Surgical Science, 2025, 16(7), 365-377
https://www.scirp.org/journal/ss
ISSN Online: 2157-9415
ISSN Print: 2157-9407

A Prototype AI Surgical Assistant for Real-Time Consultation during Laparoscopic Surgery

Savvas Hirides[1,2], Petros Hirides[3], Kalliopi Kouloufakou[3], Constantinos Hirides[3]

[1]Medical School, University of Nicosia, Nicosia, Cyprus
[2]Athens Medical Center, Athens, Greece
[3]IASO Group of Hospitals, Athens, Greece
Email: savashiridis@gmail.com

How to cite this paper: Hirides, S., Hirides, P., Kouloufakou, K. and Hirides, C. (2025) A Prototype AI Surgical Assistant for Real-Time Consultation during Laparoscopic Surgery. *Surgical Science*, **16**, 365-377.
https://doi.org/10.4236/ss.2025.167037

Received: June 19, 2025
Accepted: July 28, 2025
Published: July 31, 2025

Abstract

Recent advancements in generative AI and large language models (LLMs) have sparked new opportunities in surgical innovation. We present our prototype AI Surgical Assistant Prototype System, integrating real-time vision support for streaming video of the operative field, advanced speech recognition, multilingual natural voice interaction, both long- and short-term memory simulation, and customizable behavioral profiles. Testing showed that the system exhibited contextual awareness from visual feedback in 38% of instances, ability for verbal interaction, and dynamic memory logging throughout the procedure. This feasibility study suggests that real-time, context-aware AI support is technically viable in the OR and may serve as a basis for future clinical models.

Keywords

AI Surgical Assistant, Real-Time Video Analysis, GPT-4o in Surgery, AI in the OR, Voice-Activated Intraoperative Support

1. Introduction

The integration of AI into surgical practice has moved beyond passive decision support into the realm of interactive, real-time assistants. Large Language Models (LLMs) such as OpenAI's GPT-4 can now synthesize complex information, recognize visual patterns, and interact via speech. Recent publications underscore the expanding versatility of AI in surgical practice, suggesting a transformative potential in operative workflows [1]-[10].

We hypothesized that an AI assistant integrating GPT-based dialogue, voice recognition, and visual feedback can function in real-time during laparoscopic surgery without disrupting standard workflow. To study this hypothesis, we formulated a prototype AI Surgical Assistant Model called GePpeTto, integrating GPT-based dialogue, advanced speech recognition, and real-time vision support via streaming video of the operative field. This system enables contextual awareness, verbal interaction, memory logging, and visual feedback in the OR environment and may be used as an initial model for further clinical research. It uses three basic modules which interact with each other to leverage GPT-4's multimodal capabilities alongside natural speech recognition and actual conversation, with simultaneous direct video feedback from the laparoscopic surgery tower. We outline its architecture, implementation, use with streaming laparoscopic video, as well as drawbacks and challenges. We also provide a brief review of current literature and comparison to similar projects.

2. Materials and Methods

System Architecture: GePpeTto stands in the core of the application as a standard GPT-4o API Language Model handling both language processing and image interpretation. The model behaves similarly to standard ChatGPT but incorporates predesigned traits embedded in its system persona. Through its command listing it calls Whisper, an advanced speech interface for real-time voice command recognition and transcript generation, in order to communicate with the surgeon. Upon initiation it goes through relative literature (surgical textbooks), as well as his previous notes in the form of simple log files, aiming at creating a virtual long-term memory. During the conversation it is designed to keep notes and run back to them, in order to simulate short-term memory and maintain contextual awareness within the OR. For visual feedback it consults small reports generated by an external vision GPT-API module which runs independently in the background as a second program and analyzes every snapshot and store its analysis in a separate log file. Finally, the last module (third program running at the background) streams the laparoscopic video in VLC and automatically takes snapshots every 5 seconds. A 5-second snapshot interval was selected as a balanced compromise between temporal resolution and system performance. This frequency was sufficient to capture meaningful changes in the operative field without overwhelming the vision processing pipeline or introducing significant latency. Preliminary tests with faster intervals (e.g., 1-second sampling) resulted in noticeably increased GPU load and processing delays, particularly during simultaneous voice interactions and log updates. The 5-second interval allowed for smooth operation, consistent visual analysis, and uninterrupted surgeon-AI communication, making it the most stable configuration during early testing.

