credentials and prior contributions to research and doctoral supervision in their respective fields. They included:

- 2 senior academics specializing in AI for education and learning analytics, ensuring expertise in evaluating generative AI's role in research guidance.
- 1 professor of disaster risk management, ensuring alignment between ChatGPT-generated outputs and established DRM research frameworks.
- 1 expert in data science and predictive analytics, assessing the integration of AI-driven methodologies in doctoral research.
- 1 expert in academic writing and research methodologies, evaluating the coherence, structure, and scholarly rigor of ChatGPT's recommendations.

The selected experts were not directly involved in the PhD project under evaluation, nor affiliated with the university or supervisory team connected to the study. This independence was crucial in mitigating the risk of bias. Additionally, the evaluation was conducted through a blind review process, meaning that the experts were not informed about the specific input prompts used to generate each output. This ensured that their judgments were based solely on the content quality and academic appropriateness of ChatGPT's recommendations, rather than any preconceived expectations about AI-generated text.

The external academic experts evaluated the ChatGPT-generated outputs using a structured 5-point Likert scale assessing academic appropriateness and recommendation quality. To promote consistency across expert evaluations, raters were also guided to consider multiple evaluation dimensions: (a) factual accuracy, (b) academic depth, (c) integration with the specific PhD research context, (d) demonstration of critical thinking, (e) methodological rigor, and (f) practical applicability. While the evaluation remained based on expert judgment, these dimensions align with recent recommendations in AI-assisted academic advising literature (White et al., 2023) and may serve as the foundation for future rubric-based evaluation frameworks. The full wording of the 5-point Likert scale was as follows: 1 = Completely Inappropriate, 2 = Somewhat Inappropriate, 3 = Neutral, 4 = Appropriate, 5 = Completely Appropriate and Actionable for Doctoral Research.

Assessing Rater Consensus for AI-generated Outputs

To evaluate the consistency of expert assessments, interrater agreement was calculated using Kendall's W, a statistical measure appropriate for ordinal, non-parametric data. This analysis was conducted in Python, utilizing the pingouin and scipy libraries for statistical computation. All p-values less than 0.05 were considered statistically significant.

The prompts were input into ChatGPT (version 4.0) by a single researcher to ensure consistency in execution. Each prompt was entered into a new session to eliminate potential carryover effects from prior interactions. Outputs were systematically collected and labeled according to the input type and academic support area. To ensure comparability, the prompts were standardized in terms of wording and structure, with variations introduced only in the level of detail or instructional framework.

Results

Overview of Input and Output Characteristics

The PhD proposal analyzed in this study, *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration,* focuses on developing a decision-support framework for first responders in disaster management. This framework, driven by the student's literature review findings, aims to address key challenges such as the fragmentation of diverse data sources, the lack of real-time validation mechanisms, and the need for predictive tools to enhance situational awareness and decision-making. By integrating research paradigms and predictive analytics, the research seeks to improve the reliability of information and enable proactive responses during emergencies.

Using the full PhD proposal and contents shown in Table 1, prompts with four different levels of detail were generated. The PhD proposal provided a high-level overview without additional context.

Keyword-enhanced prompts incorporated terms selected by the PhD supervisors and ChatGPT. Topic-specific prompts included additional relevant research topics drawn from topic modeling. Finally, prompts were designed around the three research pathways identified: Geospatial Intelligence and Remote Sensing, Digital Twin Technologies, and Semantic Web Approaches, each offering a focused approach to the research challenges.

In the primary analysis, when these prompts were presented to ChatGPT, the longest output was produced in response to the Topic-specific prompts (Table 2). Conversely, the naive prompt resulted in the shortest output (general summary: 459 words, keywords by supervisors: 536 words, keywords by ChatGPT: 572 words, topic-specific: 605 words, comprehensive: 410 words). When the prompts aligned with specific research directions, such as geospatial intelligence, digital twins, or semantic web approaches, the length of the outputs varied slightly (geospatial intelligence: 404 words, digital twins: 410 words, semantic web approaches: 652 words). Overall, prompts incorporating topic modeling or domain-specific research pathways tended to generate longer outputs, likely reflecting the richer contextual information provided. However, output length was not directly associated with expert-rated appropriateness, indicating that content quality depends more on the specificity and structure of the prompt rather than on the verbosity of the response.

Appropriateness Across Input Detail Levels (Primary Analysis)

In assessing appropriateness, the responses generated from the prompt incorporating concepts from topic modeling and the naive prompt were deemed suitable by all evaluators (Table 3). However, this was not the case for the two keyword-enhanced prompts, where the appropriateness ratings were lower (keywords by supervisors: 80%, keywords by ChatGPT: 40%) (Figure 2). Among the rankings, the prompt utilizing concepts derived from topic modeling consistently received the highest scores, with four out of five experts ranking it first and one expert placing it second. The naive prompt performed well, generally ranking higher than both the supervisor- and ChatGPT-generated keyword prompts. Specifically, the naive prompt was rated second or third by most experts, demonstrating its relative strength compared to the keyword-enhanced options. Interrater consensus on these rankings was significant, as indicated by a Kendall's W value of 0.648 ($p < 0.05$), reflecting strong agreement among the experts.

Appropriateness Across Research Pathways (Secondary Analysis)

The results obtained when ChatGPT was instructed to address different research pathways were also analyzed. In their assessment of the appropriateness of the outputs, most experts found the results from these pathways to be relevant and well-aligned with the research objectives (Table 4).

In the Geospatial Intelligence and Remote Sensing pathway, ChatGPT emphasized integrating geospatial analytics with real-time data to map hazard zones, optimize resource allocation, and enhance disaster response. The recommendations included reconciling diverse spatial resolutions, employing multi-resolution fusion for hazard detection, and validating models across varied environments to ensure adaptability and reliability for first responders. When addressing the Digital Twin pathway, ChatGPT emphasized creating real-time, interactive simulations to assist first responders in anticipating disaster evolution and testing response strategies. Key recommendations included designing a layered architecture for real-time updates, integrating dynamic hazard modeling validated against historical data, and enabling scenario-based training with instant feedback. Collaboration with emergency management partners and refining user interfaces were also highlighted to ensure practical impact and usability in high-pressure workflows. The output for the Semantic Web Approaches pathway emphasized establishing data interoperability through ontology design and RDF frameworks to address data fragmentation in disaster management. Key recommendations included developing or adapting ontologies for disaster concepts, implementing real-time data validation using machine learning, and iteratively testing with diverse data sources. The outputs also highlighted involving first responders in interface design and conducting controlled simulations to ensure the system's usability, effectiveness, and impact on situational awareness and decision-making.

