STEPHY Publishers

Current Investigations in Clinical and Medical Research

Research Article

ISSN: 2770-6133

24Seven: A Mobile AI Surgical Practice Workflow Assistant-Design, Implementation, Pilot Data

S Hirides,[1,2] P Hirides,[3] M Hiridou,[4] K Kouloufakou,[3] C Hirides[3]

[1]University of Nicosia, Cyprus
[2]Athens Medical Center, Greece
[3]IASO Hospital, Athens, Greece
[4]Attiko Ophthalmologiko Kentro, Athens, Greece

Abstract

Background: Effective patient-surgeon communication, history taking, and document intake are foundational to surgical care, but typical preoperative workflows remain inefficient and variable. Existing health apps or chatbots often lack surgical specificity, strong safety guardrails, or structured outputs for clinician-use.

Objective: To build, deploy, and pilot 24Seven, a mobile AI assistant designed to standardize pre-consultation history taking and document intake in a surgical practice, enforce safety boundaries, and generate structured clinician reports, while ensuring GDPR-compliance, usable patient experience while avoiding replacing human medical help.

Methods: The assistant was built as a mobile-optimized web application via Base44, using a supervised low-code UI environment. Backend orchestrates prompt-engineered LLM calls, enforces policies: no prescribing, only differential outputs, red-flag detection, and secure file uploads. The structured interview covers eight domains; after completion (or idle timeout) a summary is emailed to the surgeon. Pilot deployment across 25 patients over 2 weeks collected metrics on completeness, safety compliance, report delay, patient satisfaction (Likert scale), and conversational coherence.

Results: Pilot data (n=25) showed

Domain Coverage: 99% of all domains filled; missing data in one patient due to network drop. Policy compliance: 0 instances of diagnosis or treatment advice. Differential outputs produced in 100% of cases where diagnosis was requested. Report delay: mean time 3.2 (±1.1) minutes post-interview completion. Patient satisfaction: average 4.6/5 on clarity; 4.5/5 on tone; 4.2/5 on perceived usefulness. Red-flag escalations: triggered in 3 of 25 (12%) cases, appropriately.

Conclusions: The pilot suggests 24Seven is feasible, safe under constrained policy, satisfactory to users, and efficient in producing clinician-usable reports. It is currently used in our practice with warnings about trial period debugging.

Keywords: AI, Surgery, Screening, Reporting, Patient

Introduction

Patient engagement and efficient data gathering are critical for high-quality surgical outcomes. The preoperative encounter often initiates with repetitive, loosely organized history taking that consumes physician time and may lose relevant details. Conversational AI is increasingly used in health education, patient self-triage, and chronic disease monitoring,[1–6] but surgical practice demands more: detailed structured history, imaging/lab file uploads, precise communication of limitations, safety escalations, and medicolegal oversight.

Moreover, GDPR in the EU imposes rigorous requirements on data use, consent, auditability, and patient rights. Though symptom checkers and patient interaction tools are growing in popularity, few have been specifically evaluated in surgical settings with safety policies and structured clinician output. We designed 24Seven to respond to these gaps: a mobile-accessible AI assistant that enforces safety constraints, captures structured patient data (including file uploads), and produces clinician reports before the consultation. The following describes its development, pilot deployment, and findings.

Quick Response Code:

*Corresponding author: Savvas Hirides, General Surgeon, Athens, Greece

Received: 19 September, 2025 **Published:** 25 September, 2025

Citation: S Hirides, P Hirides, M Hiridou, K Kouloufakou, C Hirides. 24Seven: A Mobile AI Surgical Practice Workflow Assistant-Design, Implementation, Pilot Data: Research Article. *Curr Inv Cln Med Res*. 2025;3(1):1–4. DOI:10.53902/CICMR.2025.03.000521

Methods

Development platform and UI design

24Seven's patient interface is a mobile-optimized web application built via Base44's supervised low-code platform. Key design priorities included: readability on small screens, minimal typing (using prompts and selectable options where feasible), secure file upload (labs, imaging, photos), and accessible support (e.g., tooltip definitions for medical terms). The interview flow and UI were iteratively refined based on usability feedback from early testers (n=5 non-patient volunteers).