Dry Lab testing: First 16 minutes of a single, anonymized, pre-recorded laparoscopic video of a routine elective cholecystectomy, were used for evaluation. The video had been previously recorded intraoperatively using the OR tower system and reflected standard procedural flow.

Hypothetical OR Setup: Although GePpeTto was tested outside the OR with pre-recorded laparoscopic video, the following standard OR setup would be preserved in real-life surgery. Addition of an auxiliary computer system for the project wouldn't alter OR organization and function. The surgeon would wear a single Bluetooth earbud from where they could communicate with the AI and continue operating undisturbed (Figure 1). The AI is programmed not to speak unless prompted. The system would be initialized by the OR staff upon insertion of the laparoscope.

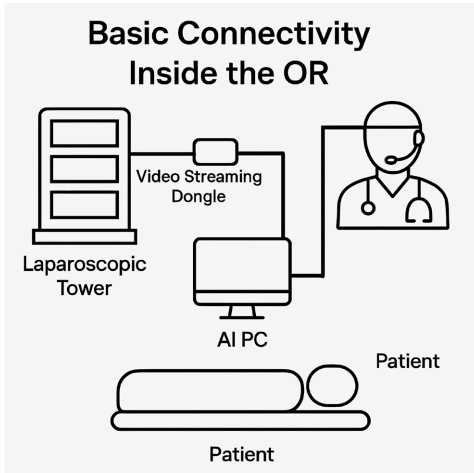

Figure 1. Basic setup of GePpeTto AI surgical assistant prototype system.

Customizable Behavioral Profiles (Persona Setup) The behavior of GPT within GePpeTto software is modulated by several key components designed to simulate a surgical assistant (Table 1). Here's a structured breakdown of the most critical elements:

Table 1. Modulation mechanisms of AI surgical assistant.

Component	Purpose	Simulates…
Personality Design	Establishes behavior of GPT via initial LLM Prompt	Assistant's tone, precision, and calmness
Vision log	Injects visual data descriptions into prompt	Visual field awareness
Memory Accumulation	Stores accumulated facts and updates in notes log text file	Short-term intraoperative memory
Extraction of important lines	Parses key phrases from replies of surgeon	Selective memory encoding
Summary	Condenses long notes	Forgetting irrelevant or outdated data
Ask GPT	Fuses all contexts into an intelligent, situational prompt	Adaptive context integration

1) Role Initialization and Personality (LLM Prompting)

Full prompt in code: {“role”: “system”, “content”: “You are Geppetto, a surgical AI assistant. Be calm, accurate, and helpful.”} This sets the initial persona of GPT as a calm, accurate, and helpful surgical assistant. The “system” prompt guides the model to avoid verbosity and focus on clinical relevance. Acts as the core behavioral modulator of GPT during all dialogue.

Dynamic Prompt Engineering with Contextual Awareness

full_prompt in code: “You are assisting in a surgical operation. This is the current video log:\n{vision_context}. These are your notes so far:\n{notes_context}. Now the surgeon asks: “Please answer and update your internal notes if needed.”

2) Visual Awareness

Visual awareness is simulated by reading from a vision log file. Context memory is accumulated from GePpeTto_notes.txt. The assistant’s replies consider prior visual findings and notes. This is crucial to simulate intraoperative continuity.

3) Memory Accumulation and Learning

After every GPT reply, the system autoparses important surgical insights (lines with keywords like note, conclude, important). These are logged as new memory entries, helping the assistant develop a cumulative understanding of the case.

4) Surgeon Interaction via Whisper & Text-to-Speech record_audio() → transcribe_audio() → ask_gpt() → speak()

Whisper software transcribes spoken surgeon input. GPT responds using surgical reasoning and context. RHVoice software speaks the answer aloud through customized voice.