Table 1. Primary analysis inputs

PhD proposal (summarized)	Keywords by PhD supervisors	Keywords by GPT	Concepts from topic modelling
The research proposal, titled *"Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration,"* aims to address critical challenges faced by emergency responders in rapidly evolving scenarios. Conducted at the Digital Systems Department of University X, the project seeks to develop a decision-support framework that integrates diverse data sources, validates their reliability, and leverages predictive analytics to provide actionable insights. A literature review highlights significant gaps in current systems, which struggle to integrate heterogeneous data from traditional sources (e.g., environmental sensors) and nontraditional sources (e.g., social media). This fragmentation hinders situational awareness and slows decision-making. The absence of real-time data validation mechanisms further compromises the reliability of information, reflecting challenges identified by the International Forum to Advance First Responder Innovation, particularly Capability Gap 4. The proposed framework addresses these issues through three objectives: (1) developing a scalable architecture to integrate heterogeneous data sources, ensuring compatibility and seamless communication; (2) incorporating real-time validation mechanisms to verify data reliability; and (3) applying predictive analytics to identify patterns, forecast incidents, and enable proactive decision-making. A user-centric design will provide customizable interfaces, allowing responders to prioritize information dynamically during high-pressure situations. Machine learning techniques and advanced data harmonization will be combined to deliver tailored insights that enhance situational awareness. The framework will be iteratively tested in simulated and real-world disaster scenarios to evaluate its effectiveness in improving decision-making speed and accuracy. This research has the potential to redefine emergency response by offering reliable, adaptable, and predictive decision-support tools. Its outcomes will advance disaster management operations, contribute to public safety innovation, and align with the mission of University X's Digital Systems Department to deliver impactful computational solutions.	• Predictive Analytics • Decision Support Systems • Data Integration • Emergency Response • Real-Time Validation • Situational Awareness • Heterogeneous Data Sources • Machine Learning • Disaster Management • Public Safety Innovation	• Disaster Risk Management • Semantic Web • Ontology Development • Data Interoperability • Machine Learning • Data Harmonization • Emergency Response • Predictive Analytics • First Responders • RDF triplestore	• Technology Awareness and Education in Disaster Management • Disaster Management Interventions Through Autonomous Systems • Capability and Capacity Building for Digital Resilience • Digital Technology-Based Monitoring and Prevention of Below-Surface Hazards/Accidents • Use of Social Media in Crisis Communication • Data Collection Through Social Media • Communication Networks and Data Applications in Disasters • Disaster Management Modeling • Emergency Response Management Systems

Table 2. Attributes of ChatGPT's input prompts and generated responses

Category	Input	Input (word count)	Output (word count)
Primary analysis:			
Naive	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout.	522	459
Keywords by PhD supervisors	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout. Please structure the recommendation using the keywords provided below: << Keywords by PhD supervisors >>	574	536
Keywords by chatGPT	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout. Please structure the recommendation using the keywords provided below: <<Keywords by chatGPT>>	572	633
Concepts from topic modeling	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout. Please structure the recommendation using the concepts provided below: <<Concepts from topic modeling>>	605	655
Secondary analysis:			
Geospatial Intelligence	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide	636	410

Category	Input	Input (word count)	Output (word count)
and Remote Sensing	recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. To strengthen the foundation of the proposed framework for disaster risk management, Geospatial Intelligence and Remote Sensing should be established as a central pillar. This pathway is justified by the critical need for first responders to anticipate incident evolution and take proactive measures, especially in rapidly changing and high-pressure scenarios. Current systems often fail to provide forward-looking insights, limiting their ability to optimize resource allocation and minimize disaster impacts. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout.		
Digital Twin technology	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. To strengthen the foundation of the proposed framework for disaster risk management, Digital Twin technology should be established as a central pillar. This approach is justified by its ability to create real-time, interactive simulations that mirror complex disaster scenarios. These virtual replicas provide first responders with a dynamic environment to monitor, predict, and experiment with potential response strategies during emergencies. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout.	627	404
Semantic Web Approaches	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. To strengthen the foundation of the proposed framework for disaster risk management, Semantic Web Approaches should be established as a central pillar. The rationale lies in the fragmented nature of the data sources that first responders must handle, including traditional sources like environmental sensors and nontraditional ones such as social media and citizen-generated reports. These sources often use inconsistent formats and terminologies, complicating interoperability and real-time integration. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout	634	652

Category	Input	Input (word count)	Output (word count)
Naive	Please analyze the following PhD proposal titled *Enhancing Decision-Making Systems for First Responders through Predictive Analytics and Advanced Data Integration*. Based on the proposal, provide recommendations on how the student should approach the research, including suggested methods, key steps, and potential challenges. Please do not use bullet points, numbered lists, or headings, but instead provide a continuous paragraph in a formal academic tone with full sentences throughout.	522	459

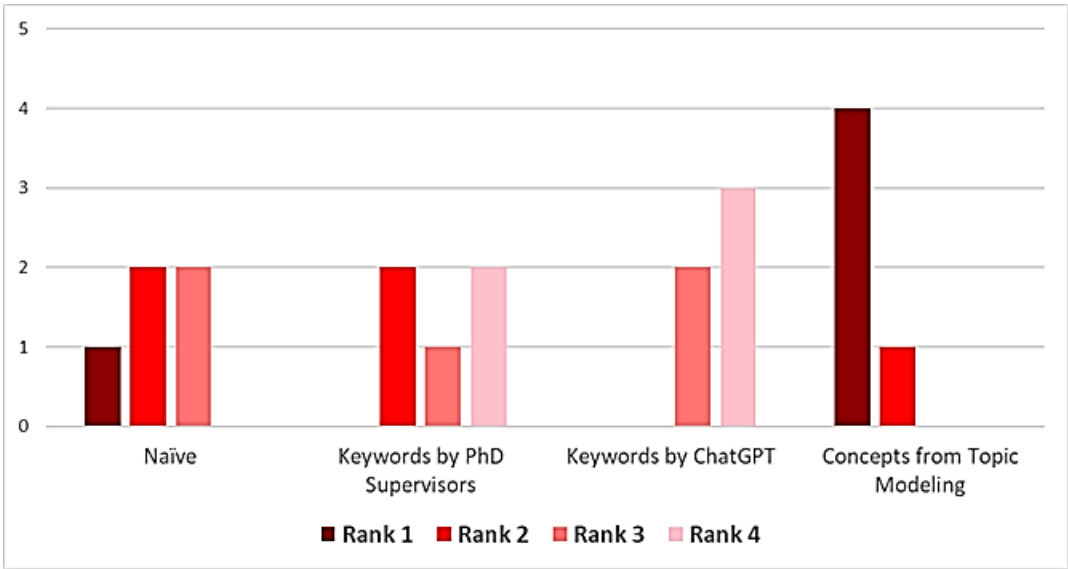

Figure 2. Ranking of ChatGPT outputs by prompt type

Each approach includes the original naive PhD proposal, with additional refinements based on PhD supervisors' keywords, ChatGPT-generated keywords, or topic modeling-derived concepts. Bars represent the number of times each approach was ranked 1st, 2nd, 3rd, or 4th across five raters.

