LINK MEDICAL JOURNAL
of Health and Community Research | LMJHCR
Volume 4, Issue 1 | June 2026 | ISSN 3007-3448

Indexed in DOAJ | CC BY 4.0
DOI: 10.61919/8xsycj49

Original Article

Trust and Consent Challenges in AI-Supported Surgical Planning: A Qualitative Study

Zeng Wenchao[1], Chrismas Gideon Bangun[1], Arna Fransisca Millyanti Purba[1]

[1] Department of Surgery, Universitas Prima, Indonesia

***Corresponding author: Chrismas Gideon Bangun, chrismasgbangun@unprimdn.ac.id**

"Cite this Article" | **Received**: 20 December 2025; **Accepted**: 23 April 2026; **Published**: 22 June 2026.

ABSTRACT

Background: Artificial intelligence is increasingly used in surgical planning through imaging interpretation, anatomical modelling, risk estimation and decision support. Although these tools may support clinical preparation, they create consent challenges when patients are not told how AI influenced the plan, how their data are used or who remains responsible for errors. Qualitative inquiry is needed to understand how patients interpret AI involvement, trust and accountability in surgical consent. Objective: To explore patient perspectives on trust and consent in AI-supported surgical planning, focusing on disclosure, explainability, surgeon responsibility, privacy, fairness and accountability. Methods: This qualitative descriptive study used semi-structured participant accounts from 12 purposively selected surgical patients and patient-related stakeholders. Participants represented variation in surgical experience, digital confidence, family decision-making roles and attitudes toward AI-supported planning. Data were analyzed using reflexive thematic analysis informed by sensitising concepts from AI ethics and inductive interpretation of participant concerns. Results: Six themes were identified: unclear boundaries between surgeon judgement and AI advice; consent without meaningful AI disclosure; explainability and patient understanding; data privacy, bias and fairness concerns; accountability for planning errors; and relational trust in the surgeon as the consent anchor. Participants did not reject AI-supported planning in principle, but acceptance depended on plain-language disclosure, visible surgeon responsibility, transparent data governance and clear institutional accountability. Conclusion: Consent for AI-supported surgical planning should be treated as a patient-centred communication process. Sustainable implementation requires proportionate AI disclosure, clinician explanation, documentation of AI use, governance safeguards and explicit reassurance that surgeons remain responsible for final clinical decisions. Keywords: artificial intelligence; surgical planning; informed consent; trust; accountability; qualitative study.

EDITORIAL INFORMATION

Author Contributions: Concept: ZW; Literature Review: CGB; Drafting: AFMP; Critical Revision and Final Approval: ZW, CGB, AFMP
Ethical Approval: Department of Surgery, Universitas Prima, Indonesia
Informed Consent: Written informed consent was obtained from all participants
Conflict of Interest: The authors declare no conflict of interest; **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

INTRODUCTION

Artificial intelligence is increasingly becoming part of surgical care through imaging interpretation, anatomical segmentation, risk estimation, operative planning support and decision assistance. In surgical planning, these functions may help clinicians integrate complex radiological, anatomical and clinical information before an operation, particularly when the procedure involves difficult anatomy, high-risk decision-making or multiple possible operative approaches. Early surgical AI literature has described both the promise and the risks of machine learning in operative practice, especially where algorithmic outputs begin to influence clinical judgement and patient-facing decisions (1). The expansion of surgical data science has further increased the relevance of this issue, as contemporary surgical AI may combine

imaging, workflow data, device-generated data and computational modelling to support decisions before and during surgery (2). Wider literature on AI in health and medicine similarly indicates that AI can support diagnosis, prediction and clinical decision-making when systems are carefully developed, evaluated and implemented within accountable clinical workflows (3).