5) Summarization Trigger for Long Notes

When memory grows too long, the system condenses it using GPT itself. This mimics a human assistant keeping only the salient facts in mind during long surgeries.

6) AskGPT

Fuses all contexts into an intelligent, situational prompt

Hardware - Software Implementation Tools: Both software development and testing took place using:

- an Intel® Core™ i7-4790 CPU
- 16GB RAM
- GPU GeForce GTX 970 4GB WINDFORCE
- Ubuntu/Linux 20.04.6LTS (64-bit)
- Python 3.13 for Linux
- OpenAI GPT API (with official key from OpenAI)
- Whisper (Medium Pack)
- and VLC & VLC Python bindings (via python-vlc).

Ethical Considerations: This study was conducted using a single de-identified, pre-recorded laparoscopic surgical video and real-time surgeon-AI voice interactions outside of a live clinical setting. No patient-identifiable information was used at any stage of development or testing. All patients sign a written consent to use of their laparoscopic video for research purposes in our practice, as long as their identity remains hidden. All procedures conform to the ethical standards of our institutional research committee.

3. Results

The test was terminated just after AI had recognized the type of the operation (16 minutes of total interaction time), (Figure 2). Up to now there was successful implementation of:

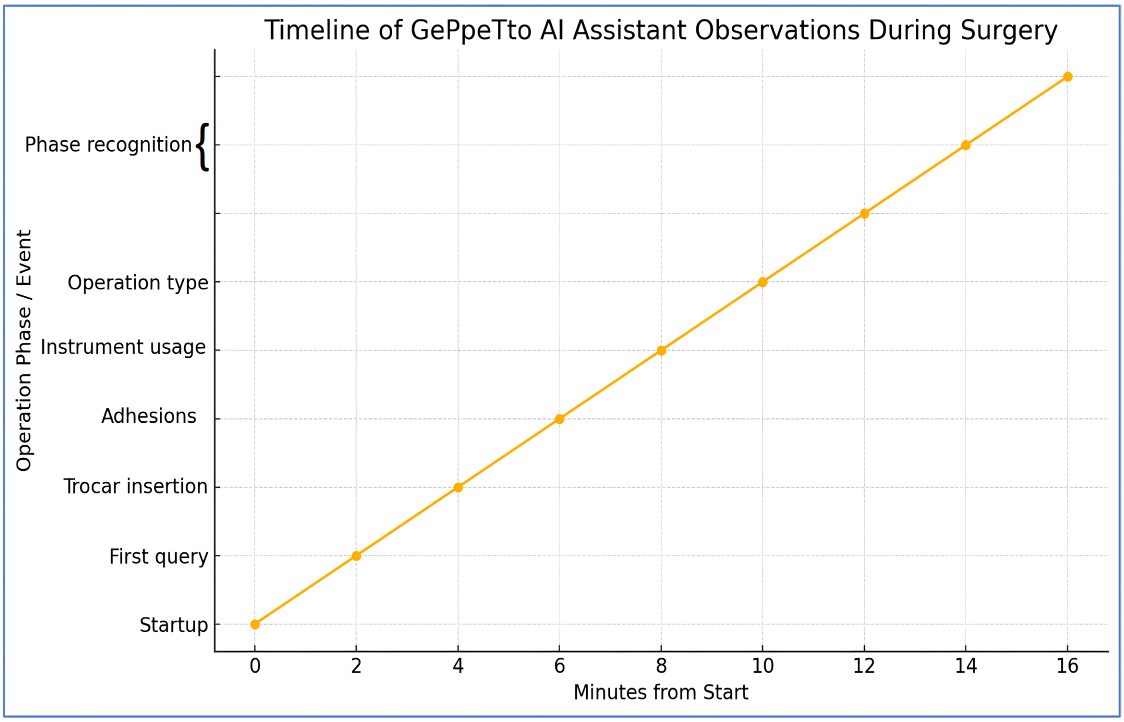

Figure 2. Timeline of GePpeTto AI assistant observations during surgery.

Real-time Video Integration: VLC retrieved laparoscopic snapshots every 1 - 5 seconds. Snapshots were sent to GPT-4 Vision API and responses were logged to file. Image interpretation managed to synthesize relative anatomy and instrument use.