Table 3. Performance outcomes for various inputs in the primary analysis

Category	Appropriateness	Key features of ChatGPT recommendations
Naive	100%	• The recommendations effectively identify the primary challenge of integrating diverse data streams. • Broad suggestions are offered for employing predictive analytics and user-centered design. • Key concerns regarding validating nontraditional sources, such as social media data, are addressed. • The suggestions indicate a reasonable grasp of the major technological gaps. • The proposed methodologies do not sufficiently interact to produce a cohesive, end-to-end decision-support framework.

Category	Appro-priateness	Key features of ChatGPT recommendations
		• The linkage of predictive analytics to real-world incident command operations remains underdeveloped. • The conceptual framework connecting data validation, system scalability, and final user adoption needs stronger emphasis. • A gap remains in translating high-level strategies into concrete, integrated solutions for first responders.
Keywords by PhD supervisors	80%	• The text covers important points about predictive analytics, decision support systems, and other key topics in emergency response. • It reads as though each paragraph was generated to match a specific keyword rather than forming a cohesive narrative. • The recommendations are valid but do not flow smoothly into one another. • The discussion feels fragmented rather than integrated. • The potential synergy among data integration, real-time validation, and user-centric interfaces is not highlighted enough. • The framework's overarching value for first responders is insufficiently articulated.
Keywords by chatGPT	40%	• The text references concepts such as data harmonization, semantic web, machine learning, and predictive analytics. • Each keyword seems to anchor its own paragraph without sufficiently demonstrating how these elements intertwine. • The output feels disjointed rather than cohesive. • A clear understanding of how machine learning and semantic web technologies can jointly enhance disaster risk management is lacking.
Concepts from topic modeling	100%	• The recommendations effectively incorporate all the specified topics related to disaster management and digital resilience. • They present a cohesive and integrated narrative. • The core challenges faced by first responders, such as data fragmentation, real-time validation, and technological literacy, are clearly articulated and systematically addressed within each section. • Central themes like interoperability, predictive analytics, and user-centric design are formulated using a structured methodology that aligns with the identified research dimensions. • Despite the absence of a specific research direction outlined in advance, the recommendations adopt a comprehensive approach. • This approach integrates technology awareness, autonomous systems, and digital resilience into a unified decision-support framework. • The alignment ensures that diverse topics are not treated as isolated components but interwoven to create a holistic solution tailored to the dynamic challenges of disaster management.

Table 4. Performance outcomes for various inputs in the secondary analysis

Research direction	Appro-priateness	Key features of ChatGPT recommendations
Geospatial intelligence and remote sensing	80%	• Establish a geo-data preprocessing pipeline that reconciles different spatial resolutions and coordinates across satellite imagery, drone footage, and sensor data. • Develop a multi-resolution fusion technique that integrates wide-coverage, lower-resolution data with narrow-coverage, high-resolution feeds to enhance hazard detection. • Integrate spatial analysis methods (GIS tools, network analysis) to tailor insights to the local geographic context, enabling first responders to identify hazard zones, routes, and resource needs. • Validate models in diverse geographic contexts (urban vs. rural, coastal vs. mountainous) to ensure adaptability and reliability across multiple disaster scenarios. • Incorporate resource allocation logic that translates predictive outputs into on-the-ground strategies, measured by real-world collaborations and user satisfaction metrics.
Digital twin technology	80%	• Design a layered architecture that separates data ingestion (sensors, IoT) from simulation modules (3D visualization, situational modeling) to ensure real-time updates and modular scalability. • Integrate dynamic hazard modeling so the virtual environment adjusts to new data (flood levels, fire perimeters), validated against historical incidents to confirm realism. • Conduct cross-layer calibration by comparing simulated conditions to real logs or sensor data, triggering alerts for major discrepancies. • Enable scenario-based training in which first responders test various tactics (evacuation routes, resource deployment) and receive instant feedback on outcomes. • Collaborate with emergency management partners for field tests and refine user interfaces to fit within high-pressure workflows, mitigating resistance and ensuring practical impact.
Semantic web approaches	100%	• Develop a prototype ontology or adapt an existing one to represent key disaster concepts, and store/query data via an RDF triplestore. • Implement real-time data validation that employs machine learning classifiers for credibility scoring and anomaly detection in sensor feeds. • Conduct iterative testing with increasingly complex scenarios, beginning with structured sensor data and gradually incorporating social media or crowdsourced inputs. • Involve first responders in the user-centered design of interfaces to ensure that the system meets practical needs, remains intuitive, and supports rapid decision-making. • Measure performance through controlled simulations and real-world trials that evaluate improvements in situational awareness and decision-making speed.

DISCUSSION

We instructed ChatGPT to generate academic guidance for DRM research using established approaches and frameworks. To our knowledge, this is the first study to systematically evaluate the use of ChatGPT to provide academic recommendations for a PhD proposal. The experiment involved varying inputs, and we observed that the most structured and relevant guidance was obtained when specific research pathways related to the PhD proposal were included. In contrast, less relevant outputs resulted from general keyword-based prompts. Additionally, ChatGPT demonstrated the capacity to adapt its outputs to diverse domains, tailoring recommendations to fit specific contexts. This finding aligns with evidence suggesting that generative AI tools excel when tasks are well-defined, as highlighted in a survey of Danish researchers, where clear instructions enhanced AI's effectiveness (Andersen et al., 2024).

While output lengths varied, appropriateness ratings indicated that content coherence and domain alignment were more important than response verbosity. Specifically, the naive output aligned with the multifaceted challenges in DRM, such as the need for semantic harmonization of heterogeneous data sources, real-time validation of nontraditional inputs like social media feeds, the integration of predictive analytics to forecast incident evolution, and geospatial intelligence to enhance situational awareness. This reflects a global need to establish frameworks that combine semantic web approaches, machine learning models, and scalable architectures to address inconsistencies, ensure data reliability, and improve decision-making for first responders (Dirgahayu et al., 2020). By capturing these elements, ChatGPT underscored the potential of AI-driven and ontology-based frameworks to support the research objectives, demonstrating its ability to generate actionable insights for the next steps of the PhD. A similar benefit of generative AI in reducing cognitive load and enhancing productivity was observed, where ChatGPT provided scaffolding for nursing PhD students (Reading Turchioe et al., 2024).

