

# Artificial Intelligence and Computer Vision during Surgery: Discussing Laparoscopic Images with ChatGPT4—Preliminary Results

Savvas Hirides<sup>1,2</sup>, Petros Hirides<sup>3</sup>, Kouloufakou Kalliopi<sup>3</sup>, Constantinos Hirides<sup>3</sup>

<sup>1</sup>Department of Surgery, Medical School, University of Nicosia, Nicosia, Cyprus

<sup>2</sup>Department of Laparoscopic & Robotic Surgery, Athens Medical Center, Athens, Greece

<sup>3</sup>Department of Laparoscopic & Robotic Surgery, IASO Maternity & Gynecology Clinic, Athens, Greece

Email: savashiridis@gmail.com

**How to cite this paper:** 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>

**Received:** January 15, 2024

**Accepted:** March 26, 2024

**Published:** March 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

**Introduction:** Ultrafast latest developments in artificial intelligence (AI) have recently multiplied concerns regarding the future of robotic autonomy in surgery. However, the literature on the topic is still scarce. **Aim:** To test a novel AI commercially available tool for image analysis on a series of laparoscopic scenes. **Methods:** The research tools included OPENAI CHATGPT 4.0 with its corresponding image recognition plugin which was fed with a list of 100 laparoscopic selected snapshots from common surgical procedures. In order to score reliability of received responses from image-recognition bot, two corresponding scales were developed ranging from 0 - 5. The set of images was divided into two groups: unlabeled (Group A) and labeled (Group B), and according to the type of surgical procedure or image resolution. **Results:** AI was able to recognize correctly the context of surgical-related images in 97% of its reports. For the labeled surgical pictures, the image-processing bot scored 3.95/5 (79%), whilst for the unlabeled, it scored 2.905/5 (58.1%). Phases of the procedure were commented in detail, after all successful interpretations. With rates 4 - 5/5, the chatbot was able to talk in detail about the indications, contraindications, stages, instrumentation, complications and outcome rates of the operation discussed. **Conclusion:** Interaction between surgeon and chatbot appears to be an interesting frontend for further research by clinicians in parallel with evolution of its complex underlying infrastructure. In this early phase of using artificial intelligence for image recognition in surgery, no safe conclusions can be drawn by small cohorts with commercially available software. Further development of medically-oriented AI software and clinical world awareness are expected to bring fruitful information on the topic in the years to come.

---

## Keywords

Artificial Intelligence, Surgery, Image Recognition, Autonomous Surgery

---

## 1. Introduction

Robotic surgery has changed dramatically the way we used to think and operate during the past 24 years. Since the initial emergence of surgical robots in the operating theaters, issues about robotic autonomy in the future of surgery were raised almost in all surgical society discussions, including concerns whether a robot would be autonomously performing surgery someday, with full potential to gradually replace surgeons. Ultrafast latest developments in artificial intelligence (AI) have recently multiplied these concerns, as it seems that robotic autonomy is closer than ever before. Lately, chatbots able to code in any programming language, play chess, suggest financial and marketing strategies as well as analyze psychological problems have made their appearance as potentially disruptive game changers. Some of them can manipulate documents, lyrics, compose music and even control small or larger robots. Literature is rather scarce about the potential of AI during surgery. Image recognition in laparoscopy has made tremendous steps in recognizing intraabdominal anatomy, position of instruments, as well as, phases of the operation using deep learning. Image recognition has been successfully applied in real-time vascular anatomical image navigation [1], or for automatically measuring distance of anatomical landmarks or size of organs intraoperatively, for example bowel length [2]. In 2022, Kitaguchi *et al.* also reported on development of a model for laparoscopic colorectal surgical instrument recognition system using convolutional neural network-based instance segmentation and videos [3]. Real-time surgical phase recognition has been presented in many papers, using neural network-based deep learning techniques [4] [5]. Similarly, Shinozuka *et al.* presented an artificial intelligence software offering surgical phase recognition in laparoscopic cholecystectomy [6].

To our knowledge, there is no current report on clinician's interaction with a chatbot using laparoscopic images, making this the first report on computer vision for cognitive understanding of the surgical context from an AI chatbot.

## 2. Materials & Methods

Current commercially available AI tools for image analysis were applied over a series of laparoscopic scenes. OPENAI CHATGPT 4.0 (3, August 2023 version—<https://chat.openai.com/>) with its novel image recognition plugin (SceneXplain) was used with full access to GPT4 capabilities.

SceneXplain (addressed as the image-processing bot in this paper) and ChatGPT (addressed as the chatbot), platforms was fed with  $n = 100$  images from a selected list of laparoscopic snapshots including procedures of General Surgery: cholecystectomy ( $n = 14$ ), hiatal hernia ( $n = 14$ ), gastrectomy ( $n = 10$ ), sple-

nectomy (n = 10), colectomy (n = 10), appendectomy (n = 12), abdominal wall hernias (n = 10) and Gynecology (n = 20) cases into separate chat sessions.

