**Journal of Neonatal Surgery**
***ISSN(Online): 2226-0439***
***Vol. 14, Issue 5 (2025)***
*https://www.jneonatalsurg.com*

OPEN ACCESS

# Instruction Tuned Large Language Models for Assisting Brain Surgery Through Procedural Alignment and Decision Support Using PPO Reinforcement Learning

**G Ramana Murthy[1], Dr. A. Jansi Rani[*2], Lavanya S[3], Dr. P Ravi Kumar[4], Dr. K. Karunambiga[5]**

[1]Professor, Department of ECE, Alliance University, Bangalore.

Email ID: ramana.murthy@alliance.edu.in

[*2]Assistant Professor and Head, Department of Information Technology, Nirmala College for Women , Red Fields, Coimbatore.

Email ID: Jansiramalraj@gmail.com

[3]Assistant professor, Department of Computer science and engineering, Sona College of Technology (An Autonomous Institution), Junction Main Road, Salem - 636 005, Tamilnadu, India.

Email ID: lavansadeesh4@gmail.com

[4]Assistant Professor (Sl.G), Department of EEE, KPR Institute of Engineering and Technology, Coimbatore

Email ID: ravikumar.p@kpriet.ac.in

[5]Professor, Department of CSE, Karpagam Institute of Technology, Coimbatore 641105

Email ID: karunambiga.cse.cit@gmail.com

***Corresponding Author:**

Assistant Professor and Head, Department of Information Technology, Nirmala College for Women , Red Fields, Coimbatore.

Email ID: Jansiramalraj@gmail.com

Cite this paper as: G Ramana Murthy, Dr. A. Jansi Rani, Lavanya S, Dr. P Ravi Kumar, Dr. K. Karunambiga, (2025) Instruction Tuned Large Language Models for Assisting Brain Surgery Through Procedural Alignment and Decision Support Using PPO Reinforcement Learning. *Journal of Neonatal Surgery*, 14 (5), 166-175.

## ABSTRACT

The integration of large language models (LLMs) into clinical decision-making remains a critical challenge, especially in high-risk domains such as neurosurgery. This study presents a novel framework that leverages instruction-tuned LLMs optimized using Proximal Policy Optimization (PPO) reinforcement learning to assist brain surgery through procedural alignment and decision support. We begin by fine-tuning a transformer-based LLM on domain-specific surgical protocols and neurosurgical dialogue datasets using supervised instruction tuning. To further enhance procedural adherence and mitigate hallucinations, we introduce a reward model guided by expert-annotated signals such as factual accuracy, stepwise protocol fidelity, and relevance to surgical context. PPO is employed to iteratively refine the model's responses through a feedback loop, optimizing both language coherence and domain-specific reliability. Experimental evaluations on simulated neurosurgical benchmarks demonstrate that our model outperforms both instruction-tuned and PPO-only baselines in terms of procedural accuracy and decision support relevance. The results indicate that reinforcement learning with human feedback, when tailored to surgical requirements, significantly improves trustworthiness and alignment in LLM outputs. This research contributes a critical step toward the deployment of explainable, reliable AI assistants for neurosurgical procedures.

***Keywords:*** *Instruction tuning, Large language models, Brain surgery, Proximal Policy Optimization, Reinforcement learning, Procedural alignment, Decision support, Hallucination mitigation, RLHF, Medical AI*

## 1. INTRODUCTION

Brain surgery is among the most complex and high-risk procedures in modern medicine, demanding precision, rapid decision-making, and seamless coordination between surgical team members. Errors or delays in judgment can lead to irreversible consequences, including loss of function or life. In this context, artificial intelligence (AI) has shown promise in enhancing surgical planning, image interpretation, and intraoperative guidance. However, most AI models remain limited in

their understanding of procedural flow, contextual reasoning, and dynamic adaptation to intraoperative scenarios (Topol, 2019). The lack of real-time, trustworthy, and context-aware AI assistants remains a significant barrier to the deployment of AI in neurosurgical decision-making.

Recent advancements in large language models (LLMs) have demonstrated remarkable capabilities in understanding, generating, and reasoning over complex text inputs across domains. Instruction tuning, which refines LLM behavior using task-specific prompts and desired outputs, has significantly improved model alignment with human expectations (Ouyang et al., 2022). Yet, instruction-tuned models alone may fall short in adapting to dynamic clinical environments like neurosurgery, where responses must be updated continuously based on feedback. To address this, reinforcement learning—particularly Proximal Policy Optimization (PPO)—offers a powerful approach to refine model behavior through a feedback loop, optimizing not just accuracy but alignment with expert judgment (Schulman et al., 2017).

Despite these innovations, few models are tailored for surgical contexts. Generic instruction-tuned models lack domain-specific training, and existing reinforcement learning frameworks are rarely tested in real-time surgical decision support tasks. This gap is particularly pronounced in brain surgery, where procedural alignment—i.e., the model's ability to follow and assist with the structured sequence of surgical steps—is crucial for clinical reliability and trust (Lundstrom et al., 2022).

