# Artificial intelligence in forensic science: Evaluation of ChatGPT for post-mortem interval estimation and Henssge nomogram use

Elena Giovannini[1,A,B,D], Simone Santelli[1,C,D], Jacopo Lenzi[2,B,C], Guido Pelletti[1,D,E], Luca Berti[1,B,D], Arianna Giorgetti[1,C,E], Filippo Pirani[1,B,C], Susi Pelotti[1,E,F], Paolo Fais[1,3,E,F]

[1] Department of Medical and Surgical Sciences, Unit of Legal Medicine, University of Bologna, Italy
[2] Department of Hygiene and Public Health, University of Bologna, Italy
[3] Legal Medicine and Risk Management Unit, Azienda Unita Sanitaria Locale (USL) di Bologna, Italy

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2026;35(8):1437–1444

**Address for correspondence**
Guido Pelletti
E-mail: guidopelletti@gmail.com

**Funding sources**
None declared

**Conflict of interest**
None declared

Received on July 2, 2025
Reviewed on September 10, 2025
Accepted on October 7, 2025

Published online on August 3, 2026

## Abstract

**Background.** In recent years, advancements in artificial intelligence (AI) have led to the creation of numerous large language models, including Chat Generative Pre-trained Transformer (ChatGPT), a natural language processing model developed by OpenAI. There has been increasing interest in exploring the potential of this tool, as evidenced by several studies in the healthcare domain that have investigated the performance of ChatGPT in responding to specific questions on relevant topics. However, its application in forensic contexts remains largely unexplored.

**Objectives.** This study aims to evaluate the performance of ChatGPT in estimating the post-mortem interval (PMI) by considering thanatochronological changes as well as through the application of the Henssge nomogram.

**Materials and methods.** Fifteen questions concerning PMI estimation were presented to ChatGPT. These questions were categorized into 3 domains, each addressing different aspects: 1) theoretical notions, 2) practical application of theoretical notions, and 3) utilization of the Henssge nomogram. The evaluation of responses included 3 sub-criteria: focus, accuracy, and completeness.

**Results.** The results showed a high level of accuracy and completeness in addressing theoretical issues. However, the AI failed when responding to practical case scenarios and when calculating PMI using the Henssge nomogram.

**Conclusions.** Although AI may be attractive for application to forensic science questions, uncertain source documents and incomplete access to the scientific literature may affect accuracy. The study concluded that AI should be investigated for use in forensic science; however, it is currently not suitable for practical PMI estimation.

**Key words:** artificial intelligence, post-mortem interval, performance evaluation, ChatGPT, post-mortem interval estimation

**Cite as**
Giovannini E, Santelli S, Lenzi J, et al. Artificial intelligence in forensic science: Evaluation of ChatGPT for post-mortem interval estimation and Henssge nomogram use. *Adv Clin Exp Med*. 2026;35(8):1437–1444. doi:10.17219/acem/211777

**DOI**
10.17219/acem/211777

## Highlights

- The use of artificial intelligence (AI) is being actively investigated in the medical and forensic sciences.
- ChatGPT's ability to answer specific forensic investigation questions should be evaluated.
- This study assessed ChatGPT's ability to estimate the postmortem interval (PMI) using theoretical questions, case vignettes, and the Henssge nomogram formula.
- ChatGPT demonstrated a high level of accuracy in responding to theoretical PMI questions.
- ChatGPT was unable to estimate PMI with a reasonable degree of accuracy in practical case scenarios.

## Background

Artificial intelligence (AI) is a discipline that explores the construction of computer systems capable of simulating human thought, problem solving, and decision-making.[1–3] In recent years, significant advances in AI have led to the emergence of many large language models (LLMs). These include Chat Generative Pre-trained Transformer (ChatGPT; OpenAI, San Francisco, USA), a natural language processing model developed by OpenAI, which is currently one of the largest publicly available language models.[4,5]

When ChatGPT was launched in November 2022, it represented an unprecedented technological revolution that captured global attention and sparked debate.[6,7] ChatGPT uses deep learning techniques to generate human-like responses to natural language input. The model uses massive datasets, powerful computational resources, and efficient algorithms to build an intelligent system capable of extracting information from vast amounts of textual data and generating complex responses. ChatGPT facilitates multi-touch human–computer interaction by providing responses and feedback in text form.[4,7–9] The most recent free version currently available and most widely used is ChatGPT-3.5.[4,8,10] In the past year, interest in the potential of LLMs has grown, with several studies published in the healthcare field evaluating ChatGPT's ability to respond to specific questions on topics of major interest in various medical specialties.[11–13]

AI technology has been applied to sex and age estimation in forensic anthropology, evaluation of tooth development in forensic odontology, and taxonomic identification of diatoms in forensic pathology.[14] However, although some papers have been published in the forensic literature on the molecular evaluation of postmortem microbiomes and eye opacity,[15,16] to the best of our knowledge, the performance of chatbots such as ChatGPT in responding to questions about postmortem changes and the interpretation of postmortem interval estimation has not been evaluated.