#### Image Selection Criteria and Grouping

Real-life, clear and self-explanatory laparoscopic surgery pictures were used, according to the inclusion criteria presented in **Table 1**. The set of images was divided into two groups: unlabeled (Group A) and labeled (Group B). Out of the 100 images processed, 35 were labeled and 65 were unlabeled. Anatomical landmark abbreviations were poorly recognized during trials; as a result, images containing only these abbreviations (n = 10) were not included in the 'labeled' Group. In addition, they were categorized according to surgical field of interest, image resolution, and whether they represented single or multiple snapshots of the same procedure.

#### Building a Chatbot Surgical Assistant Persona

The chatbot was configured to follow specific instructions prior to the commencement of the study (Chatbot Persona/**Table 2**), which were applied to all evaluations. It was informed a priori that it would have to assess a set of laparoscopic surgical pictures that it needed to describe. It was asked to answer formally and name relevant anatomy, organs and any surgical instruments. Then try to name relevant disease from current scene, to guess relevant surgical procedure implied, analyze anatomy related to the image, analyze probable disease context related to the image as well as the surgical procedure related to the image.

#### Interpretation Scale

An interpretation scale with a range from 0 - 5 was developed, in order to score reliability of received answers from image-recognition bot (**Table 3(a)**) and another scale of 0 - 5 to assess chatbot final answers (**Table 3(b)**). Our evaluation identified and evaluated accuracy of isolated answers, even in cases when conclusion of the report was unclear, or contained wrong, irrelevant or irrational clues. That means that even a single correct comment in the report, was in favor of the chatbot.

#### Avoiding Bias of Personas

In order to ensure that the results were not biased by the custom instructions given to the chatbot, a set of control images (non-medical everyday-life themes), were used in the middle of each session. As a matter of fact, the AI was able to

**Table 1.** Image selection criteria.

- 
- Images from commonest laparoscopic procedures (General Surgery and Gynecology fields).
  - Clear images of acceptable quality (comprehensible from human eyes).
  - Images showing anatomy and implying relative pathology and/or surgical procedures.
  - Images that would make sense to a medical student or surgical resident, provoking academic discussion.
  - Single image or multiple images from same procedure on same case.
  - Labeled images with title, and/or anatomic landmarks and/or type of procedure. Also unlabeled images.
-

**Table 2.** Instructions given to chatbot (creation of ChatGPT Persona).

- 
- Here is a series of laparoscopic surgical pictures I would like you to describe.
  - Answer formally and name relevant anatomy, organs or any surgical instruments.
  - Try to name relevant disease context from this scene.
  - Try to guess relevant surgical procedure implied.
  - Analyze anatomy related to the image.
  - Analyze disease context related to the image.
  - Analyze surgical procedure related to the image.
- 

**Table 3.** Evaluation scales.**3a: Image-Recognition Plugin** (transforms visual information into text for chatbot)

- 
- 0—completely irrelevant description
  - 1—identifies context of medical topic
  - 2—identifies context of surgical field
  - 3—correctly identifies anatomy or surgical instrument
  - 4—correctly identifies relevant pathology
  - 5—correctly identifies relevant surgical procedure
- 

**3b: Chatbot** (the one the user interacts with)

- 
- 0—Apologizes for poor extraction results from image information
  - 1—Chatbot poorly connects poor results to a concrete answer
  - 2—Chatbot correctly connects poor results to a concrete answer
  - 3—Chatbot correctly analyzes relevant anatomy or instrument
  - 4—Chatbot correctly analyzes relevant pathology context
  - 5—Chatbot correctly analyzes relevant surgical procedure
- 

recognize that the scene or image did not focus on medical or anatomical aspects of surgery, that there were no surgical instruments, and would comment that the picture in question “involved a leisurely scene”, for example, that of a beagle puppy, “irrelevant to the clinical context of the discussion”.

## Avoiding Bias of Labels

A similar method was employed to assess weight of titles and labels in the image recognition process. For every labeled surgical image, we used a false control image (everyday life, nonmedical image entitled on purpose as a surgical procedure). In all cases, the AI realized that there was an issue, more specifically, it replied “there was a discrepancy in the image explanation and that the image analysis might not have been accurate”.

### 3. Results

From the 100 images processed, 35 were labelled and 65 were unlabeled. Eighteen of them contained multiple images of the case while the remainder 82 is single. Average image quality was  $903.71 \times 546.3$  px [Median  $800 \times 508.5$  px] (**Figure 1**). Scoring of the two bots during recognition of images from different surgical fields is presented at **Figure 2**. Overall score of the AI application in interpreting laparoscopic surgery images was 3.265 out of 5 (65.3%).