Conversational Control: Wake word and command processing were fully functional. Whisper Voice Recognition accurately interpreted 100% of verbal commands from a non-native English speaker. A list of surgeon's calls to the AI can be found in Table 2. Pauses between input and output were implemented to give time to image analysis. Conversational memory was enabled through log parsing and summarization.

System Responsiveness: AI successfully summarized last known events (from vision logs) upon surgeon's request. Responded naturally to all surgeon's prompts (Table 2, Table 3).

Observations: AI in the form of ChatGPT led us efficiently throughout the process of materializing the preliminary phase of this project. It offered code snippets and quickly supplied all necessary technical solutions to matters of hardware and software functionality. The primary objective of this study was to assess whether the AI assistant could interpret live laparoscopic video, maintain intraoperative contextual memory, and respond meaningfully to surgeon prompts in real time. GePpeTto identified correctly (without giving him any hints), that we were performing a laparoscopic cholecystectomy within the first 10 minutes of the procedure (**Figure 2**).

Table 2. List of full surgeon's clues and relation to AI responsiveness.

	Surgeon's Triggers	Relative AI awareness
1	"Hello"	AI is only informed from its traits that "he will assist to a surgical operation"
2	"What did you see?"	AI realizes that the screen shows "an endoscopic or surgical perspective inside a body cavity or organ"
3	"This is a laparoscopic operation in the beginning"	AI recognizes the presence of intraabdominal adhesions. Also it notes that attention to the screen "highlights specific areas like the liver, gallbladder, and bowel involvement, with visible adhesions and tissue interaction". However, it did not used the word "laparoscopic" or "abdominal" before this stage.
4	"This is the phase of trocar insertion"	AI already mentioned "small incisions" but has not used the word "trocars" yet.
5	"There are unusually many adhesions in this patient"	Correctly comments "it's likely these are due to previous surgeries, inflammation, or infections"
6	"What is this operation you think?"	AI realizes "adhesiolysis" taking place and mentions "liver, gallbladder or intestines" as possible target organs, although these terms have not been implied by surgeon.
7	"What instruments do I use here?"	Mentions graspers and electrocautery, but also many other instruments not used at the specific phase of procedure (scissors, irrigation). Seems to be answering by general knowledge alone.
8	"We have created a specific exposure. What operation is it?"	AI correctly suspects "adhesiolysis", "liver and gallbladder", "exploratory laparoscopy". Even "bowel resection" possibility cannot be considered wrong, if it takes into account possible rare complications.
9	"Which organ is grabbed now from laparoscopic instruments?"	AI clearly states "the current focus involves manipulation or dissection of the gallbladder from the liver" and further reports correctly that "… the gallbladder is being dissected from the liver", immediately ending to the conclusion that the context is "possibly indicating a cholecystectomy"

Table 3. GePpeTto intraoperative conversation milestones.