The clinical promise of AI does not remove the ethical complexity of consent. When a patient agrees to a surgical plan, the patient may not know whether AI has influenced image interpretation, risk estimation, anatomical modelling, procedure selection or the surgeon's recommendation. The consent problem is therefore not simply whether AI is present in the hospital system, but whether its contribution is material to the patient's decision to proceed with surgery. Surgical consent is a communication process in which patients should understand the nature of the proposed intervention, relevant alternatives, expected benefits, material risks, uncertainties and the basis on which clinical advice is offered. If AI-supported planning remains hidden inside the clinical workflow, patients may feel excluded from an important part of the decision-making process, even when the technology is being used as an assistive rather than autonomous tool.

Global guidance on AI in health emphasizes transparency, explainability, accountability, inclusivity, privacy, fairness and human oversight as central principles for trustworthy AI implementation (4). These principles are particularly important in surgical planning because decisions are technically complex, often time-sensitive and directly affect the patient's body. Current debates about AI and informed consent ask whether patients should always be informed when AI is involved, how much information is necessary for meaningful consent and whether existing consent frameworks are sufficient for AI-enabled care (5). Surgical ethics literature also suggests that AI may alter patient expectations about responsibility, safety, explanation and professional accountability when it contributes to surgical practice (6). These concerns are not only legal or technical; they affect how patients interpret trust, autonomy and responsibility during consent conversations.

Implementation research on clinical AI shows that technical performance alone is insufficient for clinical impact because AI tools require workflow fit, local evaluation, clinical oversight and monitoring after deployment (7). In surgery, this means that an AI-generated segmentation, risk score or planning recommendation cannot replace the surgeon's responsibility to interpret the output, assess its relevance to the individual patient and explain the final plan in understandable terms. Trust in this setting is relational as well as technological. Patients may trust or distrust AI not only because of the system's accuracy, but because of how the surgeon explains it, whether the institution appears accountable and whether the patient feels able to ask questions. Literature on trust in healthcare AI therefore argues that the aim should not be uncritical patient acceptance of technology, but trustworthiness in the clinicians, institutions and systems that use it (8).

AI-supported planning also raises accountability and data governance concerns. Responsibility may become distributed across clinicians, hospitals, developers, vendors and regulators when AI contributes to clinical care (9). Regulatory developments, including risk-based governance frameworks and guidance for AI and machine-learning software as a medical device, emphasize safety, transparency, performance evaluation and lifecycle monitoring (10,11). However, regulatory compliance does not automatically produce patient understanding during consent. A system can be technically authorized and still be poorly explained to the patient. Similarly, explainability in healthcare AI cannot be reduced to technical interpretability. Patients generally need to know what the AI was used for, what it did not do, how the surgeon interpreted or challenged its output, what uncertainty remained and who remains responsible for the final plan (12,13).

Data privacy, bias and fairness are also relevant to consent because AI-supported planning depends on imaging data, medical records, operative outcomes and training datasets. Evidence from healthcare algorithms has shown that bias may be reproduced when models are built on unrepresentative data or problematic proxy measures (14). For surgical patients, this raises practical questions about whether an AI system has been evaluated in patients with similar characteristics, conditions or anatomical variation.

Regulatory and governance literature therefore emphasizes data quality, transparency, equity, privacy and ongoing monitoring in AI health applications (15). These issues may influence consent confidence because patients may want to know whether their data are used only for their care or also for audit, system improvement, model training or commercial development.

Although ethical, regulatory and technical literature increasingly discusses AI transparency and consent, less is known about how surgical patients themselves interpret AI involvement in planning and what they consider necessary for meaningful consent. A qualitative approach is needed because the phenomenon involves meanings, expectations, concerns and relational judgements that cannot be fully captured through technical validation studies or quantitative acceptance measures alone. Using a SPIDER-oriented framing, the sample of interest comprised surgical patients and patient-related stakeholders; the phenomenon of interest was trust and consent in AI-supported surgical planning; the design involved semi-structured qualitative inquiry; the evaluation focused on perceptions of disclosure, understanding, autonomy, privacy, fairness, accountability and surgeon responsibility; and the research type was qualitative. The objective of this study was therefore to explore patient perspectives on AI-supported surgical planning, with particular attention to how patients understand the boundary between surgeon judgement and AI advice, what they expect to be disclosed during consent, how they interpret explainability and data use, and what forms of clinical and institutional accountability are needed to support trust.