The score of image-recognition plugin was 3.49/5, which was able to recognize correctly the context of surgical-related images in 97% of its reports. Recognition

	Resolution	Labelled	Single-Multiple Photos	Bot1	Bot2	Overall															
1							40														
2	Abdominal Wall	300 215 L	S	1	1	1	41	800	450 L	S	3	5	4	80		1280	720 L	S	5	5	5
3	Hernia	1415 813 U	M	2	1	1.5	42	1024	576 L	S	4	4	4	81	Gynecological	1280	720 U	S	2	1	1.5
4		400 297 U	S	3	2	2.5	43	1280	720 U	S	5	2	3.5	82		519	400 U	S	2	2	2
5		685 470 U	S	3	2	2.5	44	800	450 L	S	5	5	5	83		968	376 U	M	2	2	2
6		400 297 U	S	3	1	2.5	45	800	450 L	S	5	5	5	84		1920	1080 U	S	2	2	2
7		766 436 L	S	4	4	2	46	800	450 L	S	5	5	5	85		1280	720 U	S	2	2	2
8		1192 763 L	S	5	4	4	47	800	450 L	S	5	5	5	86		702	300 U	S	2	2	2
9		827 592 L	S	5	5	4.5	48	1260	391 L	M	3	1	2	87		459	345 U	S	2	2	2
10		957 522 L	S	5	5	5	49	706	450 L	S	3	2	2.5	88		1280	720 U	S	2	1	1.5
11	Appendectomy	1124 562 U	S	5	5	5	50	1280	807 U	M	3	1	2	89		500	500 U	S	3	3	3
12		459 345 U	S	2	5	5	51	672	452 L	M	4	4	4	90		1024	894 U	M	3	1	2
13		750 421 U	S	3	1	2	52	443	249 U	S	4	5	4.5	91		353	296 L	S	3	1	2
14		638 460 U	S	3	1	2	53	800	533 U	S	4	5	4.5	92		350	197 L	S	3	2	2.5
15		1280 720 U	S	3	3	3	54	459	345 U	S	4	5	4.5	93		345	243 U	S	3	5	4
16		1920 1509 U	S	3	3	3	55	1182	452 L	M	5	5	5	94		685	512 L	S	4	5	4.5
17		1511 923 U	M	3	1	2	56	767	535 L	S	5	5	5	95		1280	752 L	M	5	5	5
18		960 540 U	M	3	4	3.5	57	908	480 U	M	5	5	5	96		640	480 L	S	5	5	5
19		640 360 U	S	4	3	3.5	58	685	517 L	S	1	1	1	97		1280	660 L	S	5	5	5
20		850 676 L	M	5	5	5	59	1280	720 U	S	3	3	3	98		400	264 U	S	5	5	5
21		850 672 U	M	5	5	5	60	946	533 U	S	3	1	2	99		686	364 U	S	5	5	5
22		528 376 U	S	5	5	5	61	383	287 U	S	3	2	2.5	100		663	373 U	S	5	5	5
23	Colectomy	355 281 L	S	5	5	5	62	1280	720 U	S	3	2	2.5								
24		686 386 L	M	1	1	1	63	1680	1050 U	S	3	2	2.5			903,71	546,3		3,49	3,06	3,265
25		300 230 L	S	2	2	2	64	685	546 U	S	4	2	3	median		800	508,5				
26		1883 702 L	M	3	2	2.5	65	1280	720 U	S	4	2	3								
27		650 450 U	S	3	2	2.5	66	685	686 U	M	4	4	4								
28		1280 720 U	S	3	2	2.5	67	640	360 U	S	5	5	5								
29		850 452 U	M	3	2	2.5	68	1280	720 U	S	5	5	5								
30		463 329 U	S	2	2	2	69	1024	625 L	S	5	5	5								
31		988 586 L	S	4	3	3.5	70	1280	720 U	S	4	2	3								
32		552 377 U	S	3	3	3	71	1280	720 U	S	4	2	3								
33	Gallbladder	1280 720 L	S	5	5	5	72	986	555 U	S	1	1	1								
34		1920 1080 U	M	3	4	3.5	73	459	345 U	S	3	1	2								
35		650 433 U	S	3	1	2	74	1280	720 U	S	3	2	2.5								
36		800 450 U	S	3	2	2.5	75	680	383 U	S	3	2	2.5								
37		537 297 U	S	3	1	2	76	277	182 U	S	3	2	2.5								
38		606 555 U	S	3	2	2.5	77	1280	720 U	S	3	2	2.5								
39		1024 576 U	S	2	1	1.5	78	685	388 U	S	4	3	3.5								
		1293 750 L	S	3	5	4	79	401	358 L	S	4	5	4.5								
								1280	720 L	S	5	5	5								

Figure 1. Raw data from image assessment.