Moreover, outputs generated using keyword-based prompts focused primarily on defining and elaborating terms such as data interoperability, predictive analytics, and validation mechanisms. While these aspects are critical to the framework, the responses often failed to connect these concepts explicitly to the core gaps identified in the proposal, such as fragmented data and real-time decision-making. Instead, they addressed technical aspects in isolation, limiting their ability to provide comprehensive, solution-oriented recommendations. This observation underscores the importance of contextualizing inputs to maximize the relevance and coherence of ChatGPT's outputs.

Compared to thematically focused inputs, such as those based on topic modeling or research dimensions like Interoperable Systems for First Responders, produced outputs that were more actionable and contextually grounded. These prompts enabled ChatGPT to align its recommendations directly with the goals and practical applications of the research dimension, enhancing the quality and relevance of the guidance. For example, structured prompts resulted in recommendations for modular ontology design, scalable frameworks, stakeholder engagement, and, in the case of geospatial intelligence, specific integration techniques for remote sensing data, all of which are critical to addressing the dynamic challenges of disaster scenarios.

Our methodology ensured consistency in the structure of prompts while modifying only the content of the background information provided. This approach allowed us to pinpoint the elements of the inputs that significantly enhanced the quality of the recommendations. The inclusion of even minimal background on key disaster management topics, such as All-Source Data Collection and Integration, resulted in more coherent and relevant guidance. For instance, semantic web approaches were contextualized to support modular ontology engineering, standards alignment, and predictive analytics, reflecting core aspects of disaster management research.

The AI-generated outputs, as applied to this PhD case study, also underscore the importance of integrating real-time validation mechanisms to enhance the reliability and accuracy of data streams. Recommendations for dynamic systems capable of validating and updating data during crises were consistent with the research objectives of the proposal. These systems were positioned as essential tools for improving situational awareness and decision-making, particularly for first responders operating under high-pressure conditions.

Within the simulated research problem space, stakeholder engagement emerged as another critical theme throughout the outputs. The recommendations consistently advocated for participatory approaches involving first responders and policymakers during framework development. This user-centric methodology ensures that the proposed solutions address operational workflows and practical needs effectively, a key requirement for real-world application in DRM (Mojtahedi & Oo, 2017). Embedding these participatory methods also resonates with sustainable research and education practicefs, as it nurtures ongoing learning communities and fosters inclusive decision-making.

Despite these strengths, some areas warrant further development as identified by external academic experts. While the outputs of this case study proposed robust solutions for fragmented data systems and real-time validation, they offered limited focus on addressing cascading effects in multi-hazard scenarios. Such scenarios, where disasters overlap or compound one another, pose unique challenges that future iterations of the research should explore (Pittore et al., 2020). Similarly, external experts noted that while ethical considerations, such as data privacy and compliance with regulations like the General Data Protection Regulation (GDPR), were acknowledged, they require deeper integration into the proposed framework to ensure responsible and secure data usage (Nussbaumer et al., 2023).

Another area for improvement lies in the integration of advanced analytics, as highlighted by academic reviewers. While the outputs recommend incorporating machine learning and AI, they could provide more explicit guidance on how these technologies would complement the proposed approaches. For example, employing machine learning-based clustering techniques to identify patterns in heterogeneous data streams could enhance the system's capability to detect emerging hotspots of activity during disaster scenarios, enabling faster allocation of resources and prioritization of critical response actions (Zhang et al., 2023).

Overall, the results demonstrate ChatGPT's ability to provide meaningful academic guidance by generating contextually relevant and methodologically grounded recommendations when applied to a complex doctoral research scenario. In this case study, ChatGPT addressed challenges such as data fragmentation, real-time validation, scalability, and the integration of geospatial intelligence within the simulated disaster risk management domain. These observations illustrate the model's capacity to engage with sophisticated research contexts, offering structured guidance that supports doctoral research design. Further refinements - including more explicit attention to multi-hazard scenarios, ethical considerations, and advanced analytics - would further enhance the applicability of such AI-assisted guidance. These findings underscore ChatGPT's broader potential as a valuable support tool in doctoral supervision, particularly for navigating interdisciplinary and technically demanding research problems.

While the use of AI chatbots in advising is still an emerging practice, findings suggest that platforms like ChatGPT have the potential to transform academic mentoring by providing flexible, on-demand guidance throughout an academic journey (Iatrellis et al., 2024). Unlike human advisors, whose availability may be limited by time and institutional responsibilities, ChatGPT offers constant accessibility, delivering feedback and recommendations instantly and at any stage of the research process. This advantage is particularly significant in the iterative nature of PhD research, where the need for timely input often arises outside conventional consultation schedules. One of the primary benefits of ChatGPT in the context of PhD advising is its ability to synthesize complex theoretical frameworks and methodologies, as demonstrated in its application to the PhD proposal on DRM. By processing and integrating diverse information sources, ChatGPT can offer structured guidance tailored to the

researcher's objectives. For example, it can align proposals with established academic standards, recommend methodologies like ontology engineering or semantic web technologies, advise on geospatial intelligence considerations, and identify relevant gaps in the literature. This adaptability allows the platform to act as a complementary resource for researchers, particularly in interdisciplinary fields where expertise may span multiple domains.

Moreover, while much of the existing literature on AI in higher education has primarily focused on automating administrative tasks or providing instant feedback on discrete learning activities, our findings suggest a broader and more integrative role for AI in doctoral research supervision. Specifically, the PhD journey may evolve into a **tripartite mentoring model**, in which the student, the supervisor, and an AI assistant collaboratively shape and refine the research process (Figure 3). This evolving framework serves as the foundation for the following model proposed in this study, which formalizes the interaction of these three actors through clearly defined principles, roles, and responsibilities.

The proposed tripartite mentoring model builds upon these findings and is grounded on seven interdependent principles that define the collaborative interaction among the student, supervisor, and AI assistant (Figure 4):

1. *Complementary Roles and Defined Responsibilities:* Each participant contributes unique expertise: the supervisor offers domain knowledge and scholarly judgment; the AI assistant delivers rapid, context-specific academic guidance; and the student engages in critical evaluation, synthesis, and application of both human and AI input. The empirical findings of this study demonstrated that ChatGPT, when guided by well-structured and context-rich prompts, was able to generate academically relevant and actionable recommendations, effectively complementing both the supervisor's expertise and the student's decision-making. This triadic configuration enables a distributed mentoring process where each actor contributes distinct, yet synergistic, forms of academic support.