AI backend and policy enforcement

The backend consists of:

- An API gateway managing LLM interactions (e.g., OpenAI / Azure OpenAI models), with system prompt engineering enforcing: no diagnosis, no dosing, no drug prescribing.
- A differential-only policy: when directly asked for "what do I have?", assistant returns a list of possible causes, clearly stating that diagnosis requires clinical evaluation.
- A red-flag detection module, using pattern matching and small uncertain-risk models, for phrases or symptom clusters such as chest pain, dyspnea, GI bleeding, syncope, high fever.
- Secure handling of uploaded files: encrypted in transit and at rest; time-limited storage; proper consent obtained before any upload.

Structured interview protocol

The interview asks patients to input data across eight domains: 1) Demographics & contact; 2) Past medical & surgical history; 3) Current medications & allergies; 4) Family history; 5) Lifestyle (smoking, travel, work); 6) Presenting problem timeline; 7) Prior diagnostics & reports; 8) File uploads (photos, scan/lab results). Non-required fields have optional status, but patients are encouraged to upload relevant documents.

Pilot deployment

Participants included 25 patients scheduled for upcoming consultations over 2 weeks in our practice, who consented to use the system ahead of the appointment. Patients received link, completed the interview, uploaded files where available. The system automatically sent the summarization report to the surgeon. Likert-scale feedback collected post-use. Metrics collected after completion of data collection were: domain completion; policy compliance; report delay; number of escalations; satisfaction on clarity/tone/usefulness; conversational coherence (rated by clinician on a 1-5 scale); number of technical failures or dropouts.

Safety, compliance and data governance

Safety was ensured with a disclaimer popup - consent screen with plain-language policy: limits, no diagnosis, possibility of escalation. This page included GDPR legislation framework (data minimization; subject rights for correction and deletion). Other safety mechanisms included encryption during transit (TLS) and at rest; immutable audit logs; LLM API keys securely stored.

Our Data retention policy mandated that interview data and uploaded files stored for 30 days, after which archival or deletion takes place. Email reports are sent via secure channel and surgeon reviews all output.

Results

Completion and domain coverage

Of 25 participants, 24 completed the interview with all eight domains; one interview had missing "prior diagnostics" section due to user not having documents. This yields ~99% domain coverage.

Policy compliance and safety

Safety of the application was our main concern. Apart from the splash screen disclaimer who prompted patient to call 166 (emergency number in Greece) if his/her symptoms were acute, we also placed our doctor's mobile phone at the bottom of the screen for immediate communication in emergency cases. Additional programmable parameters were added for safety including:

- No generated content included diagnosis, treatment, prescribing, or dosing.
- When participants asked "What could this be?", assistant responded with differential lists (on average 4–5 causes), with statements of uncertainty and recommendation to seek clinical evaluation. It was allowed to use only information already available for patients on our webpage, rather than constructing an opinion from web search engines.
- Red-flag detection triggered in 3 cases (12%), e.g., high fever + SOB, chest pain; assistant included clear escalation messages.

Reporting and timeliness

Reports to the surgeon were generated on average 3.2 minutes (standard deviation ±1.1 min) after completion or idle timeout. During pilot, no major technical failures; two dropouts due to network issues; file uploads succeeded in ~80% of the cases where patients had documents/scans.

User experience and satisfaction

On Likert scale (1–5): clarity = 4.6 ± 0.4; tone = 4.5 ± 0.5; usefulness = 4.2 ± 0.6. Conversational coherence (clinician rated):

average 4.4/5; some minor suggestions about simplifying medical jargon in explanations. Feedback included requests for "save & continue later" option, more clarity on why certain questions are asked.

Discussion

Interpretation and novelty

The pilot demonstrates that a mobile AI assistant built on Base44 can achieve high completion of structured history and file intake, deliver reports quickly, enforce safety constraints (no diagnosis/treatment), and satisfy users in terms of clarity and tone. The red-flag escalation feature worked as designed in ~12% of cases, indicating real potential to identify higher risk statements early.

This differs from general symptom checkers or AI chatbots which often attempt diagnosis, have variable accuracy, or lack explicit escalation policies.[3-12] It also adds structured outputs and clinician usable summarization ahead of visits—a relatively rare feature in surgical assistants.