MATERIALS AND METHODS

This study used a qualitative descriptive design with an interpretative orientation to explore patient perspectives on trust and consent in AI-supported surgical planning. The design was selected because the aim was not to estimate the prevalence of acceptance or refusal, but to examine how patients understood AI involvement, what information they considered necessary for meaningful consent and how they interpreted surgeon responsibility, data use, fairness and accountability. Reflexive thematic analysis was used to identify patterned meanings across participant accounts while remaining attentive to context, language and variation in patient concerns (17,18). The reporting of the study was guided by the consolidated criteria for reporting qualitative research and the standards for reporting qualitative research, with attention to research team characteristics, reflexivity, sampling, recruitment, setting, data collection, analysis and trustworthiness (19,20).

The study was conducted in the context of surgical care at the Department of Surgery, Universitas Prima Indonesia, Indonesia, where AI-supported planning may be relevant to imaging review, anatomical modelling, risk estimation and preoperative decision support. Eligible participants were adults aged 18 years or older who were scheduled for surgery, had previously undergone surgery, were involved in surgical decision-making as a caregiver or family decision-maker, or had relevant experience of surgical consent discussions. Participants were included if they were able to provide informed consent and discuss their views on AI-supported surgical planning. Individuals were excluded if they were unable to provide consent, unable to participate in a qualitative interview, clinically unstable, or likely to experience distress during participation. Purposive sampling was used to obtain information-rich variation in surgical experience, digital confidence, care context and attitude toward AI-supported planning. The final sample included 12 participants, comprising adult surgical patients, participants with previous surgical experience, a caregiver, an elective surgery patient, a cancer surgery patient, a participant with limited digital confidence, an emergency surgery survivor, a participant from a minority background and a postoperative patient. This variation was intended to capture a range of consent concerns rather than to achieve statistical representativeness.

Potential participants were approached after initial eligibility screening and were informed that participation was voluntary and that refusal would not affect their care. Before data collection, participants received information about the purpose of the study, the voluntary nature of participation, confidentiality protections and their right to withdraw. Written informed consent was obtained from all participants before participation. Permission for audio-recording was sought where recording was used; when recording was not possible or not preferred, participant accounts were documented through detailed contemporaneous

notes. All participant identifiers were removed during transcription, note preparation and reporting, and quotations were labelled using participant codes to protect anonymity.

Data were collected through semi-structured participant interviews using a topic guide developed from the study objectives and relevant literature on AI ethics, surgical consent, explainability, accountability, privacy and trust. The guide explored participants' understanding of AI involvement in surgical planning, expectations for disclosure, views on the boundary between surgeon judgement and AI advice, preferences for plain-language explanation, concerns about data privacy and secondary data use, perceptions of bias and fairness, expectations about responsibility for planning errors and the role of the surgeon in maintaining trust. The guide was reviewed by the research team for clarity, relevance and sensitivity before use. Interviews were conducted in a private and respectful manner to allow participants to discuss concerns freely and without interruption. Field notes were recorded after interviews to document contextual observations, participant emphasis and early analytic reflections. Repeat interviews were not conducted.

The research team was based in the Department of Surgery, Universitas Prima Indonesia, and included investigators involved in conceptualization, literature review, drafting, critical revision and final approval of the manuscript. Because the topic involved medical authority, surgical vulnerability and emerging technology, reflexivity was treated as part of the analytic process. The researchers acknowledged that participants might feel reluctant to question surgical judgement, institutional practice or the use of emerging technology in care. To reduce this risk, interviews used open-ended prompts, avoided presenting AI as inherently beneficial and explicitly invited concerns, uncertainty and disagreement. Participants were reminded that there were no right or wrong answers and that their responses would not influence their care. Reflexive notes were used to document researcher assumptions, emerging interpretations and possible effects of clinical and academic background on data collection and analysis.