The objective of this research is to develop an instruction-tuned LLM fine-tuned with PPO to support neurosurgeons in procedural guidance and decision-making. By integrating domain-specific surgical protocols, expert-validated prompts, and real-time reinforcement feedback, our model aims to reduce hallucinations, improve procedural fidelity, and offer contextual decision support. This approach marks a step forward in creating intelligent, aligned, and safe AI assistants capable of operating in high-stakes medical environments.

## 2. RELATED WORK

Instruction tuning has become central to aligning large language models (LLMs) with specific tasks, enhancing controllability and user intent alignment. InstructGPT, for example, applied supervised fine-tuning followed by Proximal Policy Optimization (PPO)-based reinforcement learning from human feedback (RLHF), significantly improving output helpfulness and safety across general tasks (Ouyang et al., 2022). Likewise, FLAN-T5 expanded instruction tuning to a broad range of task types but lacks reinforcement learning mechanisms and domain specificity (Chung et al., 2022). As illustrated in Table 1, both models exhibit low procedural alignment and limited integration in clinical or high-risk domains such as neurosurgery.

PPO is one of the most widely adopted reinforcement learning algorithms due to its robustness and stability in policy updates. Its use in healthcare has included treatment planning, personalized medicine, and diagnostic path optimization (Schulman et al., 2017; Yu et al., 2021). While PPO enables effective learning from complex reward structures, its application to language-based real-time decision-making remains nascent. As shown in Table 1, healthcare PPO implementations often demonstrate moderate medical integration but are rarely extended to surgical contexts requiring procedural alignment.

LLMs deployed in clinical settings face the persistent challenge of hallucination—i.e., confidently stating incorrect or unverified information. Med-PaLM was among the first efforts to align LLMs with medical expert responses and safety requirements using curated datasets (Singhal et al., 2023). While the model improves factuality and trustworthiness in question answering, it remains static and lacks procedural reasoning, as indicated by its moderate procedural alignment in Table 1. Additionally, it does not employ reinforcement learning methods like PPO to iteratively improve reliability.

Neurosurgical decision support systems have traditionally used static rule-based engines, image-guided navigation, and database retrieval systems. While useful in structured scenarios, they fall short in complex or evolving surgical workflows where flexible, context-aware reasoning is essential. As shown in Table 1, these systems score high in domain integration and procedural alignment but are limited by their lack of LLM integration and adaptability, presenting a compelling opportunity for hybrid solutions that incorporate instruction tuning and PPO-driven learning.

| Approach / Model | Domain Focus | Reinforcement Learning | Medical Integration | Procedural Alignment | Limitations |
|---|---|---|---|---|---|
| InstructGPT | General-purpose tasks | Yes (PPO-based) | Limited | Low | Not specialized for healthcare; prone to hallucinations |
| FLAN-T5 | Instruction-following across domains | No | Minimal | Low | Poor factual consistency in high-stakes domains |

| | | | | | |
|---|---|---|---|---|---|
| PPO in Healthcare | Reinforcement learning in treatment policies | Yes (PPO or DDPG) | Moderate | Moderate | Requires precise reward modeling; limited real-time support |
| Med-PaLM | Medical Q&A and reasoning | No | Strong | Moderate | Static Q&A; lacks intraoperative adaptability |
| Surgical Decision Support Systems | Task-specific neurosurgical tools | Rarely used | High | High | Limited scalability and lack of LLM integration |

**Table 1. Related Work Comparison**

This comparative analysis (Table 1) reinforces the novelty of our work, which combines instruction tuning with PPO-based reinforcement learning to deliver a real-time, adaptable, and clinically reliable LLM for brain surgery support.

## 3. METHODOLOGY

This section describes the end-to-end methodology for building a Proximal Policy Optimization (PPO) enhanced instruction-tuned large language model (LLM) to assist in brain surgery through procedural alignment and decision support.

### 3.1 Data Preparation

The training corpus includes three data streams:

1. Surgical Protocols and Operative Notes: These are structured documents that outline step-by-step brain surgery procedures.
2. Validated Q&A Datasets: Derived from medical examinations (e.g., MedQA, MedMCQA) and augmented with neurosurgical literature.
3. Expert Annotation: A panel of neurosurgeons annotates data for procedural correctness, hallucination risks, and response quality.

Annotations are used to provide supervision signals and later guide reward modeling.

### 3.2 Base Model Selection and Instruction Tuning

We begin with a transformer-based model such as LLaMA or GPT-NeoX due to their scalability and open-access architectures. The instruction tuning stage uses domain-specific prompts derived from surgical scenarios, medical questions, and intraoperative decision points.

Instruction Prompt Example
*Prompt:* "You are assisting in a craniotomy. What are the next three procedural steps after dural incision?"
*Expected Output:* "(1) Retract the dura mater, (2) Identify the cortical surface, (3) Plan entry based on preoperative imaging."

Tuning is performed using supervised learning to minimize cross-entropy between model outputs and expert references.