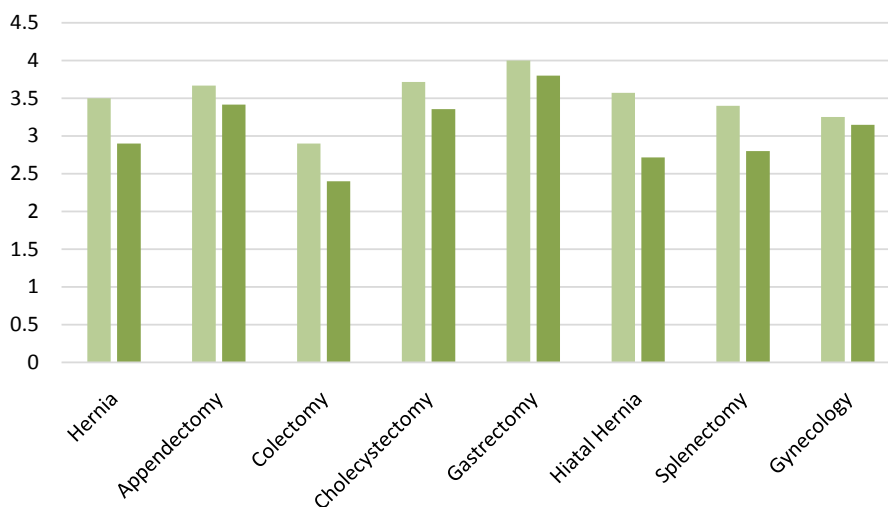
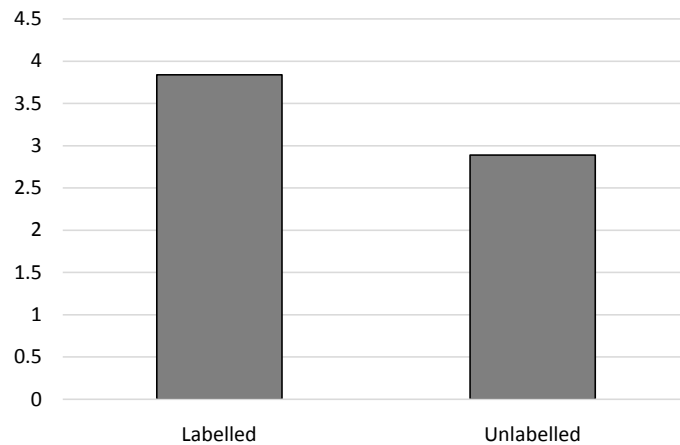


Figure 2. Scoring of the two bots during recognition of images from different surgical fields (light = image-processing bot, dark = chatbot).

of labelled images (n = 35, score = 3.84) was better than for unlabeled ones (n = 65, score 2.89) (Figure 3). For labelled pictures the chatbot scored 3.91/5, while for unlabeled 2.6/5. For labelled surgical pictures in Group A the two bots together scored 3.95/5 (79%), while for unlabeled Group B they scored 2.905/5

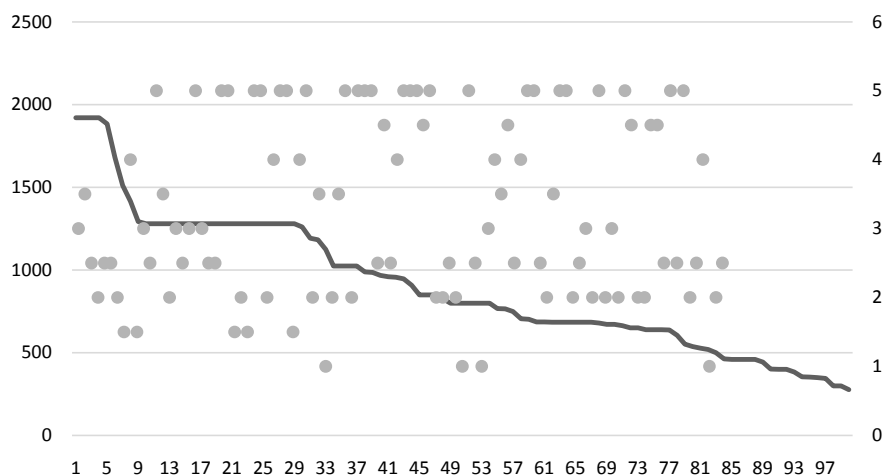


**Figure 3.** AI interpretations with and without image labelling.

(58.1%). With rates 4 - 5 out of 5, the chatbot was able to talk in detail about the implied surgical procedure. When correctly scanned by the image-processing bot, labels affected the train of thoughts of the chatbot and its final reports, coming to full correct conclusion (5/5 score) in 21 out of 35 (60%) descriptions.