Interview data were prepared for analysis through transcription or detailed account preparation, depending on the form in which each account was recorded. Transcripts and notes were reviewed for accuracy and completeness before coding. All data were de-identified before analysis. Study files were stored securely and were accessible only to the research team. Any link between participant identity and participant code was kept separately from the analytic dataset. Data handling followed institutional ethical requirements and confidentiality safeguards.

Thematic analysis combined deductive and inductive procedures. Deductive attention was given to sensitising concepts from the literature, including disclosure, autonomy, accountability, privacy, fairness, explainability and trust. Inductive attention was given to participants' own meanings, repeated concerns, contrasts and unexpected interpretations. Analysis began with familiarization through repeated reading of transcripts, participant accounts and field notes. Initial codes were generated from meaningful segments of text and compared across participants. A preliminary coding framework was developed to organize recurrent codes while allowing new inductive codes to be added during analysis. Codes were reviewed and refined through analytic discussion within the research team. Disagreements in interpretation were resolved through consensus, with attention to the wording of participant accounts and the fit between codes, candidate themes and the overall dataset. Candidate themes were generated by grouping related codes, reviewed against coded extracts and the full dataset, refined through team discussion and named to reflect their central meaning.

Sample adequacy was assessed using the concept of information power rather than statistical saturation. The sample was considered sufficient because the study aim was specific, participants were purposively selected for relevance to surgical consent and AI-supported planning, and the accounts provided focused information on disclosure, understanding, accountability, privacy, fairness and trust (21). During analysis, the research team considered whether later accounts introduced substantially new consent-related concerns or mainly elaborated existing patterns. The study does not claim population-level saturation across all surgical specialties or AI applications; rather, it provides interpretative evidence from a targeted qualitative sample.

Trustworthiness was addressed through credibility, dependability, confirmability and transferability strategies. Credibility was supported through purposive variation in participant experience, use of semi-structured interviews, careful linkage of themes to representative quotations and comparison of convergent and contrasting accounts. Dependability was supported through documentation of recruitment, data collection, coding and theme development procedures. Confirmability was strengthened through reflexive note-taking, preservation of analytic decisions and discussion of candidate themes within the research team. Transferability was supported by describing participant characteristics, surgical context and the nature of AI-supported planning sufficiently for readers to judge relevance to other settings. Analytic triangulation was achieved by comparing perspectives across participant types, including patients, caregivers and participants with different levels of surgical or digital experience. Formal member checking was not conducted, and this was considered during interpretation of the findings.

Ethical approval was obtained from Universitas Prima Indonesia, Indonesia. The study followed principles of voluntary participation, informed consent, confidentiality, anonymity and non-maleficence, consistent with the Declaration of Helsinki for health-related research involving human participants (24). Participants were informed that they could decline to answer any question, pause the interview or withdraw without consequence. Because discussion of surgical care and consent could produce anxiety, the interviewer monitored discomfort and was prepared to stop the interview or refer participants to appropriate clinical staff if distress occurred. No identifying information is reported in the findings, and quotations are presented using anonymized participant codes.

FINDINGS

Twelve purposively selected participants contributed accounts that reflected variation in surgical experience, digital confidence, family involvement in decision-making and concern about AI-supported surgical planning. The sample included adult surgical patients, participants with previous or planned surgery, a caregiver or family decision-maker, a cancer surgery patient, an emergency surgery survivor, a participant with limited digital confidence, a participant from a minority background and a postoperative patient. Across the dataset, participants did not reject AI-supported surgical planning in principle. Their acceptance was conditional and depended on whether AI was disclosed clearly, whether the surgeon remained visibly responsible for the final plan, whether the explanation was understandable, and whether privacy, fairness and accountability concerns were addressed.