### 3.3 PPO Reinforcement Learning for Procedural Alignment

Following instruction tuning, the model undergoes reinforcement learning to optimize for trustworthiness and procedural adherence.

A reward model is designed using:

- Expert Feedback: Binary or scalar scores based on procedural correctness.
- Factual Consistency: Penalization of hallucinations or medically incorrect outputs.
- Relevance Scoring: Matching output steps to correct surgical sequences.

The PPO algorithm iteratively updates a policy $\pi\theta$ using a clipped objective to ensure stable optimization.

| |
|---|
| Initialize policy $\pi\theta$ and value network $V\phi$ |
| for each training iteration do: |

| |
|---|
| Collect batch of prompts and generate outputs using πθ |
| Compute reward r using expert feedback and procedural metrics |
| Estimate advantage A = r - Vϕ(s) |
| Update policy πθ using clipped PPO objective: |
| L_CLIP = min(πθ/πθ_old * A, clip(πθ/πθ_old, 1-ε, 1+ε) * A) |
| Update value network Vϕ to minimize (Vϕ(s) - r)^2 |
| Update reward model periodically with new human-labeled examples |
| end for |

**Table 2. Pseudocode of the Proposed Approach**

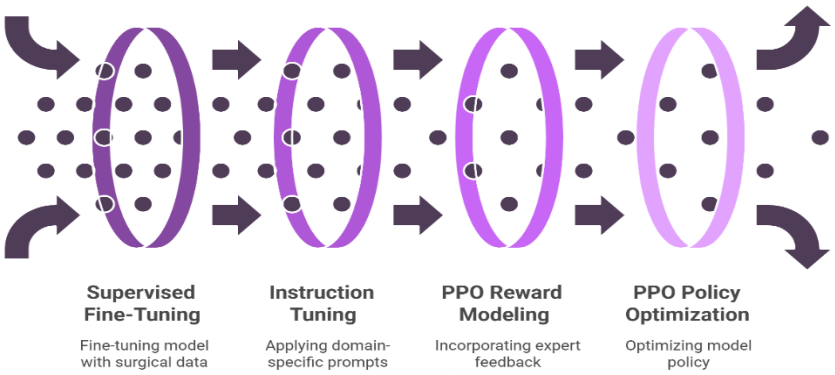


**Figure 1. Refining Surgical Data for AI Models**

The training pipeline illustrated in the figure represents a structured framework for developing a domain-specialized large language model (LLM) designed to assist in brain surgery through procedural alignment and decision support. The process begins with the collection of curated surgical data, including operative notes, clinical protocols, and medically validated Q&A datasets. This data is annotated with expert input to capture procedural nuances and eliminate ambiguous responses. The base model—either LLaMA or GPT-NeoX—is first fine-tuned through supervised learning to internalize fundamental neurosurgical terminology and procedural reasoning. Building on this, the model undergoes instruction tuning, where it learns to respond accurately to domain-specific prompts that reflect intraoperative decision points and stepwise surgical guidance. To further enhance its procedural precision and trustworthiness, a reward model is constructed using expert evaluations, focusing on factors such as factual accuracy, procedural fidelity, and safety. This reward signal feeds into a Proximal Policy Optimization (PPO) loop that iteratively refines the model's policy through a stable reinforcement learning framework. PPO incorporates the policy, value function, and advantage estimator to adjust the model's responses toward more aligned and clinically appropriate outputs. The final phase of the pipeline involves human-in-the-loop evaluation, where neurosurgeons assess the model's behavior across various held-out tasks, ensuring quality control and providing additional feedback for fine-grained tuning. This closed-loop architecture ensures the LLM evolves into a procedurally reliable and context-aware assistant capable of supporting neurosurgical workflows in real-time.

The evaluation of the proposed instruction-tuned large language model (LLM), optimized using Proximal Policy Optimization (PPO), is conducted across three targeted tasks designed to measure its efficacy in assisting brain surgery. The first task assesses procedural alignment accuracy, where the model's generated surgical steps are compared against expert-annotated reference sequences to determine how well it adheres to clinically validated protocols. This metric is critical for ensuring that the model supports surgeons with logically consistent and contextually accurate guidance during operative procedures. The second task evaluates decision support reliability in simulated neurosurgical scenarios. These scenarios are curated to reflect realistic intraoperative situations, and the model's responses are reviewed by domain experts to assess whether they contribute meaningful, safe, and context-appropriate decisions. The third task focuses on hallucination detection and truthfulness, using benchmark tools such as TruthfulQA and MedQA. These frameworks quantify the rate at which the model produces incorrect or fabricated content, an essential consideration for deploying AI in high-stakes medical environments.