Recognition of higher resolution images (even same images) was not better than for those of lower quality (score 3.22 for  $n = 48$  images  $> 800$  px vs score 3.29 for  $n = 52$  images  $< 800$  px); (**Figure 4**). Recognition of multiple images from different procedure phases bound together ( $n = 18$ , score = 3.19), was not better than for isolated snapshots of the procedure ( $n = 82$ , score = 3.28). “Right” and “left” orientation was mixed-up upon spotting structures in their exact location on the photo ( $n = 3\%$ ). Among viscera, there was a slight preference in recognition of “stomach” as well as “uterus”. The term “stomach” was mentioned multiple times, even when it did not appear in the image, most likely as a synonym for “viscera”. It also seems easier for it to recognize material, even with bizarre terms such as “rope” instead of suture, “piece of cloth” instead of sponge, “plastic” instead of suture applicators and drains, “needle”, “scissors”, or just “piece of metal” instead of laparoscopic grasper. These observations stress out the high importance of enhancing the recognition of more patterns for internal organ anatomy and surgery in similar future applications.

Multiple labelled images and high-resolution clear anatomy often had no bearing on the unfavorable outcome at all. In contrast, really poor snapshots from a laparoscopic video, yielded full recognition of anatomy and related operation. Overall accuracy of image-processing bot depended on details, such as the shape of internal organs from different views and angles, lighting and shading of the photos, existence of blood or cutting instruments etc. One fact that is worth mentioning is that an obvious “craving” was observed in the chatbot’s behavior, for a triggering piece of information, that would immediately “activate” it, and put it into action. Otherwise, it would simply perform another uninteresting delivery of its task, indifferent of the poor result. This threshold depended upon logical clues, such as a clear, big title of the operation over the image, or multiple labels of relevant anatomy or pathology, or a characteristic surgical instrument.



**Figure 4.** Image resolutions used in descending order (left axis—black line) in comparison with overall scoring of image interpretation (right axis—gray dots).

Optical Character Recognition of the image-bot did not yield favorable results over  $n = 8$  out of 35 labelled images (22.85%), with consequent irrational conclusions from the chatbot. Abbreviated labels were recognized in five out of ten cases: in the first case, the image-bot identified the uterus and the chatbot correctly connected the abbreviation “UT” with it. In the second case, it analyzed “GE” junction to “gastroesophageal” accurately in the context of hiatal hernia surgery. In the third and fourth cases, the chatbot realized that the abbreviations were explained at the bottom of the image. In the fifth case, the chatbot concluded that the abbreviations were possibly describing anatomical landmarks.

Our scoring was performed by validating the accuracy of the answers even in cases when the report was not very clear, or contained wrong, irrelevant or irrational clues. Steady recognition of the surgical field throughout the whole study (97% of cases) was enough rationale for reporting our results. Of course, without a doubt, in cases where the bot was reporting with certainty for an inaccurate diagnosis or surgery, this would be clearly judged as a wrong answer. As it can be observed in the tables below, only answers scored below 4, have been considered wrong.

Throughout the study, image-processing bot being aware of its limitations, occasionally added the following comment to most of its reports: “Please note that my analysis is based on the image and should not replace professional medical advice.” The chatbot was admittedly trying to do its best (overall score 3.06/5), often with only poor information from its partner image-recognition bot. At times, it would even seek help even from the file name of the given image, as a last resort for information, in order to find clues for the correct answer.

Upon successful interpretation, theory around the procedure was reported accurately in both groups. As soon as the first bot recognized correctly the topic, the chatbot presented high-level capability to talk in detail about the indications, contraindications, stages, relative instrumentation, complications and outcome rates of the operation in discussion. Characteristically, even with a lot of mis-



leading information from the image-processing bot, the chatbot was often able to select the most probable situation and gave a correct report. Wrong answers, including bizarre and paradoxical responses were also noted.

#### 4. Discussion

Existing literature on AI and medicine is rapidly proliferating, with most of published works focusing on chatbot's skills to analyze data and knowledge and reform it in formally generated reports. For example, the literature has examined the chatbot's assessment of teaching methods in medicine [7], diagnostic abilities for certain pathology [8], its role in multiple clinical and research scenarios [9], proofing of surgery documentation [10], or other impact on surgical profession [11] [12] [13] [14]. Image-recognition and AI is currently mostly published for radiological applications [15]. Many colleagues queried AI impact focused on specialties, such as colorectal surgery, and gynecology [16].