Minutes from AI boot	Phase of Procedure	AI answers in simple questions ("What do you see?" Or "What procedure is this?")	AI awareness/ Correct Replies
2:00	Initial Inspection	…Hello, Doctor. How can I assist you in the current surgical operation? Let me know if you need guidance on tissue identification, procedural reminders, or anything else pertaining to the images or the surgery. …indicate the presence of adhesions within the abdominal cavity, which may be a focal point for procedures during this surgery. …also highlight specific areas like the liver, gallbladder, and bowel involvement, with visible adhesions and tissue interaction	Abdominal Surgery Context
4:00	Trocar Insertion	The operation is currently in the trocar insertion phase, crucial for establishing instrument access in laparoscopic surgery. • Port placement is essential for optimal visualization and maneuverability within the abdominal cavity. • Care should be taken to avoid injury to internal organs and vessels during trocar placement.	Trocar Insertion (revealed verbally from surgeon)
6:00	Adhesiolysis	Based on the extensive video log and descriptions of the surgical images, it appears that you are performing a laparoscopic procedure, possibly involving the gastrointestinal tract, with a focus on addressing adhesions in the abdominal cavity. The presence of multiple adhesions suggests that the surgery might be aimed at their removal (adhesiolysis) to alleviate pain or prevent bowel obstruction … Possible exploration for pathologies involving the liver, gallbladder, or intestines	Adhesiolysis
8:00	Gallbladder identification	Based on the descriptions of the surgical scenario involving extensive adhesions and possibly inflamed or infected tissue in a laparoscopic setting, the following instruments would be beneficial: • **Laparoscopic Scissors:** For cutting tissue or releasing adhesions efficiently. … • **Laparoscopic Forceps (graspers):** For holding and manipulating tissue safely, allowing separation and repositioning without causing additional trauma. • **Laparoscopic Suction/Irrigation:** To maintain a clear field by removing fluids, blood, or debris, and to irrigate areas to minimize the risk of infection. • **Laparoscopic Electrocautery or Harmonic Scalpel:** To assist with cutting and coagulation, minimizing bleeding while removing adhesions. • **Laparoscope:** To provide a clear and magnified view of the surgical area, crucial for precise work within the abdominal cavity. • **Laparoscopic Retractors:** To keep tissues out of the way and provide a better view of the operative field.	Instruments Used (answered correctly based on general knowledge, not from visual information)

Continued

10:00	Gallbladder Retraction	Based on the extensive series of images and descriptions provided, the operation seems to be focused on the abdominal cavity with significant attention to adhesions, likely involving procedures related to: 1) **Adhesiolysis:** The presence of multiple images showing adhesions strongly suggests that part of the procedure involves dissecting and removing these adhesions to alleviate symptoms or prevent complications like bowel obstruction. 2) **Liver and Gallbladder:** Several images indicate involvement of the liver and gallbladder, which could suggest a cholecystectomy (gallbladder removal) or liver biopsy if abnormal tissue is present.	Operation type identified
12:00	Hemostasis	There appears to be inflammation or changes in the area, possibly indicating a cholecystectomy (removal of the gallbladder) or exploration due to adhesions or other complications. - Attention is being paid to carefully handle tissue to minimize bleeding and preserve surrounding structures.	Hemostasis identified
14:00	Partial mobilization of gallbladder from liver bed	The laparoscopic instruments are likely manipulating the gallbladder, as images like snap_126.png and snap_127.png suggest the gallbladder is being dissected from the liver.	Gallbladder Separation from Liver Bed identified

Evaluation of verbal output: A phrase was labeled as "context-aware" if it incorporated one of the following: surgical tool recognition, current phase of operation, anatomic reference, or intraoperative reasoning. Two human raters independently scored all responses; disagreements were resolved by consensus. The 38% metric reflects only those responses judged to have inferred correct visual context based on image history alone, without surgeon cues.

Quantitative Results: Our 16 minute conversation yielded 9 answers from AI which could be further breakdown into 18 paragraphs. Out of these 18 paragraph-level response units, 7 were unanimously scored as context-aware, resulting in a context-aware rate of 38%. These paragraphs demonstrated contextual integration of both speech prompts and corresponding visual analysis, referencing specific instruments, anatomy, or operative steps based on earlier observations or visual reports. Classification of these paragraphs according to their relation to AI responses can be found in Table 4.

Limitations: Interrupt Functionality: Mid-sentence speech interruption was attempted at first. It was supposed to halt AI instantly and reset it to standby when the surgeon wanted to clarify something or simply skip long texts of AI commentary. However, it was abandoned for later revisit, because it often led to freezing of the program.

Rigid or detached responses: Due to API constraints from the manufacturers, the intraoperative AI assistant was more rigid than anticipated, with obvious lack of emotion (enthusiasm, encouragement, humor quality) and a disturbing lack of interest. The assistant often reported contextually adjacent but clinically non-applicable information, analogous to a junior trainee with theoretical knowledge but limited contextual judgment. This made a strong contrast to classic chatGPT which presents with higher standards of cognition, presence, and co-experience.

Table 4. Quantitative evaluation.