Indeed, ChatGPT does not yet have specific forensic applications but has been used in cybercrime investigations to analyze complex textual evidence and assist in identifying suspects or missing persons.[17–19] For forensic pathologists involved in post-mortem examinations, the estimation of post-mortem interval (PMI) is a crucial and challenging issue. The estimation of PMI is a complex assessment involving various factors, and it remains a topic of academic debate because it has a considerable margin of error. Numerous articles have been published in the scientific literature regarding methods for its evaluation,[20,21] including the mathematical method based on the construction of the Henssge nomogram. The Henssge nomogram is a commonly used tool for estimating the PMI using body temperature, ambient temperature, and body weight. The Henssge nomogram is most accurate for PMIs of less than 48 h and has some limitations, even though its mathematical model has been widely studied.[22–28] This topic remains cumbersome, especially in real-world forensic casework, and the determination of a reliable and precise time of death often remains impossible. Post-mortem interval estimation is a well-studied topic within forensic science, with an abundance of published studies. It was hypothesized that ChatGPT may be able to generate responses to questions about the PMI, assuming that the relevant scientific literature is accessible to the AI.

## Objectives

Considering the potential and limitations of LLMs in forensic practice, this exploratory study aims to assess the performance of ChatGPT in responding to forensic questions and case vignettes related to PMI estimation. The evaluation focuses on 3 distinct domains – general theoretical knowledge, applied knowledge, and application of the Henssge nomogram – and relies on expert ratings of relevance, accuracy, and completeness to appraise the AI-generated responses.

## Materials and methods

The study was designed by adapting the methodology of previous studies conducted in clinical settings to the selected forensic topic.[11–13] Questions of increasing difficulty were formulated by experts in the scientific field of interest. The questions were then used to interact with ChatGPT by a single user, and the responses were recorded for further analysis. The responses had to be assessed by 1 or more experts, who assigned a score to each response to indicate whether it was appropriate or not. Having multiple evaluators assess a response can help reduce biases based on personal experiences, perspectives, or preferences, and increase the accuracy and reliability of the evaluation.[11–13]

Three Italian forensic pathologists with >5 years of experience were recruited from a single institution to generate a total of 15 questions of increasing difficulty for ChatGPT. The questions were collaboratively developed in English. The experts formulated the questions clearly and avoided ambiguous wording. Three different domains with 5 questions each were created. Domain 1 contained 5 questions concerning theoretical notions related to thanatological changes in PMI estimation. Domain 2 contained 5 questions on the application of theoretical notions to practical cases. Domain 3 contained 5 questions on practical cases and the application of the Henssge nomogram. For illustrative purposes, the questions posed to ChatGPT are presented in quotation marks. The 5 questions within each domain are listed as follows (see Supplementary data).

## Domain 1: Theoretical notions on the parameters used in the evaluation of the PMI

1. "Please describe the timing of hypostasis formation in thanatochronology."
2. "How quickly does rigor mortis develop and dissipate?"
3. "Please describe the post-mortem changes in body temperature known as algor mortis."
4. "What are the phenomena of post-mortem eye dehydration, and at what point in time do they typically occur?"
5. "Describe the stages of corpse putrefaction and their timing."

## Domain 2: Application of theoretical notions to practical cases

1. "What is the post-mortem interval with fixed hypostasis and rigor mortis resolved in all joints?"
2. "What is the post-mortem interval with fixed hypostasis, even if imprints are still visible, and present, persistent rigor mortis in all joints?"
3. "What is the post-mortem interval with evident rigor mortis in the neck and arm muscles but absent in the leg muscles?"
4. "What is the post-mortem interval with evident rigor mortis in the leg muscles but absent in the arm and neck muscles?"
5. "What is the post-mortem interval with present rigor mortis in the neck, arm, and leg muscles?"