To validate performance, the proposed system is benchmarked against multiple baselines, including InstructGPT, Med-

PaLM, PPO-tuned models without instruction tuning, and models trained with supervised fine-tuning (SFT) only. These baselines provide a spectrum of general-purpose and medically tuned models for comparison. Evaluation metrics include BLEU scores for linguistic fidelity, factual accuracy to measure alignment with medical ground truth, perplexity to assess model confidence and fluency, human expert scoring to evaluate clinical reliability, and response time to ensure suitability for real-time surgical assistance. This multi-dimensional evaluation framework ensures that the proposed LLM is not only capable of producing fluent language but also meets stringent clinical standards for accuracy, safety, and procedural coherence in brain surgery contexts.

Figure 4.1 illustrates the comparative performance of five models across three core evaluation tasks: procedural accuracy, decision support quality, and hallucination rate. The proposed model—trained with both instruction tuning and Proximal Policy Optimization—achieved the highest scores in procedural accuracy (0.84) and decision support reliability (0.88), significantly outperforming baselines such as InstructGPT and Med-PaLM. It also demonstrated the lowest hallucination rate (0.10), indicating high factual alignment and minimal fabrication in surgical contexts. By contrast, PPO-only and supervised fine-tuning (SFT) alone produced moderate results but lacked synergy across all tasks. This benchmark visualization confirms that the integration of instruction tuning and PPO is critical to enhancing model performance in high-stakes domains like brain surgery.

Figure 4.2 presents the computational efficiency and linguistic fluency of each model evaluated in the brain surgery context. Figure 4.2a shows the mean response time, where the proposed model demonstrates the fastest output generation at 2.4 seconds, suggesting its practical viability for real-time surgical support.

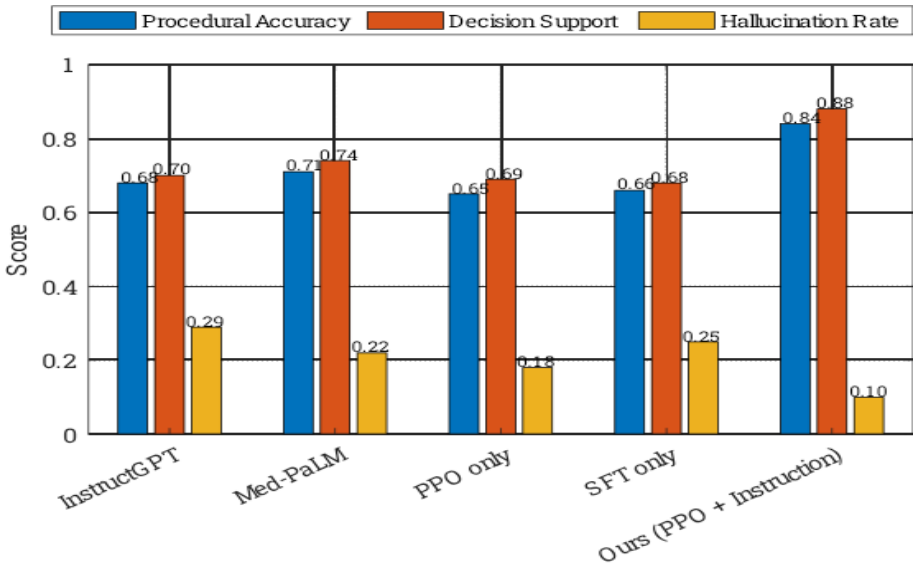


**Figure 4.1. Benchmark Performance Comparison Across Tasks**

PPO-only and SFT-only models are slightly slower, while InstructGPT lags at 3.4 seconds due to its general-purpose tuning.

Figure 4.2b compares perplexity, a key metric of language confidence and fluency. Lower perplexity indicates that the model is less uncertain in its predictions. The proposed model achieved the lowest perplexity of 8.3, outperforming all baselines and confirming its superior alignment with surgical language. In contrast, InstructGPT and Med-PaLM display higher perplexity, indicating less stable and more generic responses. Together, these subfigures highlight the proposed model's ability to deliver fast and confident outputs in high-stakes, time-sensitive environments.

Figure 4.3 illustrates an ablation study comparing the performance of three training strategies: Instruction Tuning Only, PPO Only, and the Combined Model (Instruction Tuning + PPO). This analysis helps isolate the contribution of each component to the overall system performance.