Realization of	OR context	Laparoscopic context	Trocar Insertion	Adhesiolysis	Cholecystectomy	Instruments	Phase of Procedure	Total Replies
Nr of completely irrelevant replies								
Nr of relative replies due to software preprogrammed traits (not acquired during surgery), general knowledge or surgeon's guidance during conversation	1		1					11%
Nr of wrong assumptions within context					1	1		11%
Nr of replies implying awareness					6			33%
Nr of replies with full context awareness		3	1	4				38%
%								
Total	18							

Anticipation and Perceptual Bias: While analyzing the AI assistant's behavior, it became evident that certain contextual inferences were formulated by the system before explicit verbal cues were given by the surgeon (Table 2). For instance, during the initial phases of dissection, the assistant referenced "the inside view suggests a close focus on either a cavity or organ", "clear images of the intestine and surrounding tissues", prior to the operator mentioning "laparoscopic" or "trocar". Although GePpeTto had no direct access to telemetry or operative metadata, its multimodal exposure (video, language prompt history, and prior conversational memory) enabled it to anticipate procedural steps and highlight areas of interest. This phenomenon, previously observed in simulated environments, was now confirmed under realistic operating conditions. The assistant's partial yet increasingly accurate guesses often aligned with the real focus of surgical activity—later validated by verbal confirmations from the surgeon. This pattern suggests a nontrivial emergence of perceptual bias toward procedural relevance, which, while sometimes premature or incorrect, demonstrates proto-clinical judgment rather than random output. Such behavior indicates that GePpeTto is not merely reactive, but begins to display anticipatory awareness, a hallmark of effective surgical assistance. This property also implies that reinforcement training and model refinement could yield significant gains in contextual precision. While the assistant does not yet "understand" the procedure in a human sense, its early inferences reveal that low-level probabilistic modeling of visual-language interaction may already simulate aspects of clinical anticipation.

4. Discussion

GePpeTto represents an early implementation of an AI surgical assistant. Unlike fixed automation protocols, the system builds a narrative of the procedure using sequential vision and conversation. This provides a form of "situational memory" that enhances intraoperative decision support. While not autonomous, GePpeTto embodies an augmented human-in-the-loop framework.

Key challenges include: temporal image interpretation limitations (no true frame-by-frame video analysis yet) and need for multimodal memory consolidation across text, speech, and image inputs. Nonetheless, this project clearly demonstrates that a fully Python-based AI assistant can operate within OR constraints without requiring complex robotics or expensive external hardware, thereby opening the door for widespread experimentation and deployment.

Similar Ongoing Projects: In 2024 Chen, Luo *et al.* [11] presented their plan to build an intelligent and versatile surgical assistant expected to accurately understand the surgeon's intentions and accordingly conduct the specific tasks to support the surgical process. In this work, by leveraging advanced multimodal large language models (MLLMs), the authors proposed a Versatile Surgery Assistant (VS-Assistant) [12] that can accurately understand the surgeon's intention and complete a series of surgical understanding tasks, e.g., surgical scene analysis, surgical instrument detection, and segmentation on demand. Specifically, to achieve superior surgical multimodal understanding, they devised a mixture of projectors (MOP) module to align the surgical MLLM in VS-Assistant to balance the natural and surgical knowledge. Moreover, they devised a surgical Function-Calling Tuning strategy to enable the VS-Assistant to understand surgical intentions, and thus make a series of surgical function calls on demand to meet the needs of the surgeons. Extensive experiments on neurosurgery data confirmed that VS-Assistant can understand the surgeon's intention more accurately than the existing MLLM. Last December, Wu, Liang *et al.* [13] presented SurgBox project, ie an agent-driven sandbox framework to systematically enhance the cognitive capabilities of surgeons in immersive surgical simulations. Specifically, their SurgBox leveraged large language models (LLMs) with tailored Retrieval-Augmented Generation (RAG) to authentically replicate various surgical roles, enabling realistic training environments for deliberate practice. In particular, they devised Surgery Copilot, an AI-driven assistant to actively coordinate the surgical information stream and support clinical decision-making, thereby diminishing the cognitive workload of surgical teams during surgery. By incorporating a novel Long-Short Memory mechanism, their Surgery Copilot could effectively balance immediate procedural assistance with comprehensive surgical knowledge. Extensive experiments using real neurosurgical procedure records validated SurgBox framework in both enhancing surgical cognitive capabilities and supporting clinical decision-making. Tobias and Paschali *et al.* introduced OperA [14], a transformer-based model that accurately predicted surgical phases from long video sequences. A novel attention regularization loss encourages the model to focus on high-quality frames during training. Moreover, the attention weights are utilized to identify characteristic high attention frames for each surgical phase, which could further be used for surgery summarization. OperA was thoroughly evaluated on two datasets of laparoscopic cholecystectomy videos.