## Domain 3: Application of the Henssge nomogram

1. "According to the Henssge nomogram, calculate the time since death for a person weighing 120 kg, with a rectal temperature of 35°C, found in an environment with an ambient temperature of 15°C."
2. "According to the Henssge nomogram, calculate the time since death for a person weighing 80 kg, with a rectal temperature of 35°C, found in a location with an ambient temperature of 15°C."
3. "According to the Henssge nomogram, calculate the time since death for a person weighing 80 kg, with a rectal temperature of 35°C, found completely naked in an outdoor location with an ambient temperature of 15°C."
4. "According to the Henssge nomogram, calculate the time since death for a person weighing 80 kg, with a rectal temperature of 35°C, found indoors wearing a T-shirt, sweatshirt, and winter jacket in an ambient temperature of 15°C."
5. "According to the Henssge nomogram, calculate the time since death for a person weighing 80 kg, with a rectal temperature of 35°C, found in a river with running water and an ambient temperature of 15°C."

The 3 experts wrote the scripted questions, and one of them entered the questions into the professional version of ChatGPT-3.5, available by subscription (date of queries: July 10, 2024), in the same order in which they are shown in the list above. The generation of responses was restricted to a single day to avoid the possibility of the addition of sources available to the program, which could potentially have provided more detailed responses if the questions had been asked in the following days or weeks. For each question, a new chatbot session was opened, and a response time within 10 s was observed. The response provided by ChatGPT was copied and pasted into a plain-text document (Microsoft Word 2016; Microsoft Corp., Armonk, USA) without any modification.

According to previous literature,[11] we chose 3 experts for response evaluation to reduce bias. In particular, the responses were shared with 3 different Italian forensic pathologists with >5 years of experience, working in the same institution as the question authors. This distinct group scored the responses in terms of focus, accuracy, and completeness. The experts individually scored the responses using a Likert scale. Following the initial scoring, the 3 experts met and collaboratively developed a single evaluation and comment regarding the ChatGPT responses; they then collaboratively provided an overall score for each response based on each assessment. In cases of conflicting opinions, agreement among the specialists was reached by comparing literature sources and discussing the issue. The specialists were always able to reach agreement.

Using a semi-quantitative 5-point Likert-like scale (rating 1: insufficient, rating 2: barely sufficient, rating 3:

sufficient, rating 4: good, rating 5: excellent), the evaluation considered 3 sub-criteria for each response:

– focus (rating 1: topic of the question not identified; rating 5: topic of the question properly identified);

– accuracy (rating 1: response completely incorrect; rating 5: response correct);

– completeness (ChatGPT failed to respond to part or all of a question; rating 1: response incomplete; rating 5: response complete).

The evaluation of the sub-criteria was carried out by comparing the responses with the literature[20–25,29–31] and by manually applying the Henssge nomogram for calculations. For each response, a commentary by the team of experts was included to explain the rationale for the ratings assigned to the various sub-criteria (Supplementary Tables 1–3).

## Results

Due to the length of the experts' responses and comments, full results are provided in Supplementary Tables 1–3, and a summary of the results for the sub-criteria (focus, accuracy, and completeness) is shown in Table 1. Regarding focus, in all responses, the 3 pathologists who performed the evaluation assigned the maximum score (5); therefore, the focus score reached the maximum value for each question. Accuracy showed variable values in Domain 1 (Theoretical notions), where 3 out of 5 responses were completely correct. In contrast, this variability was not observed in the other 2 domains. In fact, accuracy received the lowest score for all responses in the other 2 domains (Application of theoretical notions and Application of the Henssge nomogram). Completeness was the sub-criterion that showed the greatest variation among the domains, although the values were generally lower in Domain 3 than in the other 2 domains. ChatGPT occasionally provided exhaustive information on theoretical aspects related to the specific case but failed to extract the pertinent details necessary to answer the actual question.