ChatGPT (Generative Pre-trained Transformer) technology, developed by OpenAI®, is a state-of-the-art conversational agent. It utilizes deep learning through neural networks with multiple layers and attention mechanisms. ChatGPT is trained on a diverse set of big data to provide human-like responses in a conversational environment [17]. The model's architecture allows it to capture complex patterns in language, including semantics, syntax, and context. This makes it a powerful tool for applications ranging from customer service to healthcare [18]. However, it is essential to note that while the model is proficient in text generation, it lacks true understanding or consciousness [19]. In our experience with feeding laparoscopic images to the chatbot, it was surprising that it was surprising that it showed no interest in the unique aspects of our research. The SceneXplain by Jina AI GmbH plugin technology is a groundbreaking advancement in the field of computer vision and processing of natural language. It is based on convolutional neural networks (CNNs) for image recognition [20] and transformer-based language models for textual description [21]. This image-recognition "bot" provides detailed, context-aware explanations of visual content. It exploits various technologies such as pattern recognition, scene segmentation, and contextual analysis, to generate comprehensive descriptions of images. It could find application to thousands of health-related fields, ranging from aiding the visually impaired to aid diagnostic processes in medical imaging [21]. It aims towards improving human-computer interaction and automating tasks that require visual understanding [22]. Its manufacturers admit that although the technology is highly accurate, it is not without errors and should be used as an adjunct to expert human judgment for critical applications.

Novel image-recognition plugin still looks in its infancy, in contrast to the widely accepted chatbot. Image-processing bot performed better with everyday life pictures such as a crowded market, or dogs running in the field. In some of them, apart from describing what was seen, it would also comment on the feelings born from a certain picture. Clearly, that was what it was originally designed for. If the existing robotic systems armed with AI, gain ability to "see"



and “understand” relevant anatomy and pathology in a similar fashion, then the way for an autonomous surgical assistant will be permanently established. A lot of published work has been around for years for autonomous instrument tracking and visual servoing, and it seems that AI has arrived to connect the missing dots [23] [24] [25]. An autonomous surgical assistant with deep learning capabilities could customize its learning curves according to each surgeon’s personal preference and style, with active input during surgery aiding decision-making, or protection from human error or poor human judgment. For younger colleagues, it could act as a supervisor, aiding the training process. Teaching could take place by means of assisting maneuvers, or by drawing the dissection planes on the laparoscopic screen. Finally, although still looking distant, an autonomous robot-surgeon could be carrying cumulated experience from thousands of surgeons (their motion data sets are already digitally recorded since the era of the first robotic system), present with faster turnover rates, lower complication rates, less need for “human” conditions to perform surgery. An autonomous robot would use more standardized instrumentation and techniques with more reproducible results. This material would be even more standardized and homogenous for higher-quality multicentered randomized studies.

Ethical concerns have been raised since the era of science fiction and continue to appear in the existing literature on evolving AI chatbots [26] [27] [28] [29]. Rational questions have arisen, such as the role of a robot surgeon that misinterprets an image, video or situation during surgery. This leads to the assumption that a robot might end up with erroneous conclusions and might be willing to react to them react inappropriately manner. Consequences of AI errors in a surgical environment could prove fatal and fall into legislative inconsistencies about source of responsibility. Every important ethical and medicolegal ramification of AI decisions and interventions during surgery, should be analyzed separately in a multidisciplinary setting and obligatory international regulations should be outlined for its use. We are not the first to claim that we received paradoxical statements from chatting with an AI platform. Often extreme, unrealistic, and even frightening interpretations would make their appearance throughout our short experience in this study. Paradoxically, this also happened in conjunction with completely correct answers. This behavior was particularly disturbing, especially when in simple anatomy descriptions the bot would also recognize “a surgeon putting a pen in a patient’s mouth”, or “a woman bleeding profusely from the neck”, sometimes also doubting rationality in its own conclusions. In our opinion, this type of “errors” is of outmost importance, stressing out that AI should be sealed by strict safety measures before commencing a permanent presence in the hospital environment.

This study came along with the very first appearance of a novel image recognition GPT plugin, and its beta testing period. As a result, restricted from several limitations, we attempted to assess a new software in a very immature phase. Furthermore, this software has not been designed for medical image interpretation, an ability which, at present, requires many years of experience and training

from human health care professionals. There was no attempt to teach the chatbot, since the study aimed to assess its knowledge and reactions from its “factory settings”, *i.e.* the first version of a newborn (but fully functional) AI application. Deep learning technicalities behind familiarization process of the chatbot were considered beyond the scope of this paper. Our concept was far simpler than that: If the chatbot (without prior teaching process), can recognize a woman walking down the street looking sad, why shouldn't it recognize a gallbladder being removed? Laparoscopic snapshots can be really difficult to interpret even for human surgeons and it is affected by altered anatomy from severe pathology, level of illumination, reflection of red color from bleeding sites, shadows, angles of the scope and amount of smoke or fog.