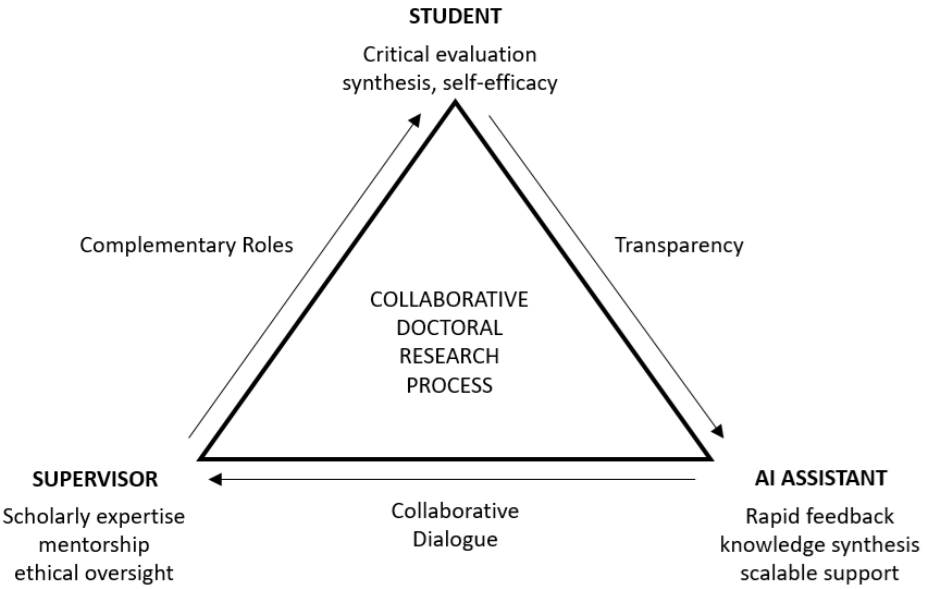

Figure 3. The tripartite mentoring model

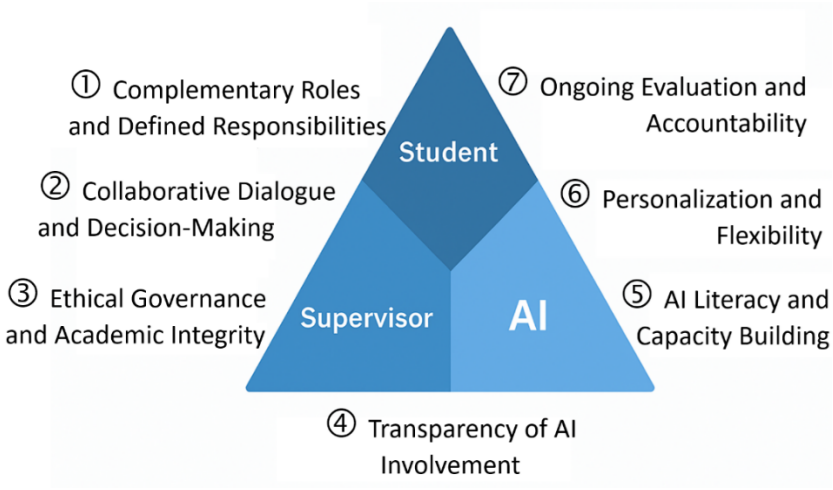

Figure 4. Principles of the tripartite mentoring model

2. *Collaborative Dialogue and Decision-Making:* A dynamic feedback loop fosters iterative refinement of research ideas, methodologies, and writing through transparent interaction among all parties. The study's results indicated that prompt-driven interactions with ChatGPT allowed for iterative cycles of refinement, where AI-generated suggestions offered additional pathways and alternatives for consideration, encouraging both supervisors and students to critically engage with new research possibilities and integrate them into the doctoral trajectory.
3. *Ethical Governance and Academic Integrity:* The model embeds ethical guidelines addressing responsible AI use, data privacy, fairness, and protection against over-reliance or algorithmic bias. External expert evaluations in this study emphasized the importance of integrating ethical oversight when leveraging AI-generated content in doctoral supervision, particularly to safeguard academic integrity, avoid over-dependence on AI outputs, and ensure that critical academic judgment remains firmly rooted in human expertise.
4. *Transparency of AI Involvement:* AI's role remains visible and accountable within the supervisory relationship, preventing covert or unsanctioned use of AI-generated content. The study underscores the necessity of openly acknowledging AI contributions in the mentoring process, ensuring that ChatGPT operates as an explicitly integrated advisory tool rather than an unregulated background influence, thereby preserving transparency in both supervision and academic authorship.
5. *AI Literacy and Capacity Building:* Both students and supervisors develop skills to effectively prompt, interpret, and critically evaluate AI-generated outputs, ensuring informed decision-making. The findings demonstrate that the quality of ChatGPT's academic guidance was highly dependent on the sophistication and contextualization of the prompts, highlighting the importance of building AI literacy and prompting skills among doctoral stakeholders to fully leverage AI capabilities while avoiding misleading outputs.
6. *Personalization and Flexibility:* The model adapts to the student's developmental stage, allowing AI support to scale and evolve as research progresses, while maintaining human mentoring as the guiding authority. As shown in the study, ChatGPT demonstrated adaptive capabilities when provided with progressively richer background information, generating increasingly relevant and customized academic guidance that aligned with the evolving needs of the PhD research journey.

7. *Ongoing Evaluation and Accountability:* The model incorporates continuous assessment mechanisms to monitor student progress, evaluate research outcomes, and refine supervisory processes. This includes tracking milestones (e.g., proposal approval, conference papers), applying rubric-based assessments, and collecting satisfaction data from both students and supervisors. The introduction of multi-dimensional evaluation criteria during expert assessment in this study offers an initial framework for such monitoring, suggesting that future implementations of tripartite supervision could incorporate structured rubrics to systematically evaluate AI-generated contributions alongside traditional performance indicators.

With the advent of GenAI, this model has permeated universities, emerging as a framework worthy of further exploration, driven by the evolving academic ecosystem. This aligns with other research highlighting the transformative potential of AI in reshaping supervisory roles and integrating collaborative approaches to doctoral mentoring, emphasizing the need for frameworks that adapt to technological advancements while maintaining the relational aspects of effective supervision (Andersen et al., 2024; Caillaud & Skec, 2024; Harding & Boyd, 2024; Zou & Huang, 2023). While this study contributes initial insights into the conceptualization of the tripartite mentoring model, its full operationalization, long-term effects, and institutional implementation remain important areas for future empirical investigation.