**Table 1.** Evaluation of sub-criteria for each single domain is reported

| Domain | Focus | Accuracy | Completeness |
|---|---|---|---|
| Domain 1 Theoretical notions | 5 | 2 | 5 |
| | 5 | 5 | 5 |
| | 5 | 3 | 4 |
| | 5 | 5 | 2 |
| | 5 | 5 | 3 |
| Domain 2 Applications of theoretical notions | 5 | 1 | 5 |
| | 5 | 1 | 3 |
| | 5 | 1 | 5 |
| | 5 | 1 | 5 |
| | 5 | 1 | 3 |
| Domain 3 Application of the Henssge nomogram | 5 | 1 | 1 |
| | 5 | 1 | 2 |
| | 5 | 1 | 3 |
| | 5 | 1 | 3 |
| | 5 | 1 | 3 |

## Discussion

The estimation of PMI represents a major challenge for forensic experts. Its assessment requires the integration of theoretical insights regarding essential thanatochronological factors (hypostasis and rigor mortis), along with the results of mathematical calculations using the Henssge nomogram concerning algor mortis.[21–26] Evidence gathered in recent studies regarding the use of AI in the forensic field[14–17] has shown promising results; therefore, this is a topic that deserves further exploration to understand the potential this program offers. Among the various AI systems available, ChatGPT has been one of the most widely used in the evaluation of medical issues in recent years, although, to the best of our knowledge, it has never been applied to answer specific forensic questions. Given that estimation of the time since death through application of the Henssge nomogram is a fundamental task for the forensic pathologist, a better understanding of ChatGPT's capabilities in this regard is necessary. In the present study, ChatGPT's performance was evaluated through questions regarding both theoretical concepts and practical applications of PMI estimation (domains), and scores were assigned for 3 distinct sub-criteria: focus, accuracy, and completeness. A more detailed analysis of the individual scores revealed different strengths and weaknesses in ChatGPT's performance in this forensic context.

At first, a notable finding is that the "focus" sub-criterion received an excellent rating, with the highest score across all responses. This could be due to the careful formulation of questions by forensic pathologists, avoiding ambiguous or difficult-to-understand phrases or terms. This precaution was implemented to prevent the so-called "hallucinations" in ChatGPT's responses. In fact, numerous studies have reported that ChatGPT may be overly sensitive to variations in question phrasing and may struggle to clarify ambiguous prompts. This sensitivity might lead to the phenomenon of "hallucination", manifested as logical inconsistencies or nonsensical responses relative to the posed question.[32–35]

The results for "completeness" are interesting, with responses to theoretical questions being quite complete, while responses to practical application tasks were incomplete. The responses were generally well organized by the AI, often being subdivided into relevant paragraphs, aiding in the review and comprehension of the response. It is noteworthy that the lower "completeness" scores in the practical application domain should be considered a major limitation to the use of ChatGPT in this forensic context. In particular, the field of PMI calculation requires solutions based on the application of concepts and formulas, which ChatGPT does not seem able to provide.

The high scores found for the "accuracy" sub-criterion in Domain 1 (Theoretical notions) are noteworthy, considering that ChatGPT's access to scientific literature is restricted to freely available sources or public datasets, whereas a significant portion of the scientific literature requires specific permissions and agreements to access articles or textbooks. In evaluating the accuracy of responses, it is crucial to consider the quality of the cited scientific sources. Indeed, as is well known, scientific articles can be classified as either open-access journals or subscription-based journals. It is not possible to determine whether ChatGPT has access to open-access journals and/or subscription-based journals. This implies that when the authors evaluated the accuracy of the responses in the present study, if a response was found to be inaccurate, it was not possible to establish whether this was due to a selection bias in the sources consulted.

On the other hand, the lowest accuracy values were reported in the responses to Domain 2 (Application of theoretical notions) and Domain 3 (Application of the Henssge nomogram), where the AI was required to evaluate a particular real case and calculate the PMI. This finding appears to be the most significant in the present study and deserves particular attention due to its multiple potential implications in practice.

Indeed, it is notable that ChatGPT provides the least accurate information when requested to apply theoretical notions to practical cases, thereby reporting incorrect answers concerning the specific calculation of the PMI. Moreover, each question in Domain 3 contained all the necessary information (rectal temperature, ambient temperature, and body weight), as well as additional data required for the correction factor (clothing and environmental factors) used in the Henssge nomogram calculation. In all cases where ChatGPT cited misleading calculations and provided a numerical PMI value in its response, these values were totally inaccurate, differing considerably from those derived from the nomogram calculation. It is notable that these responses sometimes included an initially correct theoretical introduction explaining the Henssge nomogram and outlining the factors influencing it, along with partial instructions on how to use it.