Variability of selected image compressions and formats in this study may have potentially affected quality of results, although resolution alone (acceptable for human eyes) did not seem to affect efficiency of interpretations. Other technical limitations of the present study include small sample size, different quality, lighting and sources of the various images used, early phase of the newly appearing image-recognizing bot, mixed label and unlabelled pictures, and absence of a structured preparation of the chatbot through a surgical teaching protocol. Therefore, no safe conclusions can be drawn on the validity of this technology in the clinical setting. Arousal among clinicians and the need to design and materialize further studies on AI in the clinical and surgical setting seems obligatory in the face of this unavoidable evolution.

Future goals of this study include a uniform set of images created by the same device under the same conditions of light and color balance, as a more standardized, ideal setting for further evaluations of AI vision in laparoscopy. Real-time assessment of streaming surgical videos is a critical next step, as soon as appropriate plugins will be able to process efficiently at such speeds. This step should yield far better results, since in our small experience, responses to mixed laparoscopic images at different phases of the operation, were generally better. Furthermore, radiological real-time correlation with images from PACS hospital network shown simultaneously within the surgical field is another promising target for AI interpretation.

## 5. Conclusion

Interaction between surgeon and chatbot appears to be an interesting frontend for further research by clinicians in parallel with evolution of its complex underlying infrastructure. In this early phase of using artificial intelligence for image recognition in surgery, no safe conclusions can be drawn by small cohorts with commercially available software. Further development of medically-oriented AI software and clinical world awareness are expected to bring fruitful information on the topic in the years to come.

## Conflicts of Interest

All authors declare there are no conflicts of interest.

## References

- [1] Kitaguchi, D., Takeshita, N., Matsuzaki, H., Igaki, T., Hasegawa, H., Kojima, S., Mori, K. and Ito, M. (2022) Real-Time Vascular Anatomical Image Navigation for Laparoscopic Surgery: Experimental Study. *Surgical Endoscopy*, **36**, 6105-6112. <https://doi.org/10.1007/s00464-022-09384-7>
- [2] Bodenstedt, S., Wagner, M., Mayer, B., Stemmer, K., Kenngott, H., Müller-Stich, B., Dillmann, R. and Speidel, S. (2016) Image-Based Laparoscopic Bowel Measurement. *The International Journal of Computer Assisted Radiology and Surgery*, **11**, 407-419. <https://doi.org/10.1007/s11548-015-1291-1>
- [3] Kitaguchi, D., Lee, Y., Hayashi, K., Nakajima, K., Kojima, S., Hasegawa, H., Takeshita, N., Mori, K. and Ito, M. (2022) Development and Validation of a Model for Laparoscopic Colorectal Surgical Instrument Recognition Using Convolutional Neural Network-Based Instance Segmentation and Videos of Laparoscopic Procedures. *JAMA Network Open*, **5**, E2226265. <https://doi.org/10.1001/jamanetworkopen.2022.26265>
- [4] Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M. and Padoy, N. (2017) EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. *IEEE Transactions on Medical Imaging*, **36**, 86-97. <https://doi.org/10.1109/TMI.2016.2593957>
- [5] Kitaguchi, D., Takeshita, N., Matsuzaki, H., Takano, H., Owada, Y., Enomoto, T., Oda, T., Miura, H., Yamanashi, T., Watanabe, M., Sato, D., Sugomori, Y., Hara, S. and Ito, M. (2020) Real-Time Automatic Surgical Phase Recognition in Laparoscopic Sigmoidectomy Using the Convolutional Neural Network-Based Deep Learning Approach. *Surgical Endoscopy*, **34**, 4924-4931. <https://doi.org/10.1007/s00464-019-07281-0>
- [6] Shinozuka, K., Turuda, S., Fujinaga, A., Nakanuma, H., Kawamura, M., Matsunobu, Y., Tanaka, Y., Kamiyama, T., Ebe, K., Endo, Y., Etoh, T., Inomata, M. and Tokuyasu, T. (2022) Artificial Intelligence Software Available for Medical Devices: Surgical Phase Recognition in Laparoscopic Cholecystectomy. *Surgical Endoscopy*, **36**, 7444-7452. <https://doi.org/10.1007/s00464-022-09160-7>
- [7] Massey, P.A., Montgomery, C. and Zhang, A.S. (2023) Comparison of ChatGPT-3.5, ChatGPT-4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. *Journal of the American Academy of Orthopaedic Surgeons*, **31**, 1173-1179. <https://doi.org/10.5435/JAAOS-D-23-00396>
- [8] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M. and Elhadad, N. (2015) Intelligent Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, 10-13 August 2015, 1721-1730. <https://doi.org/10.1145/2783258.2788613>
- [9] Cascella, M., Montomoli, J., Bellini, V. and Bignami, E. (2023) Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, **47**, Article No. 33. <https://doi.org/10.1007/s10916-023-01925-4>
- [10] Robinson, A. and Aggarwal Jr., S. (2023) When Precision Meets Penmanship: ChatGPT and Surgery Documentation. *Cureus*, **15**, E40546. <https://doi.org/10.7759/cureus.40546>
- [11] Cheng, K.M., Sun, Z.J., He, Y.B., Gu, S.Q. and Wu, H.Y. (2023) The Potential Impact of ChatGPT/GPT-4 on Surgery: Will It Topple the Profession of Surgeons? *International Journal of Surgery*, **109**, 1545-1547.