GENERALIZATION AND BROADER APPLICABILITY

Although the present study focused on a single PhD proposal in disaster risk management, the findings offer valuable principles that can be generalized to other doctoral contexts. By design, PhD research is highly individualized, often making broad-scale comparisons challenging; nonetheless, this case illustrates how a thoughtful, iterative approach to prompt design and the inclusion of rich contextual details can yield more coherent and domain-relevant AI outputs. These practices - tailoring prompt structure, integrating established frameworks, and validating results against expert feedback - are readily adaptable to diverse fields such as healthcare, engineering, or the social sciences. Examining how these approaches function in various institutional and disciplinary settings can shed additional light on whether similar patterns of prompt effectiveness and limitations emerge. Over time, such comparative work may inform robust practices for harnessing generative AI tools like ChatGPT in doctoral mentorship and academic innovation.

ETHICAL CONSIDERATIONS

The implementation of AI tools such as ChatGPT in doctoral supervision raises important ethical challenges that must be addressed to ensure responsible use. Data privacy remains a primary concern, particularly in contexts where sensitive or unpublished research data may be processed by third-party AI systems. Institutions must ensure strict adherence to privacy regulations such as the GDPR. Additionally, inherent biases in AI-generated content may unintentionally shape research directions, requiring supervisors and students to critically review AI outputs rather than accepting recommendations uncritically. The risk of over-reliance on AI-generated suggestions also highlights the need to preserve student autonomy and supervisor oversight in research decision-making. Furthermore, academic integrity must remain central, with clear guidelines to ensure that AI tools support, but do not replace, original scholarly thinking, authorship, and rigorous methodological development. Institutions should consider developing explicit policies and training programs to guide the ethical integration of generative AI into PhD mentoring practices.

LIMITATIONS

There are several limitations to our study. First, we utilized only one research proposal as the basis for evaluating ChatGPT's ability to provide academic guidance in the context of PhD advising. While

the chosen proposal was carefully selected to reflect real-world research challenges in disaster risk management, our findings would benefit from validation using a broader range of proposals across diverse academic fields. Expanding the dataset in future studies to include multiple proposals across various disciplines would significantly enhance the generalizability of the results.

Second, the evaluation of ChatGPT's outputs involves an inherent degree of subjectivity. While we applied consistent criteria to assess the relevance and appropriateness of the recommendations, the absence of widely accepted benchmarks for academic guidance in this context poses a challenge. Future studies could benefit from the development of structured evaluation metrics tailored to academic advising scenarios, focusing on aspects such as clarity, coherence, and practical applicability.

Third, the prompts designed for generating recommendations were informed by structured disaster management concepts and theoretical frameworks. Although care was taken to avoid overly prescriptive inputs, the inclusion of specific domain knowledge might have influenced the nature of the outputs. Future research incorporating a wider array of prompts, including more open-ended, exploratory, and less domain-specific inputs, could further clarify ChatGPT's general capabilities and limitations in generating academic guidance.

Fourth, there is a possibility that the training data used to develop ChatGPT included elements related to disaster risk management, which might have influenced the outputs. However, given the vast scale of the training dataset, it is unlikely that any single piece of information could have a significant impact on the results. Future research could explicitly investigate the extent to which pre-training domain exposure influences AI-generated outputs in specific academic advising scenarios.

Lastly, this study focused exclusively on English-language outputs and was conducted within the context of a specific case study. While the findings are relevant to academic advising in general, they may not fully account for variations in academic norms, linguistic diversity, or disciplinary expectations. Future studies should explicitly consider cross-disciplinary comparisons, multilingual evaluations, and variations in cultural and institutional expectations. Such research would provide a more comprehensive understanding of generative AI's effectiveness and appropriateness across diverse doctoral supervision contexts.

CONCLUSIONS

We used ChatGPT to generate recommendations for a PhD proposal focused on developing advanced frameworks for academic research, revealing its capacity to produce relevant outcomes even with minimal input. More structured and context-specific prompts, however, yielded richer, domain-centered guidance. In the case of disaster risk management, this approach helped connect theoretical constructs like semantic web frameworks and geospatial intelligence to real-world challenges, underscoring the importance of thoughtful prompt design. Such strategies can be readily extended to doctoral mentoring more broadly. By beginning with a broad research overview and gradually introducing specialized concepts, educators can guide AI toward deeper discipline-specific insights. Students and supervisors can then evaluate these outputs together, contrasting them with established scholarly standards and refining methodologies or literature reviews accordingly. Throughout this process, it remains essential to uphold responsible usage, including data privacy considerations and ethical guidelines, to ensure that AI serves as an assistive rather than a decisive authority.

While our findings highlight the potential benefits of generative AI tools like ChatGPT for doctoral mentoring-such as expedited iterative feedback, stimulation of novel ideas, and reduced supervisory burdens-these results should be interpreted with caution given the exploratory nature of this study. Critical limitations include possible biases in AI-generated content, the risk of over-reliance on automated recommendations, and uncertainties regarding the effectiveness of such tools across different disciplines, cultures, and languages. Future research should investigate these limitations explicitly, evaluating ChatGPT's application across diverse academic and multilingual contexts. Further studies

should also explore the formal integration and evaluation of the proposed "tripartite mentoring model," where generative AI collaborates systematically with supervisors and students to enhance doctoral research practices sustainably and ethically.

Data Availability

The data (ChatGPT responses) used in this study can be accessed at https://doi.org/10.5281/zenodo.14976424

Acknowledgement

Supported by the INVEST project under the ERASMUS EDU-2023-EUR-UNIV program (Project No.: 101124598, www.invest-alliance.eu). Special thanks to the other project partners. This work has been conducted within the RES-Q Living Lab, which was established through the INVEST project framework.

References

Andersen, J. P., Degn, L., Fishberg, R., Graversen, E. K., Horbach, S. P. J. M., Schmidt, E. K., Schneider, J. W., & Sørensen, M. P. (2024). *Generative Artificial Intelligence (GenAI) in the research process – A survey of researchers' practices and perceptions*. SocArXiv Papers. https://doi.org/10.31235/osf.io/83whe

Anwar, B., Shoaib, M., & Shahid, U. (2024). Academic performance of the students in higher education in Pakistan: The role of information and communication technology. *Pakistan Social Sciences Review*, *8*(2), 472–482.