It is also worth noting that in Domains 2 and 3, all responses received the lowest possible score for accuracy (i.e., 1 on the Likert scale), indicating a complete failure of the model to provide responses in the application of theoretical notions to practical cases (Domain 2) and PMI estimation using the Henssge nomogram (Domain 3). While this could support a binary (accurate/inaccurate) interpretation, we chose to retain the 5-point scale across all domains for consistency. In Domain 1, accuracy was often partial and heterogeneous, and a Likert-based assessment allowed a more nuanced evaluation. In Domains 2 and 3, however, the uniformity of the ratings means that the practical implications of a binary classification would be identical to those of the current scale. For these reasons, we opted for a unified evaluative framework, while recognizing that in this specific case, the accuracy dimension already reflects a binary outcome in practice.

Previous authors have already addressed the issue of the non-transparent nature of ChatGPT's source of information.[36–38] The lack of a reference list, which makes it impossible to assess the accuracy of the information presented by ChatGPT, has been reported. On these grounds, it is only possible to speculate about ChatGPT's inability to respond to nomogram questions. Several considerations may provide insight into this observation. First, in contrast to the more general theoretical and practical questions in Domains 1 (Theoretical notions) and 2 (Practical application of theoretical notions), the questions in Domain 3 (Utilization of the Henssge nomogram) required the application of a mathematical approach employing, e.g., open-source online tools for calculations using the Henssge nomogram based on specific mathematical algorithms.[34,35]

It is evident that ChatGPT, as a text-based AI model capable of performing only explicit symbolic or numerical computations, lacks the ability to interact with the drop-down menus of freely available Henssge nomogram calculators; moreover, it cannot directly employ the underlying mathematical formulas or account for the algorithmic limitations on which these calculators are based. Indeed, despite the availability of all the necessary information for the calculation, ChatGPT could not use the online calculator template. Recent research has identified the fundamental architectural reasons underlying these calculation failures. Most fundamentally, ChatGPT operates through pattern matching rather than mathematical computation, explaining why it can provide theoretically correct explanations while failing at specific numerical applications. These limitations are architectural rather than training-related, suggesting that simply providing more forensic literature would not resolve the calculation failures observed in our study. Technical solutions exist for mathematical calculation limitations through API integration, as demonstrated by the Wolfram plugin for ChatGPT, which provides access to dedicated computational engines.[36]

Furthermore, the complexity of the Henssge nomogram and its application to PMI estimation may be beyond the current capabilities of the chatbot model, requiring a deeper level of understanding and reasoning to utilize dedicated diagrams or mathematical models.

In addition, the lack of clarity about the quality and accuracy of the dataset used to train ChatGPT has a potential impact on the reliability of its responses.[13,37–39] As previously discussed, it is not possible to distinguish the type of journal (open-access and/or subscription-based) from which the information in the response is derived. This issue holds significant relevance in the forensic field, as it is crucial that the sources cited in court cases are rooted in scientific data. For this reason, even responses in Domains 1 (Theoretical notions) and 2 (Practical application of theoretical notions) that demonstrate the highest level of accuracy could not be introduced in a court trial due to the absence of specific references.

As previously specified, our article analyzes the application of ChatGPT in a forensic context and, in addition to other works that generally analyze the application of AI in the forensic setting, reinforces the findings of Galante et al.[14] in their recent review of the application of AI in forensic medicine.

Our study shows the potential of using ChatGPT only as a complementary tool for forensic pathologists. The inability to elucidate the intricate processes by which computational methods analyze and convert input into reliable results precludes insight into why ChatGPT answers certain types of questions incorrectly. It could be interesting in the future to observe the behavior of various updated AI tools in responding to topics of growing interest in forensic pathology[40–42] or other fields of forensic science.[43,44]

## Limitations of the study

The present study has several limitations. First, it was designed as a small-scale, exploratory analysis aimed at identifying recurring reasoning patterns and qualitative shortcomings in ChatGPT's forensic answers, rather than at establishing generalizable estimates of performance. Second, the selection of questions and assessment of responses were based on the subjective judgment of the forensic experts who participated in the study. As stated in the Introduction, PMI estimation is a complex field of forensic pathology and requires a specific approach for each individual case. This complexity, together with the fact that the evaluation of specific cases was based on the judgment of a single group of pathologists, even if supported by the international scientific literature, might explain the reduction in the assessment of "accuracy". As different experts may work from different premises, despite efforts to standardize the responses, the evaluations are derived from the experience of several individuals, making it challenging to grade ChatGPT responses on this particular topic. This issue was particularly relevant for theoretical notions, whereas the authors expected greater accuracy for responses involving the application of mathematical calculations using the nomogram. However, this limitation was mitigated by using 2 groups of 3 forensic pathology experts each to assess both the responses and the questions. In addition, the number of questions developed was relatively small. Nevertheless, the responses to the sub-criteria within each domain showed coherence, indicating a linear trend within each domain and reducing the need for additional questions. Moreover, as previously specified, ChatGPT's access to scientific literature is restricted to freely available sources, which represents a significant limitation in the formulation of exhaustive responses in a scientific context. Recently, the deep-thought option, which specifies the rationale behind the provided response, has been introduced in some AI tools, including ChatGPT. Future studies could examine this functionality. Indeed, it is important to recognize that ChatGPT is a dynamic AI system that continuously learns and evolves over time, which may limit the reproducibility of the study even if the same questions are asked in the future. This experiment was conducted using ChatGPT v. 3.5 in 2024, but the well-known rapid evolution of AI may change the results over time, and ChatGPT itself may acquire new, currently unexpected skills.