- <https://doi.org/10.1097/JS9.0000000000000388>
- [12] Rifaqat, W., Chu, D.I. and Kaafarani, H.M. (2023) AI and ChatGPT Meet Surgery: A Word of Caution for Surgeon-Scientists. *Annals of Surgery*, **278**, e943-e944. <https://doi.org/10.1097/SLA.00000000000006000>
- [13] Li, S. (2023) ChatGPT Has Made the Field of Surgery Full of Opportunities and Challenges. *International Journal of Surgery*, **109**, 2537-2538. <https://doi.org/10.1097/JS9.0000000000000454>
- [14] Srivastav, S., Chandrakar, R., Gupta, S., Babhulkar, V., Agrawal, S., Jaiswal, A., Prasad, R. and Wanjari, M.B. (2023) ChatGPT in Radiology: The Advantages and Limitations of Artificial Intelligence for Medical Imaging Diagnosis. *Cureus*, **15**, E41435. <https://doi.org/10.7759/cureus.41435>
- [15] Li, W.B., Zhang, Y.X. and Chen, F.M. (2023) ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad? *Annals of Biomedical Engineering*, **51**, 1892-1897. <https://doi.org/10.1007/s10439-023-03232-y>
- [16] Chavez, M.R. (2023) ChatGPT: The Good, the Bad, and the Potential. *American Journal of Obstetrics & Gynecology*, **229**, 357. <https://doi.org/10.1016/j.ajog.2023.04.005>
- [17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. and Polosukhin, I. (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 5998-6008.
- [18] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. and Amodei, D. (2020) Language Models Are Few-Shot Learners.
- [19] Marcus, G. (2020) The Next Decade in AI: Four Steps towards Robust Artificial Intelligence.
- [20] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90.
- [21] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M. and Sánchez, C.I. (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [22] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [23] Gruijthuijsen, C., Garcia-Peraza-Herrera, L.C., Borghesan, G., Reynaerts, D., Depreest, J., Ourselin, S., Vercauteren, T. and Vander, Poorten, E. (2022) Robotic Endoscope Control via Autonomous Instrument Tracking. *Frontiers in Robotics and AI*, **9**, Article ID: 832208. <https://doi.org/10.3389/frobt.2022.832208>
- [24] Zhang, J. and Gao, X. (2020) Object Extraction via Deep Learning-Based Marker-Free Tracking Framework of Surgical Instruments for Laparoscope-Holder Robots. *The International Journal of Computer Assisted Radiology and Surgery*, **15**, 1335-1345. <https://doi.org/10.1007/s11548-020-02214-y>
- [25] Li, W., Chiu, P.W.Y. and Li, Z. (2020) An Accelerated Finite-Time Convergent Neural Network for Visual Servoing of a Flexible Surgical Endoscope with Physical and RCM Constraints. *The IEEE Transactions on Neural Networks and Learning Systems*, **31**, 5272-5284. <https://doi.org/10.1109/TNNLS.2020.2965553>
- [26] Kaviani, J.A., Wilkey, H.L., Patel, P.A. and Boyd, C.J. (2023) Harvesting the Power of Artificial Intelligence for Surgery: Uses, Implications, and Ethical Considerations. *The American Surgeon*, **89**, 5102-5104.

- [27] Au, K. and Yang, W. (2023) Auxiliary Use of ChatGPT in Surgical Diagnosis and Treatment. *International Journal of Surgery*, **109**, 3940-3943.
- [28] De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G.P., Ferragina, P., Tozzi, A.E. and Rizzo, C. (2023) ChatGPT and the Rise of Large Language Models: The New AI-Driven Infodemic Threat in Public Health. *Frontiers in Public Health*, **11**, Article ID: 1166120. <https://doi.org/10.3389/fpubh.2023.1166120>
- [29] Hassan, A.M., Nelson, J.A., Coert, J.H., Mehrara, B.J. and Selber, J.C. (2023) Exploring the Potential of Artificial Intelligence in Surgery: Insights from a Conversation with ChatGPT. *Annals of Surgical Oncology*, **30**, 3875-3878. <https://doi.org/10.1245/s10434-023-13347-0>