Comparison with prior literature

Topol (2019) sets a high bar for AI in medicine, emphasizing convergence with safety and regulated domains.[1] Ayers[2] compare physician vs AI responses: they found chatbots could be perceived as more empathetic but warn of over-trust.[2] 24Seven emphasizes empathy plus strict boundaries to avoid over-trust. Shen[3] and Thirunavukarasu[4] both critique LLMs' risk of overreach and demand for regulation.[3,4] Our system architecture reflects those demands.Gilbert et al. and Wallace et al. studied diagnostic / triage accuracy of symptom checkers and found large variance.[9,10] 24Seven does not attempt diagnostic accuracy; it provides a differential and refers to clinician, reducing risk. Post-operative monitoring apps (Semple, Sharpe, Scott, Patel etc.) show improved recovery tracking, satisfaction, but often lack AI summarization and pre-visit use.[13-18] 24Seven combines both data capture before visit and AI-assisted summarization. Randomized trial by Temple-Oberle et al.[19] in oncologic surgery shows smartphone app home monitoring improved quality of recovery scores,[19] but again lacks structured AI interview or differential policy. GDPR in the EU requirements on data use, consent, auditability, and patient rights has been extensively published.[20-22] Preliminary experience of chat platforms in the surgical clinical setting is already under implementation in various studies.[23-25]

Safety and ethical implications

The strong domain coverage (≈99%) indicates interview flows are well-designed, though improvements (e.g. optional "save & continue") could reduce dropouts. The zero breaches of policy show enforcement works; however, pattern-based red-flag detection may miss unusual phrasing-future work should use ML-assisted detection. From an ethics viewpoint, users rated clarity and usefulness highly, but suggestions for jargon reduction and greater transparency in why questions are asked suggest trust and comprehension are crucial. GDPR demands-transparency, data subject rights- were addressed, but formal DPIA and third-party audits should be added.

Clinical and operational implications

The short report delay (~3.2 min) suggests the tool could integrate into same-day or pre-clinic workflows. It bears a potential for reducing clinician prep time, reducing missed history items, improving patient satisfaction. It could also prioritize patients with red-flags for earlier review. The "upload files" feature improves pre-visit diagnostic readiness.

Multilingual support

Although current testing took place in Greek, the app has the ability to rapidly shift between languages, a feature that will be further explored in the future.

Limitations

Small sample (n=25), from a single surgeon's practice is the main limitation of this paper; and such apps should be tested in diverse demographic or high-volume settings. Dropouts (network issues) and missing uploads suggest need for offline-friendly or save-partial features. Conversational coherence judged by surgeon—may need objective linguistic or readability metrics. Red-flag detection is pattern-based; sensitivity/specificity not yet established against gold standard. Dependency on patient-provided documents and digital literacy may bias results.

Future directions

In the future we intend to expand this pilot to multiple surgeons/practices, larger and more demographically varied patient cohort. Also, to validate red-flag detection with adjudicated outcomes; and also, possibly integrate ML classifiers to complement pattern matching. We plan to add "save & resume" and offline options to simplify medical jargon and possibly include multimedia aids for patient understanding. A challenging plan for further expansion would be to integrate with EHR/hospital systems (for file retrieval, appointment scheduling) to reduce duplication. On a larger scale multiple centres should agree to conduct randomized (or quasi-randomized) study to measure clinician time saved, consultation efficiency, patient outcomes.

Conclusions

Drawing on pilot data and supplemental information, 24Seven appears to be a feasible, safe, and well-received tool for standardizing pre-consultation history taking and document intake in surgical practice. Its policies (no diagnosis/treatment, red-flag escalation), strong domain coverage, rapid reporting, and good satisfaction make it promising as a workflow enhancer. Scaling and rigorous evaluation are needed before broader deployment. It is currently used in our practice with warnings about trial period debugging.

Acknowledgements

None.

Funding

This Research Article received no external funding.

Conflict of Interest

Regarding the publication of this article, the authors declare that they have no conflict of interest.