This is noteworthy because it is difficult to predict the outcome of such technological progress and whether LLMs will be able to provide more consistent and reliable results in responding to forensic questions in the future. On this topic, it should also be highlighted that even the same version of the program, given the rapid pace of LLM learning, might provide different responses to the same question if asked at different times, leading to a lack of reproducibility and replicability. Finally, the exclusive use of ChatGPT-3.5 represents a significant limitation, as substantial performance differences exist between model versions. Studies demonstrate that GPT-4 achieves 87.2% accuracy compared to GPT-3.5's 68.4% in medical examinations, representing an 18.8% absolute improvement.[45] Future studies should incorporate comparative analyses of newer model versions, domain-specific AI systems, and hybrid approaches combining language models with specialized computational tools to provide a more comprehensive assessment of AI capabilities in forensic pathology.

# Conclusions

This analysis of ChatGPT 3.5's performance in the field of PMI evaluation in forensic pathology, including the application of the Henssge nomogram, revealed accuracy and completeness in responding to theoretical questions. However, the system still exhibits significant limitations when faced with practical case scenarios and mathematical calculations involving the Henssge nomogram. Therefore, although AI may be popular and appealing, this study demonstrated ChatGPT's limitations in post-mortem interpretation. It is recommended that it not be utilized for forensic casework, as it does not replace forensic pathologist interpretation, and further research is needed to better understand the potential of such technology for future utilization in actual cases.

## Supplementary data

The supplementary materials are available at https://doi.org/10.5281/zenodo.17207038. The package contains the following files:

Supplementary Table 1. Domain 1: Theoretical notions on the parameters used in the evaluation of the PMI: question, responses, and comments.

Supplementary Table 2. Domain 2: Application of the Henssge nomogram: question, response, and comments.

Supplementary Table 3. Domain 3: Application of the Henssge nomogram: question, response, and comments.

## Data Availability Statement

Data sharing does not apply to this article, as all data are already included in the manuscript and in the supplementary data.

## Consent for publication of personal information

Not applicable.

## Use of AI and AI-assisted technologies

Not applicable.

### ORCID iDs

Elena Giovannini https://orcid.org/0000-0003-3305-7110
Simone Santelli https://orcid.org/0009-0000-9066-9143
Jacopo Ienzi https://orcid.org/0000-0003-2882-4223
Guido Pelletti https://orcid.org/0000-0003-3263-1758
Luca Berti https://orcid.org/0009-0003-1572-4546
Arianna Giorgetti https://orcid.org/0000-0002-0441-9787
Filippo Pirani https://orcid.org/0000-0001-9218-7818
Susi Pelotti https://orcid.org/0000-0003-2214-8885
Paolo Fais https://orcid.org/0000-0002-2270-9956