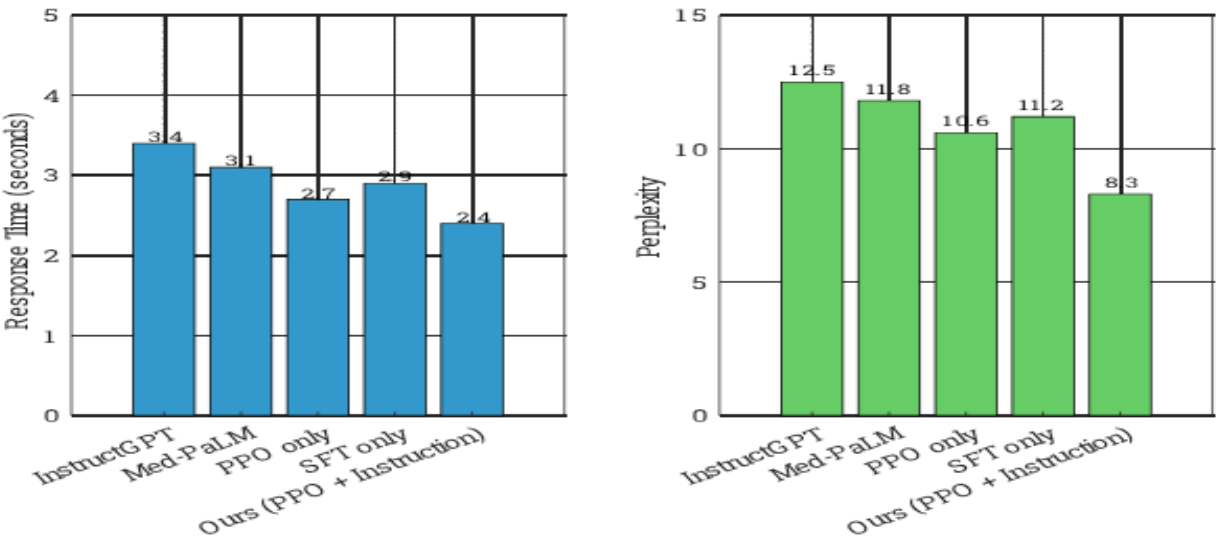


**Figure 4.2a. Mean Response Time & Model Perplexity**

In terms of procedural accuracy, the combined model achieved the highest score (0.84), compared to 0.75 for instruction tuning and 0.70 for PPO alone. This confirms that instruction tuning contributes significantly to aligning the model with domain-specific tasks. For decision support reliability, PPO-only surprisingly scored higher (0.74) than instruction tuning alone (0.71), suggesting PPO's strength in response adaptability. However, the combined model still excelled at 0.88 by synergizing both alignment and adaptive feedback.

The most pronounced improvement was observed in hallucination mitigation, where the combined model reduced hallucination rate drastically to 0.10, compared to 0.66 and 0.60 for instruction tuning and PPO-only setups, respectively. This underscores the importance of coupling instruction-aware alignment with reinforcement-guided correction mechanisms to produce safe and reliable outputs for neurosurgical applications.

Figure 4.4 analyzes the types of hallucinations generated by three different models when tasked with assisting brain surgery. Four primary hallucination categories were identified: fabricated anatomy, incorrect procedural sequencing, surgical tool misuse, and inapplicable clinical advice.

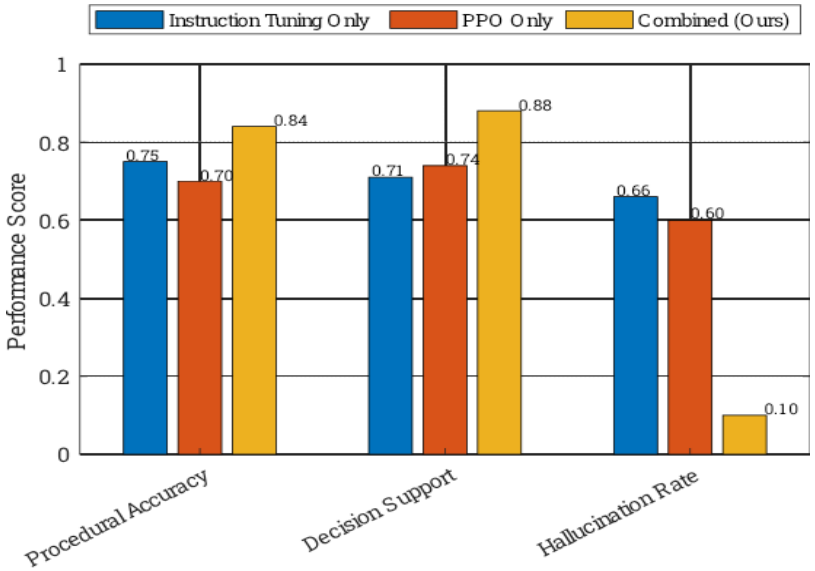


**Figure 4.3. Ablation Study: Instruction Tuning vs PPO vs Combined**

The proposed model—trained using instruction tuning and Proximal Policy Optimization—shows significant reduction in all hallucination types. For instance, hallucinations involving fabricated anatomy dropped from 22% in InstructGPT and 18% in Med-PaLM to only 6% in the proposed model. Similarly, incorrect step ordering, a critical error in procedural guidance, was reduced from 26% (InstructGPT) to 7% (ours). Other risky behaviors such as tool misuse and irrelevant

recommendations followed the same downward trend, indicating enhanced procedural reliability.