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, opportunities, and threats. *Proceedings of the Systems and Information Engineering Design Symposium, Charlottesville, VA, USA,* 274–279. https://doi.org/10.1109/SIEDS58326.2023.10137850

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, Nusa Dua, Bali*, 675–718. https://doi.org/10.18653/v1/2023.ijcnlp-main.45

Bania, A., Iatrellis, O., & Samaras, N. (2022, August). Information Communication Technologies (ICTs) and Disaster Risk Management (DRM): Systematic literature review. *Proceedings of the 6th Conference on Sustainable Urban Mobility, Skiathos Island, Greece*, 1779-1794.

Bania, A., Iatrellis, O., Samaras, N., & Panagiotakopoulos, T. (2024). *Fires: A semantic model for advanced querying and prediction analysis for first responders in post-disaster response plans*. SSRN. https://doi.org/10.2139/ssrn.4527923

Caillaud, E., & Skec, S. (2024, September). Supervision of Design PhD students in an era of generative artificial intelligence. *Proceedings of the International Conference on Engineering and Product Design Education, Aston University, Birmingham, United Kingdom*, 420–425. https://doi.org/10.35199/epde.2024.71

Cardona, T. A., & Cudney, E. (2019). Predicting student retention using support vector machines. *Procedia Manufacturing*, *39*(2019), 1827–1833. https://doi.org/10.1016/j.promfg.2020.01.256

Chan, Z. C. Y., Chan, H. Y., Chow, H. C. J., Choy, S. N., Ng, K. Y., Wong, K. Y., & Yu, P. K. (2019). Academic advising in undergraduate education: A systematic review. *Nurse Education Today*, *75*, 58–74. https://doi.org/10.1016/j.nedt.2019.01.009

Cheng, S. L., Tsai, S. J., Bai, Y. M., Ko, C. H., Hsu, C. W., Yang, F. C., Tsai, C. K., Tu, Y. K., Yang, S. N., Tseng, P. T., Hsu, T. W., Liang, C. S., & Su, K. P. (2023). Comparisons of quality, correctness, and similarity between ChatGPT-generated and human-written abstracts for basic research: Cross-sectional study. *Journal of Medical Internet Research*, 25, e51229. https://doi.org/10.2196/51229

Dirgahayu, T., Hendrik, & Setiaji, H. (2020). Semantic web in disaster management: A systematic literature review. *IOP Conference Series: Materials Science and Engineering*, *803*, 012043. https://doi.org/10.1088/1757-899X/803/1/012043

Ghosh, C. (2023). GIS and geospatial studies in disaster management. In A. Singh (Ed.), *International handbook of disaster research* (pp. 701–708). Springer. https://doi.org/10.1007/978-981-19-8388-7_214

Han, X. (2024). Generative artificial intelligence for future education: Current research status, hot spots, and research trends. *Proceedings of the 8th International Conference on Algorithms, Computing and Systems* (pp. 56–61). Association for Computing Machinery. https://doi.org/10.1145/3708597.3708606

Harding, D., & Boyd, P. (2024). Generative AI and PhD supervision: A covert third wheel or a seat at the table? In C. Thomson, & C. J. Neilson (Eds.), *Proceedings of the 40th Annual ARCOM Conference* (pp. 91–100). Association of Researchers in Construction Management.

Hasanuzzaman, M., Hossain, S., & Shil, S. K. (2023). Enhancing disaster management through AI-driven predictive analytics: Improving preparedness and response. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(1), 533–562.

Higgins, B., & Thomas, I. (2016). Education for sustainability in universities: Challenges and opportunities for change. *Australian Journal of Environmental Education*, *32*(1), 91–108. https://doi.org/10.1017/aee.2015.56

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), Article 42. https://doi.org/10.1145/3703155

Iatrellis, O., Kameas, A., & Fitsilis, P. (2020). EDUC8 pathways: Executing self-evolving and personalized intra-organizational educational processes. *Evolving Systems*, *11*, 277–240. https://doi.org/10.1007/s12530-019-09287-4

Iatrellis, O., Samaras, N., Kokkinos, K., & Panagiotakopoulos, T. (2024). Leveraging generative AI for sustainable academic advising: Enhancing educational practices through AI-driven recommendations. *Sustainability*, *16*(17), 7829. https://doi.org/10.3390/su16177829

International Forum to Advance First Responder Innovation. (2018, August). *Capability Gap 4 "deep dive" analysis synopsis*. https://www.internationalresponderforum.org/sites/default/files/2023-02/gap4_analysis.pdf

Kabra, G., Ghosh, V., & Mukherjee, R. (2024). Discovering latent topics and trends in digital technologies and disaster management research: A structural topic modeling approach. *Engineering Management Journal*. https://doi.org/10.1080/10429247.2024.2345526

Kouatli, I. (2020). The need for social and academic responsibility advisor (SARA): A catalyst toward the sustainability of educational institutes. *Social Responsibility Journal*, *16*(8), 1275–1291. https://doi.org/10.1108/SRJ-04-2019-0151

Kuhail, M. A., Al Katheeri, H., Negreiros, J., Seffah, A., & Alfandi, O. (2023). Engaging students with a chatbot-based academic advising system. *International Journal of Human-Computer Interaction*, *39*(10), 2115–2141. https://doi.org/10.1080/10447318.2022.2074645

Maher, M. A., Gilmore, J. A., Feldon, D. F., & Davis, T. E. (2013). Cognitive apprenticeship and the supervision of science and engineering research assistants. *Journal of Research Practice*, *9*(2), Article M5.

Maphosa, V., & Maphosa, M. (2023). Fifteen years of recommender systems research in higher education: Current trends and future direction. *Applied Artificial Intelligence*, 37(1). https://doi.org/10.1080/08839514.2023.2175106

McCallin, A., & Nayar, S. (2012). Postgraduate research supervision: A critical review of current practice. *Teaching in Higher Education*, *17*(1), 63–74. https://doi.org/10.1080/13562517.2011.590979

Melián, E., Reyes, J. I., & Meneses, J. (2023). The online PhD experience: A qualitative systematic review. *The International Review of Research in Open and Distributed Learning*, *24*(1), 137–158. https://doi.org/10.19173/irrodl.v24i1.6780

Mian, S. H., Salah, B., Ameen, W., Moiduddin, K., & Alkhalefah, H. (2020). Adapting universities for sustainability education in industry 4.0: Channel of challenges and opportunities. *Sustainability*, *12*(15), 6100. https://doi.org/10.3390/su12156100

Miller, M., White, E., Nutt, C., & Drake, J. (2015). The role of academic advising programs. *CAS Standards Contextual Statement*, *16*(3), 8–12.