GePpeTto distinguishes itself through its integration of speech recognition, visual analysis, and contextual memory within a real-time framework. SurgBox team also fashioned their own mechanism to establish long-short memory, as we also did with the project in this paper. Unlike systems like OperA, which focus solely on video-based phase recognition, or ActivSight [15], which enhances imaging without interactive capabilities, GePpeTto seems to offer a more holistic assistant experience. While VS-Assistant and SurgBox Copilot showcase advanced functionalities like function-calling and cognitive support, they remain in research or simulation phases without clinical deployment. They also focused on the more structured field of neurosurgery which offers easy localization fiducials (bony framework) to compare to preoperative information from imaging modalities. GePpeTto differs because it is challenged within an actual streaming laparoscopic setting, which can never be fully anticipated and remains constantly alternating even for human surgeons (**Table 5**).

Table 5. Comparative analysis of AI surgical assistants today.

System	Modality	Core Capabilities	Real-Time Interaction	Multimodal Input	Clinical Deployment
GePpeTto	LLM-based (GPT-4o), Whisper	Voice-guided assistance, vision snapshot analysis, contextual memory logging	✅	✅	✅ (Prototype stage)
VS-Assistant	Multimodal LLM with surgical tuning	Scene analysis, instrument detection, function-calling based on surgeon intent	✅	✅	❌ (Research phase)
SurgBox Copilot	LLM with RAG and memory modules	Simulated surgical training, cognitive support, decision-making assistance	✅	✅	❌ (Simulation only)
OperA	Transformer-based vide analysis	Surgical phase recognition from video sequences	❌	✅ (Visual only)	❌ (Offline analysis)
ActivSight	Al-enhanced imaging device	Real-time vessel identification during laparoscopic procedures	✅	✅ (Visual only)	✅ (Commercial use)

5. Design for Future Expansions

Use in actual OR setting: This proof-of-concept study was conducted using prerecorded video outside of the live OR. Future phases will include real-time testing in a simulated OR environment, followed by monitored use during actual laparoscopic procedures.

Optimization: GePpeTto needs to become faster, situational-aware and more human-like and should offer answers and hints that would be welcomed from the surgical team. Dynamic adjustment of waiting intervals and snapshot frame rates may optimize performance further. Structured memory and context cleaning may also be further optimized. In addition, attempts for contextual learning of AI Surgical Assistant via reinforcement or expert correction are an interesting field for further research. To establish that we intend to use video overlay software that will enable the surgeon to illustrate anatomy and pathology landmarks on screen at real-time, but also enable AI to show these structures on screen upon request.

Integration with EMR data of the patient: This is a mandatory next step because it may assist the model seeking for the correct anatomy and pathology during surgery, but also taking into consideration OR times, positioning, amount of blood loss and weigh their importance according to tailored patient needs.

Ambient awareness: Our system is not restricted to the OR context alone. It could be extended to pre-, intra-, and postoperative care, following the surgical workflow. In this way, with proper equipment, AI could accompany the surgeon during patient history-taking, examining the patients pre and postoperatively, even when preparing the patient and docking the robot for an operation. And it could be learning from all these processes. Ambient awareness may affect AI behavior intraoperatively in ways that we cannot imagine yet.