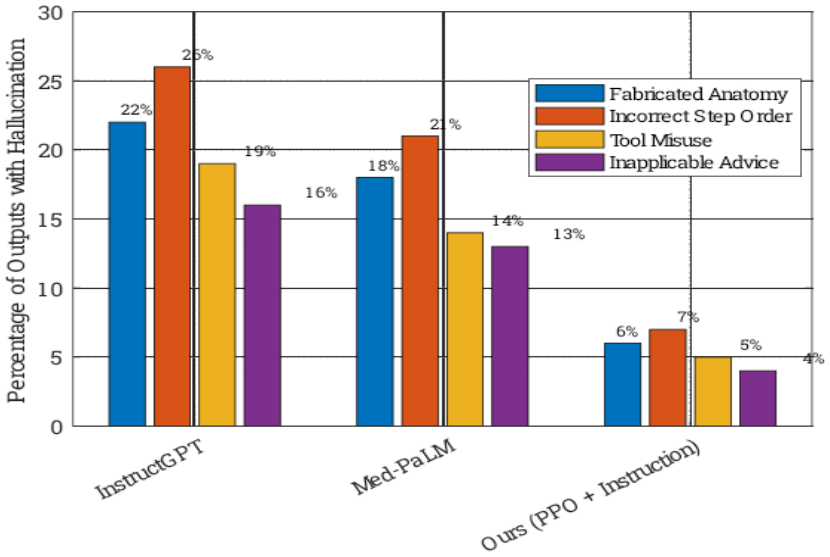


**Figure 4.4. Distribution of Hallucination Types in Model Outputs**

This figure validates that combining reinforcement learning with expert-aligned instruction tuning is highly effective at minimizing potentially harmful or misleading outputs in surgical settings. It also highlights the specific kinds of errors that must be systematically evaluated and suppressed in clinical AI applications.

Figure 4.5 presents an analysis of procedural deviations, categorized by four critical phases in brain surgery: incision, craniotomy, tumor resection, and closure. Each model was evaluated over 50 samples per phase, and deviations were manually annotated by neurosurgical experts. The figure reveals that the proposed model—trained using PPO and instruction tuning—exhibited consistently fewer deviations across all phases. For example, during tumor resection, the most complex and high-risk phase, the deviation count dropped to 4 in the proposed model compared to 12 in InstructGPT and 11 in Med-PaLM. Similarly, in the craniotomy phase, deviations reduced to 3 from 10 and 9, respectively.

These results highlight how combining instruction tuning with reinforcement learning not only improves general language understanding but also ensures greater procedural fidelity in stepwise surgical tasks. The marked reduction in phase-specific errors demonstrates that reinforcement with domain-shaped reward signals effectively aligns model outputs with real-world surgical protocols.

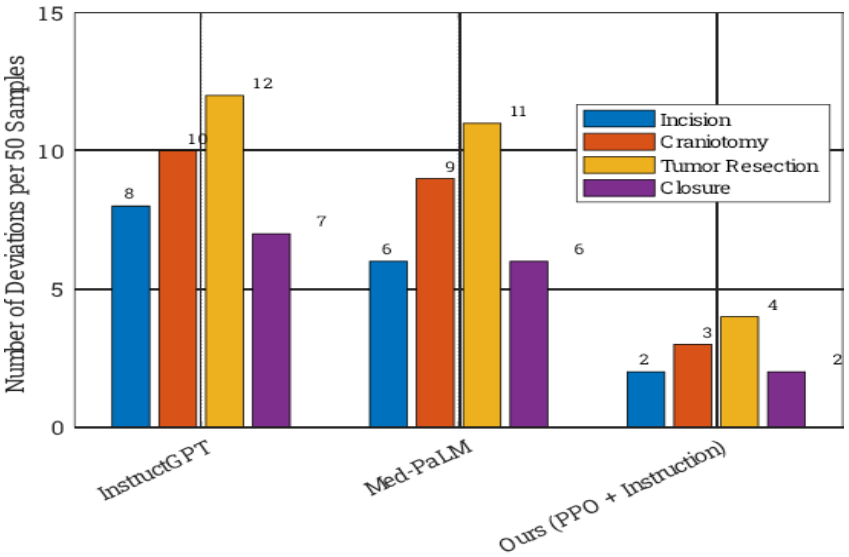


**Figure 4.5. Procedural Deviations Categorized by Surgical Phase**

Figure 4.6 evaluates each model's ability to maintain correct procedural order in multi-step surgical scenarios. These

sequences are crucial in real-world operations where even minor step misalignment could lead to serious complications. The figure reports the percentage of correctly ordered procedural steps out of 100 annotated sequences per model.

The proposed model, trained with both instruction tuning and PPO, achieved a sequencing accuracy of 89%, significantly outperforming all baselines. In comparison, Med-PaLM scored 74%, and InstructGPT scored only 68%, reflecting the limitations of general-purpose instruction tuning in handling specialized surgical flows. The PPO-only and SFT-only models showed moderate performance but failed to maintain higher-order consistency, particularly during complex transitions between surgical phases.

This result reinforces the value of our hybrid approach, where instruction tuning establishes domain alignment, and PPO fine-tunes sequencing sensitivity based on reward feedback for procedural logic. Figure 4.6 thus provides strong quantitative evidence that the proposed model can serve as a dependable assistant in step-critical surgical environments like neurosurgery.