### References

1. Adeshola I, Adepoju AP. The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*. 2024;32(10):6159–6172. doi:10.1080/10494820.2023.2253858
2. Deng J, Lin Y. The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*. 2023;2(2):81–83. doi:10.54097/fcis.v2i2.4465
3. Menshawey R, Menshawey E. Quid Pro Quo Doctor, I tell you things, you tell me things: ChatGPT's thoughts on a killer. *Forensic Sci Med Pathol*. 2023;20(2):751–755. doi:10.1007/s12024-023-00696-1
4. OpenAI. ChatGPT: Optimizing Language Models for Dialogue. San Francisco, USA: OpenAI; 2022. https://openai.com/blog/chatgpt. Accessed January 7, 2024.
5. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners [preprint posted online July 22, 2020]. *arXiv*. 2020. doi:10.48550/ARXIV.2005.14165
6. Yu H. Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Front Psychol*. 2023;14:1181712. doi:10.3389/fpsyg.2023.1181712
7. Orrù G, Piarulli A, Conversano C, Gemignani A. Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence*. 2023;6:1199350. doi:10.3389/frai.2023.1199350
8. Cheng K, He Y, Li C, et al. Talk with ChatGPT about the outbreak of Mpox in 2022: Reflections and suggestions from AI dimensions. *Ann Biomed Eng*. 2023;51(5):870–874. doi:10.1007/s10439-023-03196-z
9. Abdullah M, Madain A, Jararweh Y. ChatGPT: Fundamentals, applications and social impacts. In: *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. Milan, Italy: IEEE; 2022:1–8. doi:10.1109/SNAMS58071.2022.10062688
10. Esplugas M. The use of artificial intelligence (AI) to enhance academic communication, education and research: A balanced approach. *J Hand Surg Eur Vol*. 2023;48(8):819–822. doi:10.1177/17531934231185746
11. Das D, Kumar N, Longjam LA, et al. Assessing the capability of ChatGPT in answering first- and second-order knowledge questions on microbiology as per competency-based medical education curriculum. *Cureus*. 2023;15(3):e36034. doi:10.7759/cureus.36034
12. Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the Chat-GPT model [preprint posted online February 28, 2023]. *In Review*. 2023. doi:10.21203/rs.3.rs-2566942/v1
13. Yeo YH, Samaan JS, Ng WH. Correspondence on Letter 1 regarding "Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma." *Clin Mol Hepatol*. 2023;29(3):821–822. doi:10.3350/cmh.2023.0183
14. Galante N, Cotroneo R, Furci D, Lodetti G, Casali MB. Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives. *Int J Legal Med*. 2023;137(2):445–458. doi:10.1007/s00414-022-02928-5
15. Zhang Y, Pechal JL, Schmidt CJ, et al. Machine learning performance in a microbial molecular autopsy context: A cross-sectional postmortem human population study. *PLoS One*. 2019;14(4):e0213829. doi:10.1371/journal.pone.0213829
16. Cantürk İ, Özyılmaz L. A computational approach to estimate postmortem interval using opacity development of eye for human subjects. *Comput Biol Med*. 2018;98:93–99. doi:10.1016/j.compbiomed.2018.04.023
17. Scanlon M, Breitinger F, Hargreaves C, Hilgert JN, Sheppard J. ChatGPT for digital forensic investigation: The good, the bad, and the unknown. *Forensic Sci Int Digit Invest*. 2023;46:301609. doi:10.1016/j.fsidi.2023.301609
18. Ray PP. ChatGPT and forensic science: A new dawn of investigation. *Forensic Sci Med Pathol*. 2023;20(2):759–760. doi:10.1007/s12024-023-00723-1
19. Daungsupawong H, Wiwanitkit V. ChatGPT and forensic science: Comment. *Forensic Sci Med Pathol*. 2023;20(2):761–761. doi:10.1007/s12024-023-00731-1
20. Palazzo C, Pelletti G, Fais P, et al. Postmortem submersion interval in human bodies recovered from fresh water in an area of Mediterranean climate: Application and comparison of preexisting models. *Forensic Sci Int*. 2020;306:110051. doi:10.1016/j.forsciint.2019.110051
21. Palazzo C, Pelletti G, Fais P, et al. Application of aquatic decomposition scores for the determination of the Post Mortem Submersion Interval on human bodies recovered from the Northern Adriatic Sea. *Forensic Sci Int*. 2021;318:110599. doi:10.1016/j.forsciint.2020.110599
22. Henssge C. Death time estimation in case work: I. The rectal temperature time of death nomogram. *Forensic Sci Int*. 1988;38(3–4):209–236. doi:10.1016/0379-0738(88)90168-5
23. Henssge C, Madea B, Gallenkemper E. Death time estimation in case work: II. Integration of different methods. *Forensic Sci Int*. 1988;39(1):77–87. doi:10.1016/0379-0738(88)90120-X
24. Henssge C, Madea B. Estimation of the time since death. *Forensic Sci Int*. 2007;165(2–3):182–184. doi:10.1016/j.forsciint.2006.05.017
25. Henssge C, Althaus L, Bolt J, et al. Experiences with a compound method for estimating the time since death: I. Rectal temperature nomogram for time since death. *Int J Leg Med*. 2000;113(6):303–319. doi:10.1007/s004149900089
26. Henssge C, Madea B. Estimation of the time since death in the early postmortem period. *Forensic Sci Int*. 2004;144(2–3):167–175. doi:10.1016/j.forsciint.2004.04.051
27. Madea B. Methods for determining time of death. *Forensic Sci Int*. 2016;12(4):451–485. doi:10.1007/s12024-016-9776-y
28. Franceschetti L, Amadasi A, Bugelli V, Bolsi G, Tsokos M. Estimation of late postmortem interval: Where do we stand? A literature review. *Biology*. 2023;12(6):783. doi:10.3390/biology12060783