Mojtahedi, M., & Oo, B. L. (2017). Critical attributes for proactive engagement of stakeholders in disaster risk management. *International Journal of Disaster Risk Reduction*, *21*, 35–43. https://doi.org/10.1016/j.ijdrr.2016.10.017

Moundridou, M., Matzakos, N., & Doukakis, S. (2024). Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans. *Computers and Education: Artificial Intelligence*, 7, 100277. https://doi.org/10.1016/j.caeai.2024.100277

Nussbaumer, A., Pope, A., & Neville, K. (2023). A framework for applying ethics-by-design to decision support systems for emergency management. *Information Systems Journal*, *33*(1), 34–55. https://doi.org/10.1111/isj.12350

Pittore, M., Gómez Zapata, J. C., Brinckmann, N., Weatherill, G., Babeyko, A., Harig, S., Mahdavi, A., Proß, B., Rosero Velasquez, H. F., Straub, D., Krautblatter, M., Frimberger, T., Langbein, M., Geiß, C., & Schoepfer, E. (2020). Towards an integrated framework for distributed, modular multi-risk scenario assessment. *EGU General Assembly*. https://doi.org/10.5194/egusphere-egu2020-19097

Polkinghorne, M., Taylor, J., Knight, F., & Stewart, N. (2023). Doctoral supervision: A best practice review. *Encyclopedia*, 3(1), 46–59. https://doi.org/10.3390/encyclopedia3010004

Reading Turchioe, M., Kisselev, S., Fan, R., & Bakken, S. (2024). Returning value from the All of Us research program to PhD-level nursing students using ChatGPT as programming support: Results from a mixed-methods experimental feasibility study. *Journal of the American Medical Informatics Association*, *31*(12), 2974–2979. https://doi.org/10.1093/jamia/ocae208

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of risks posed by language models. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, *Seoul, Republic of Korea,* 214–229. https://doi.org/10.1145/3531146.3533088

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. PsyArXiv. http://arxiv.org/abs/2302.11382

Yu, D., & He, Z. (2022). Digital twin-driven intelligence disaster prevention and mitigation for infrastructure: advances, challenges, and opportunities. *Natural Hazards, 112*(1), 1–36. Springer Science and Business Media B.V. https://doi.org/10.1007/s11069-021-05190-x

Yun, S. J., Kwon, J. W., & Kim, W. T. (2022). A novel digital twin architecture with similarity-based hybrid modeling for supporting dependable disaster management systems. *Sensors, 22*(13). https://doi.org/10.3390/s22134774

Zhang, T., Wang, D., & Lu, Y. (2023). Machine learning-enabled regional multi-hazards risk assessment considering social vulnerability. *Scientific Reports, 13*(1), 1–14. https://doi.org/10.1038/s41598-023-40159-9

Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. *Education Sciences, 14*(12). https://doi.org/10.3390/educsci14121302

Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology, 14*(October), 1–9. https://doi.org/10.3389/fpsyg.2023.1259531

AUTHORS

Omiros Iatrellis holds a PhD in Educational Technology from the School of Science & Technology, Hellenic Open University (Greece), with his doctoral research focusing on AI-supported career advising systems. Since 2004, Dr. Iatrellis has been active in higher education teaching and research, with a particular emphasis on leveraging digital technologies to enhance learning and guidance processes. He is currently an Assistant Professor in the Department of Digital Systems, University of Thessaly. His research focuses on intelligent academic and career guidance, educational data mining, and technology-enhanced learning systems. He has led or contributed to several national and European projects involving the design of ICT-based solutions that incorporate artificial intelligence, big data, and semantic web technologies to support personalized learning and career development. He has published over 50 peer-reviewed articles in international journals and conferences on innovative methodologies and digital tools for higher education.

Areti Bania is a PhD candidate in the Department of Digital Systems, University of Thessaly, in Larissa, Greece. She holds a bachelor's degree in Computer Science from Aristotle University of Thessaloniki and two master's degrees: one in Information and Telecommunication Systems and another in Agile Management. Her research interests include Disaster Risk Management, Semantic Web technologies, Machine Learning, Artificial Intelligence, and Decision Support Systems.

Nicholas S. Samaras is a Professor at the Department of Digital Systems at the University of Thessaly. Prof. Samaras obtained his BS and MS degrees from the City College of the City University of New York, NY, and his PhD degree from the University of Pittsburgh, Pittsburgh, PA, USA, in 1986, 1987, and 1999, respectively, all in Electrical Engineering. Prof. Samaras has a distinguished career involving numerous National and European research projects focused on educational innovations and methodologies. For over 15 years, he has developed and implemented various distributed control and automation systems globally. He has authored over 55 publications in major refereed international journals, conferences, and book chapters, many of which explore educational technologies and learning systems. Additionally, Prof. Samaras has played a significant role as the local coordinator for the INVEST European University project funded by ERASMUS+ and the INVEST4EXCELLENCE project funded by H2020, both aimed at advancing educational systems in Europe.

Ioanna Kosmopoulou is a PhD candidate at the Department of Digital Systems, University of Thessaly, in Larissa, Greece. She holds a bachelor's degree in Computer Engineering and Telecommunications from the University of Thessaly and a master's degree in Computer Science from the same institution. Her research interests include Educational Data Mining (EDM), Learning Analytics (LA), and the application of Artificial Intelligence in education. She is particularly focused on promoting personalized learning through the use of innovative methodologies.

Theodoros Panagiotakopoulos is an Assistant Professor in the Department of Management Science and Technology at the University of Patras, specializing in "Internet of Things Applications." In 2006, he received his diploma from the Department of Electrical and Computer Engineering at the University of Patras, and in 2011, he obtained his PhD in Pervasive Systems from the same department. Since 2011, he has been a research fellow in the Laboratory of Mobile and Pervasive Computing, Quality, and Ambient Intelligence of the Hellenic University, and since 2020, a visiting Professor at the University of Nicosia. His research interests include Internet of Things (IoT) technologies (cloud platforms, middleware, machine learning, deep learning, etc.), IoT applications with a focus on smart cities, energy, commerce, healthcare, and transportation, as well as continuous professional training of the workforce in key emerging technologies through e-learning systems. He has published numerous scientific papers in reputable international journals and conferences. He has also participated in more than 25 European and National research projects (e.g., H2020, Horizon Europe, and Erasmus+), holding key roles at research, technical, and managerial levels.