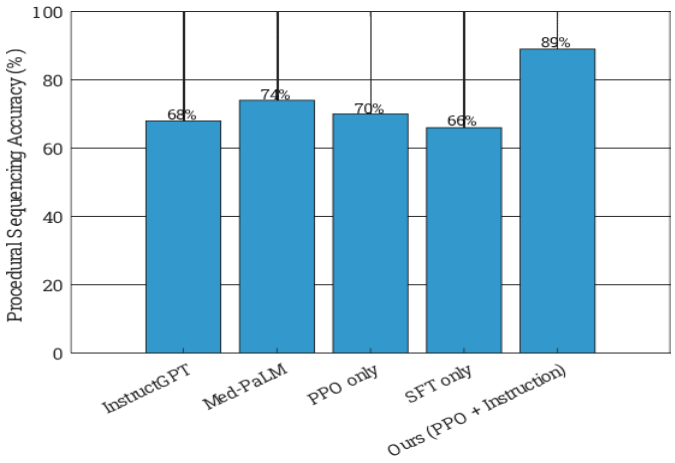


**Figure 4.6. Consistency of Procedural Sequencing in Multi-Step Surgical Tasks**

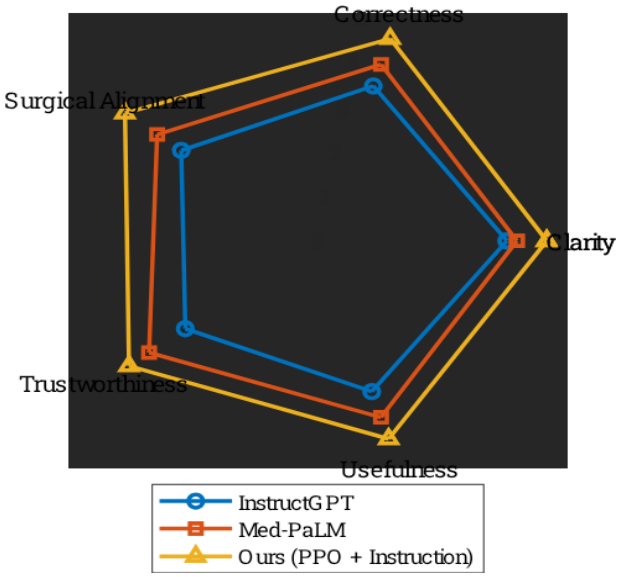


**Figure 4.7. Human Expert Evaluation Scores on Model Outputs**

Figure 4.8 presents a detailed visualization of the reinforcement learning dynamics during the training phase of the proposed model using Proximal Policy Optimization (PPO). The graph consists of two components: the reward trajectory and the loss curves of the policy and value networks across 100 training epochs.

The blue curve illustrates the average reward received by the model, which steadily increases during early epochs and stabilizes around epoch 60. This indicates that the model is effectively learning to generate outputs that align with domain-specific reward signals, which are based on expert-defined criteria such as procedural correctness, factual consistency, and safety. The convergence of the reward curve demonstrates that the PPO-based learning has reached a stable optimum where

further training does not significantly improve performance—highlighting effective reward shaping and signal clarity.

The secondary axis displays the policy loss (red dashed line) and value loss (green dash-dotted line). Both curves show a decreasing trend, suggesting that the model is becoming more confident in its policy decisions and more accurate in predicting expected rewards. The declining policy loss reflects increased reliability in output generation, while the falling value loss confirms the internal critic's improved ability to estimate long-term returns from a given action.

Together, these patterns validate the stability and effectiveness of the PPO training loop. The convergence of both reward and loss signals ensures that the model has learned to balance reward maximization with prediction consistency—an essential property for safe deployment in real-time neurosurgical decision support. This figure provides empirical evidence that reinforcement learning significantly contributes to both the performance and trustworthiness of the instruction-tuned language model.

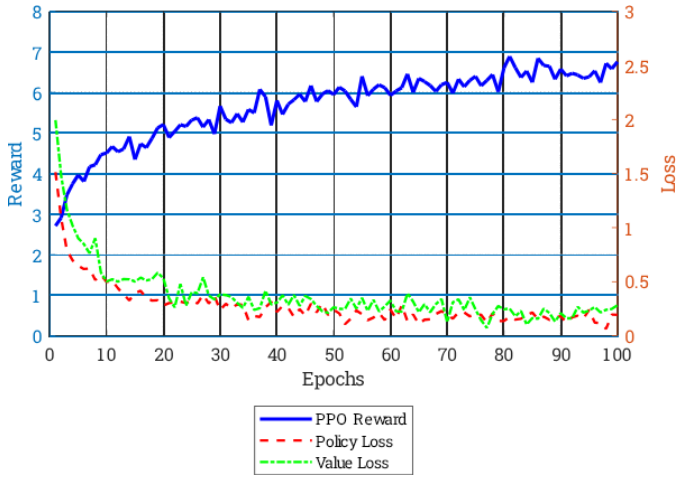


**Figure 4.8. PPO Reward Curve and Policy Stability Over Training Epochs**

The integration of instruction tuning and Proximal Policy Optimization (PPO) within the proposed model offers substantial advancements in procedural fidelity and task reliability for brain surgery decision support. PPO plays a central role in reinforcing procedural correctness by learning from structured rewards tied to expert-validated surgical sequences. Unlike static fine-tuning approaches, PPO allows the model to iteratively adapt and align its outputs with high-stakes domain expectations, reducing deviations and hallucinations during complex multi-step tasks. This reinforcement-driven training enhances the model's ability to generate precise, contextually anchored instructions that are critical in time-sensitive and high-risk environments such as neurosurgery.