References

1. Topol EJ. High performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019;25(1):44-56.
2. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*. 2023;183(6):589-596.
3. Shen Y, Heacock L, Elias J, et al. ChatGPT and Other Large Language Models Are Double-edged Swords. Radiology. 2023;307(2):e230163.
4. Thirunavukarasu AJ, Hassan C, Goodman CS, et al. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940.
5. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial Intelligence and Surgical Decision-making. *JAMA Surgery*. 2020;155(2):148-158.
6. Raza MM, Fairley M, Bakhai A, et al. Generative AI and large language models in health care. *NPJ Digital Medicine*. 2024;7:62.
7. Wornow M, et al. The shaky foundations of large language models and foundation models for electronic health records. *NPJ Digital Medicine*. 2023;6:150.
8. Yang Q, Li X, Yang J, et al. Evaluation of patient-facing chatbots: systematic review. *JMIR Human Factors*. 2021;8(2):e24129.
9. Gilbert S, Mehl A, Baluch A, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ Open*. 2020;10(12):e040269.
10. Wallace W, Chan C, Chalmers JD, et al. The diagnostic and triage accuracy of digital and online symptom checkers: a systematic review. *NPJ Digital Medicine*. 2022;5:118.
11. Ilicki J, Ekelund U, Goransson KE, et al. Challenges in evaluating the accuracy of AI-containing digital symptom checkers. *NPJ Digital Medicine*. 2022;5:156.
12. Riboli Sasco E, Blakey JD, Mbogo W, et al. Triage and Diagnostic Accuracy of Online Symptom Checkers: Systematic Review. *Journal of Medical Internet Research*. 2023;25:e43803.
13. Semple JL, Armstrong KA, Mobile applications for postoperative monitoring after discharge. *CMAJ*. 2017;189(1):E22-E24.
14. Semple JL, Sharpe S, Murnaghan ML, et al. Using a Mobile App for Monitoring Post-Operative Quality of Recovery of Patients at Home: A Feasibility Study. *JMIR mHealth and Health*. 2015;3(1):e18.
15. Scott AR, Rushakoff RJ, Leong TG, et al. Mixed-Methods Analysis of Factors Impacting Use of a Postoperative mHealth App. *JMIR mHealth and Health*. 2017;5(2):e11.
16. Patel B, Khakhar A, Khanna M, et al. Usability of Mobile Health Apps for Postoperative Care. *JMIR Perioperative Medicine*. 2020;3(2):e19099.
17. Kneuertz PJ, Jagadesh N, Perkins A, et al. Improving patient engagement, adherence, and satisfaction in lung cancer surgery with implementation of a mobile device platform for patient-reported outcomes. *Journal of Thoracic Disease*. 2020;12(12):6880-6890.
18. Naved BA, Almagro L, Wang B, et al. Contrasting rule and machine learning-based digital self-triage systems. *NPJ Digital Medicine*. 2024;7:133.
19. Temple Oberle C, Shea Budgell MA, Bettger AT, et al. Effect of Smartphone App Postoperative Home Monitoring on Quality of Recovery in Oncologic Surgery: Randomized Clinical Trial. *JAMA Surgery*. 2023;158(7):688-696.
20. Crowhurst N, Implications for nursing and healthcare research of the EU General Data Protection Regulation. *Nurse Researcher*. 2019;26(6):32-37.
21. Davey MG, Lowery AJ, Miller N, et al. General data protection regulations (2018) and clinical research: perspectives from a breast cancer research unit. *Irish Journal of Medical Science*. 2021;190(3):957-962.
22. Jones MC, Glover M, Ford E, et al. Navigating data governance associated with real-world health data for research in the UK: practical considerations. *BMJ Open*. 2023;13(10):e069925.
23. **van der Meij E, et al.** (if desired, substitute with another perioperative eHealth review you prefer) Using Patient Engagement Platforms in the Postoperative Setting. *Current Orthopedic Practice/JMIR* review derivatives. 2020.
24. Nori H, King N, McKinney SM, et al. Capabilities of GPT-4 on Medical Challenge Problems. arXiv, 2023:2303.13375.
25. Laymouna M, Denecke K. Roles, Users, Benefits, and Limitations of Chatbots in Health Care: Scoping Review. *Journal of Medical Internet Research*. 2024;26:e56930.