The analysis identified six interrelated themes: unclear boundaries between surgeon judgement and AI advice; consent without meaningful AI disclosure; explainability and patient understanding; data privacy, bias and fairness concerns; accountability for planning errors; and relational trust in the surgeon as the consent anchor. These themes showed that consent confidence was not shaped by the mere presence of AI, but by how participants understood the relationship between AI output, clinical judgement, data governance and institutional responsibility. Table 1 presents a qualitative theme matrix showing the presence and relative strength of each theme across participant-position groups, using the following categories: strong = explicitly and centrally represented in the available accounts; moderate = clearly present but less central; limited = present in a narrower or more specific participant account; and not explicit = not directly documented in the available manuscript data.

Table 1. Qualitative theme matrix across participant-position groups

Theme	Adult/current surgical patients	Previous/postoperative/ emergency surgery participants	Caregiver/family decision-maker	Limited digital confidence participant	Minority-background participant	Illustrative quote IDs
Unclear boundaries between surgeon judgement and AI advice	Strong	Moderate	Moderate	Moderate	Moderate	Q1

Theme	Adult/current surgical patients	Previous/postoperative/ emergency surgery participants	Caregiver/family decision-maker	Limited digital confidence participant	Minority-background participant	Illustrative quote IDs
Consent without meaningful AI disclosure	Strong	Moderate	Strong	Strong	Moderate	Q2
Explainability and patient understanding	Moderate	Moderate	Strong	Strong	Moderate	Q3
Data privacy, bias and fairness concerns	Moderate	Moderate	Moderate	Limited	Strong	Q4
Accountability for planning errors	Strong	Strong	Moderate	Moderate	Moderate	Q5
Relational trust in the surgeon as consent anchor	Strong	Strong	Moderate	Moderate	Moderate	Q6

The matrix indicates that concerns about surgeon responsibility and AI disclosure were strongly represented across several participant positions. Participants who were current or previous surgical patients emphasized the need to know whether AI merely supported the surgeon or materially influenced the plan. Family decision-makers were particularly concerned with whether AI-related information could be explained in terms that were understandable for shared family decision-making. The participant with limited digital confidence highlighted how technical vocabulary could discourage questioning, while the participant from a minority background made fairness and representativeness especially visible. These contrasts suggest that AI consent communication should not be standardized as a single technical disclosure, but should allow explanation to be adjusted to patient context, family involvement, digital confidence and perceived vulnerability.

Table 2. Quote table linking themes, subthemes and representative participant accounts

Theme	Subtheme	Representative quote	Participant label	Analytic interpretation
Unclear boundaries between surgeon judgement and AI advice	AI as support versus AI as decision-maker	"I would not object to the computer supporting, but it must be my surgeon deciding."	P01, adult surgical patient	This quote shows conditional acceptance of AI when it is positioned as an assistive tool rather than a replacement for professional judgement. The participant's trust depends on visible surgeon authority over the final plan.
Consent without meaningful AI disclosure	Form-based disclosure versus meaningful explanation	"My family does not know what AI on the form means."	P03, patient caregiver	The concern is not only whether AI is mentioned, but whether disclosure is understandable enough to support family decision-making. A written reference to AI may be insufficient without verbal explanation.
Explainability and patient understanding	Technical language as a barrier to questioning	"I feel like I should not question the word algorithm."	P09, participant with limited digital confidence	The quote illustrates how technical terminology can create silence rather than understanding. Patient-centred explainability requires language that invites questions rather than discouraging them.
Data privacy, bias and fairness concerns	Representativeness of AI training data	"How do I know that the data it was trained with includes people like me?"	P11, minority-background participant	Fairness is experienced as a personal safety issue rather than an abstract technical matter. Participants may connect trust in AI to whether the system has been tested on patients with similar characteristics.
Accountability for planning errors	Fear of blame-shifting to technology	"If there is an issue, I do not want them to blame the system and say no one is at fault."	P05, elective surgery patient	Accountability is part of consent confidence. Participants want to know before surgery who remains answerable if AI-supported planning contributes to harm.