6. Conclusion

Our prototype suggests that an AI-based surgical assistant model is feasible and can be safely tested within actual OR conditions. It can already integrate speech, hearing, vision, short-term and long-term memory functions, and adopt programmable behavior patterns according to the surgical team preferences. This paper invites further exploration into how generative AI can coexist with the surgical team—not as a replacement, but as a context-aware assistant that learns and adapts in real time, with the aim of advising the surgeon, enhancing the optimal outcome for the patient and preventing complications.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Hashimoto, D.A., Rosman, G., Rus, D. and Meireles, O.R. (2018) Artificial Intelligence in Surgery: Promises and Perils. *Annals of Surgery*, **268**, 70-76. https://doi.org/10.1097/sla.0000000000002693

[2] Maier-Hein, L., Vedula, S.S., Speidel, S., Navab, N., Kikinis, R., Park, A., *et al.* (2017) Surgical Data Science for Next-Generation Interventions. *Nature Biomedical Engineering*, **1**, 691-696. https://doi.org/10.1038/s41551-017-0132-7

[3] OpenAI (2023) GPT-4 Technical Report. arXiv:2303.08774. https://arxiv.org/abs/2303.08774

[4] Radford, A., Kim, J., Xu, T., *et al.* (2022) Robust Speech Recognition via Large-Scale Weak Supervision. Whisper by OpenAI. https://openai.com/research/whisper

[5] Chen, J., Zhu, D., Shen, X., Li, X., Liu, Z., Zhang, P., Krishnamoorthi, R., Chandra, V., Xiong, Y. and Elhoseiny, M. (2023) Minigpt-v2: Large Language Model as a Unified Interface for Vision-Language Multi-Task Learning. arXiv:2310.09478.

[6] Chen, Z., Guo, Q., Yeung, L.K.T., Chan, D.T.M., Lei, Z., Liu, H., *et al.* (2023) Surgical Video Captioning with Mutual-Modal Concept Alignment. In: Greenspan, H., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2023, Springer, 24-34. https://doi.org/10.1007/978-3-031-43996-4_3

[7] Chiang, W.L., Li, Z., Lin, Z., Sheng, Y., Wu, Z., Zhang, H., Zheng, L., Zhuang, S., Zhuang, Y., Gonzalez, J.E., Stoica, I. and Xing, E.P. (2023) Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality. https://lmsys.org/blog/2023-03-30-vicuna/

[8] Das, A., Khan, D.Z., Williams, S.C., Hanrahan, J.G., Borg, A., Dorward, N.L., *et al.* (2023) A Multi-Task Network for Anatomy Identification in Endoscopic Pituitary Surgery. In: Greenspan, H., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2023, Springer, 472-482. https://doi.org/10.1007/978-3-031-43996-4_45

[9] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H. (2020) NNU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nature Methods*, **18**, 203-211. https://doi.org/10.1038/s41592-020-01008-z

[10] Szeliski, R. (2010) Computer Vision: Algorithms and Applications. Springer Science & Business Media. https://doi.org/10.1007/978-1-84882-935-0

[11] Chen, Z. and Luo, X. (2024) VS-Assistant: Versatile Surgery Assistant on the Demand of Surgeons. arXiv: 2405.08272

[12] Hirides, S., Hirides, P., Kalliopi, K. and Hirides, C. (2024) Artificial Intelligence and Computer Vision during Surgery: Discussing Laparoscopic Images with ChatGPT4—Preliminary Results. *Surgical Science*, **15**, 169-181. https://doi.org/10.4236/ss.2024.153017

[13] Wu, J., Liang, X., Bai, X. and Chen, Z. (2024) SurgBox: Agent-Driven Operating Room Sandbox with Surgery Copilot. arXiv:2412.05187.

[14] Czempiel, T. and Paschali, M. (2021) OperA: Attention-Regularized Transformers for Surgical Phase Recognition. arXiv:2103.03873.

[15] Mehrotra, S., Liu, Y.Z., Nwaiwu, C.A., Buharin, V.E., Stolyarov, R., Schwaitzberg, S.D., *et al.* (2023) Real-Time Quantification of Bowel Perfusion Using Laparoscopic Laser Speckle Contrast Imaging (LSCI) in a Porcine Model. *BMC Surgery*, **23**, ArticleNo. 261. https://doi.org/10.1186/s12893-023-02161-w