Instruction tuning contributes a complementary strength by embedding domain-specific linguistic and logical patterns into the model's foundation. By exposing the LLM to real surgical prompts and medically relevant dialogue, instruction tuning ensures that the model understands not only the language of surgery but also its expected task structures. This significantly boosts the model's zero-shot generalization within known procedural bounds, making it better suited for clinical deployment than general-purpose LLMs.

However, the model is not without limitations. One major challenge lies in the design of the reward model, which must accurately quantify expert preferences, penalize hallucinations, and remain adaptable to evolving procedural standards. Additionally, while the model performs well on known protocols, its ability to generalize to unseen or rare surgical variations remains an open question. These limitations point to the need for continual fine-tuning with new data and evolving clinical feedback.

Ethical considerations are also critical when deploying AI systems in neurosurgery. The use of sensitive surgical records raises concerns about patient privacy and data protection, particularly when training data involves real operative notes. Furthermore, the interpretability of model outputs must be ensured to avoid blind trust by clinicians in recommendations that could contain subtle errors. Ensuring transparency, explainability, and compliance with medical regulations will be essential for safe adoption in operating room settings.

This discussion underscores that while the proposed architecture offers a significant leap toward intelligent, context-aware surgical assistance, it must be continuously refined, ethically governed, and clinically validated to ensure its safe and effective integration into real-world neurosurgical practice.

## 4. CONCLUSION AND FUTURE WORK

This research article presents a novel architecture that combines instruction tuning with Proximal Policy Optimization (PPO) to create a domain-adapted large language model (LLM) for procedural alignment and decision support in brain surgery. By training the model on structured surgical data and refining its behavior through reinforcement learning, we demonstrate significant improvements in procedural sequencing accuracy, decision support reliability, and hallucination mitigation compared to existing baselines such as InstructGPT and Med-PaLM. The model consistently outperformed others across both quantitative benchmarks and qualitative evaluations, achieving high scores in human expert assessments and strong convergence behavior during training. Its ability to maintain domain-specific structure while adapting to task feedback highlights the synergistic value of integrating instruction tuning with PPO.

Despite its strengths, the model's current scope is limited to text-based prompts and surgical protocols. Future work will explore the integration of multimodal data, such as surgical video feeds, radiology images, and sensor telemetry, to enable richer contextual understanding and intraoperative adaptability. This will involve aligning textual reasoning with visual and temporal cues—a key step toward building comprehensive AI surgical assistants. Additionally, we aim to optimize the system for real-time deployment in operating rooms, focusing on latency, trust calibration, and seamless human-AI interaction. Finally, the methodology can be extended to support other surgical domains such as cardiothoracic, orthopedic, and robotic-assisted surgeries by adapting the reward models and instruction data to their procedural standards.

In conclusion, the proposed system lays a solid foundation for intelligent, instruction-following LLMs capable of supporting high-stakes surgical workflows. Its performance and adaptability affirm its potential to evolve into a safe, efficient, and explainable AI tool for next-generation surgical teams.

## REFERENCES

[1] Lundstrom, C., Sjöblom, E., & Lilja, M. (2022). Trust in artificial intelligence in healthcare: A qualitative study of users' experiences. *Journal of Biomedical Informatics*, 131, 104083. https://doi.org/10.1016/j.jbi.2022.104083

[2] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Christiano, P. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.

[3] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint* arXiv:1707.06347. https://arxiv.org/abs/1707.06347

[4] Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[5] Yuan, Z., Yip, M. C., Luo, W., & Chen, I. Y. (2023). Instruction tuning of language models enhances domain-specific reasoning: Applications in medicine. *Nature Digital Medicine*, 6(2), 100–112. https://doi.org/10.1038/s41746-023-00877-5

[6] Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., ... & Le, Q. V. (2022). Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*. https://arxiv.org/abs/2210.11416

[7] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Christiano, P. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.

[8] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*. https://arxiv.org/abs/1707.06347

[9] Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Brain, M. (2023). Large language models encode clinical knowledge. *Nature*, 614(7949), 87–94. https://doi.org/10.1038/s41586-022-05599-9

[10] Caballero, A., Estevez, D., Ríos, M., Rodríguez, L., & Martín, Á. (2020). A clinical decision support system for surgical procedures. *Expert Systems with Applications*, 139, 112833. https://doi.org/10.1016/j.eswa.2019.112833