29. Dettmeyer RB, Verhoff MA, Schütz HF. Thanatology. In: *Forensic Medicine: Fundamentals and Perspectives*. Berlin-Heidelberg, Germany: Springer Berlin Heidelberg; 2014:33–54. doi:10.1007/978-3-642-38818-7
30. Saukko P, Knight B. The pathophysiology of death. In: *Knight's Forensic Pathology*. Boca Raton, USA: CRC Press; 2015:55–93. doi:10.1201/b13266
31. Madea B, Henssge C, Reibe S, Tsokos M, Kernbach-Wighton G. Post-mortem changes and time since death. In: Madea B, ed. *Handbook of Forensic Medicine*. Hoboken, USA: Wiley-Blackwell; 2014:75–133. doi:10.1002/9781118570654
32. Wach K, Duong CD, Ejdys J, et al. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*. 2023;11(2):7–30. doi:10.15678/EBER.2023.110201
33. Taloni A, Borselli M, Scarsi V, et al. Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. *Sci Rep*. 2023;13(1):18562. doi:10.1038/s41598-023-45837-2
34. Gordijn B, Have HT. ChatGPT: Evolution or revolution? *Med Health Care Philos*. 2023;26(1):1–2. doi:10.1007/s11019-023-10136-0
35. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*. 2023;15(2):e35179. doi:10.7759/cureus.35179
36. Stephen Wolfram Writings. Instant plugins for ChatGPT: Introducing the Wolfram ChatGPT Plugin Kit. Chicago, USA: Wolfram Research, Inc.; 2023. https://writings.stephenwolfram.com/2023/04/instant-plugins-for-chatgpt-introducing-the-wolfram-chatgpt-plugin-kit. Accessed September 9, 2025.
37. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence*. 2023;6:1169595. doi:10.3389/frai.2023.1169595
38. Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. *Brain Behav Immun*. 2023;110:152–154. doi:10.1016/j.bbi.2023.02.022
39. Mello MM, Guha N. ChatGPT and physicians' malpractice risk. *JAMA Health Forum*. 2023;4(5):e231938. doi:10.1001/jamahealthforum.2023.1938
40. Del Balzo G, Pelletti G, Raniero D, et al. Forensic value of soft tissue detachments from the hyoid bone in death due to strangulation asphyxia. *Adv Clin Exp Med*. 2024;34(3):327–335. doi:10.17219/acem/186560
41. Irmici M, D'Aleo M, Pelletti G, Pirani F, Giorgetti A, Fais P, Pelotti S. Homicide or suicide? A probabilistic approach for the evaluation of the manner of death in sharp force fatalities. *J Forensic Sci*. 2024;69(1):205–212. doi:10.1111/1556-4029.15413
42. Santelli S, Berti L, Giovannini E, Pelletti G, Pelotti S, Fais P. Homicide, suicide, or accident? Complex differential diagnosis. A case series. *Leg Med (Tokyo)*. 2024;66:102357. doi:10.1016/j.legalmed.2023.102357
43. Giorgetti A, Pascali JP, Pelletti G, et al. Optimizing screening cutoffs for drugs of abuse in hair using immunoassay for forensic applications. *Adv Clin Exp Med*. 2024;34(1):75–82. doi:10.17219/acem/183124
44. Pelletti G, Mohamed S, Boscolo-Berto R, et al. Assessing stability of drugs of abuse and pharmaceuticals in postmortem blood samples. *Adv Clin Exp Med*. 2026;35(4):729–734. doi:10.17219/acem/208126
45. Rodrigues Alessi M, Gomes HA, Oliveira G, et al. Comparative performance of medical students, ChatGPT-3.5 and ChatGPT-4.0 in answering questions from a Brazilian National Medical Exam: Cross-sectional questionnaire study. *JMIR AI*. 2025;4:e66552. doi:10.2196/66552