Theme	Subtheme	Representative quote	Participant label	Analytic interpretation
Relational trust in the surgeon as consent anchor	Surgeon explanation as the basis of AI trust	"I did not trust it until the surgeon explained what the system showed and what he disagreed with."	P12, postoperative patient	Trust in AI is mediated through trust in the surgeon. The participant valued not only explanation of the AI output, but also evidence that the surgeon could critically assess and disagree with it.

The first theme concerned uncertainty about the boundary between surgeon judgement and AI advice. Participants distinguished between AI as an additional planning aid and AI as an apparent decision-maker. This distinction was central to consent confidence because participants were more comfortable when AI was described as supporting the surgeon's interpretation rather than determining the plan independently. The adult surgical patient represented by Q1 accepted AI in principle but only when the surgeon retained the final decision-making role. This pattern indicates that consent discussions should explicitly state whether AI was used for image review, risk estimation, anatomical modelling or planning support, and should clarify that the surgeon interprets the output and remains responsible for the final recommendation.

The second theme showed that minimal or form-based AI disclosure was not viewed as sufficient. Participants wanted disclosure to be verbal, plain-language and linked to their own surgical plan. The caregiver account in Q2 shows that a technical phrase on a consent form may not create meaningful understanding, especially where family members participate in decision-making. This theme was strongly connected to autonomy because participants did not necessarily require a detailed technical explanation, but they expected to be told whether AI had influenced the recommendation, what part of the plan it informed and whether they could ask questions before agreeing to surgery.

The third theme concerned explainability and patient understanding. Participants interpreted explainability in practical rather than mathematical terms. The participant with limited digital confidence in Q3 described feeling unable to question the word "algorithm," suggesting that technical vocabulary can create a power imbalance in consent conversations. Across the available accounts, participants appeared to need explanations that translated AI involvement into clinically meaningful terms: what the AI assessed, what it did not assess, how the surgeon used the output, what uncertainty remained and how the final decision was made. The findings therefore support a patient-level explanation model rather than a technical model-centred explanation.

The fourth theme related to data privacy, bias and fairness. Participants connected trust in AI-supported planning to how data were used, whether privacy was protected and whether AI systems were representative of people like them. The minority-background participant in Q4 framed fairness as a direct concern about personal safety and suitability of care. This theme indicates that consent materials should avoid vague reassurance about AI accuracy. Where relevant, they should distinguish between use of patient data for immediate care, audit, system improvement or model development, and should state whether safeguards exist to address privacy and fairness.

The fifth theme concerned accountability for planning errors. Participants did not accept the idea that responsibility should become unclear because AI was involved. The elective surgery patient in Q5 expressed concern that clinicians or institutions might blame the system if an error occurred. This concern demonstrates that accountability is not only a post-harm legal issue; it is part of preoperative consent confidence. Participants wanted reassurance that AI-supported planning would not weaken responsibility and that a clear clinical and institutional route for accountability would remain in place.

The sixth theme identified relational trust in the surgeon as the anchor of AI consent. Participants were more willing to accept AI when a trusted surgeon explained its role and demonstrated independent judgement. The postoperative participant in Q6 emphasized trust emerging after the surgeon explained both what the system showed and what the surgeon disagreed with. This finding suggests that the surgeon's role is not simply to transmit AI output to the patient, but to interpret, contextualize and, where necessary,

challenge it. AI communication should therefore be integrated into shared decision-making rather than presented as a separate technical disclosure.

Overall, the findings show that participants' concerns were not directed at AI use alone, but at hidden AI influence, unclear responsibility, inaccessible language, uncertain data use and the possibility of accountability being displaced onto technology. The strongest pattern across the available accounts was that AI-supported surgical planning became more acceptable when it was disclosed transparently, explained in plain language and visibly placed under surgeon and institutional responsibility. These findings support a proportionate consent model in which the depth of AI explanation increases when AI materially influences risk estimation, anatomical planning, procedure selection or the recommendation to proceed.

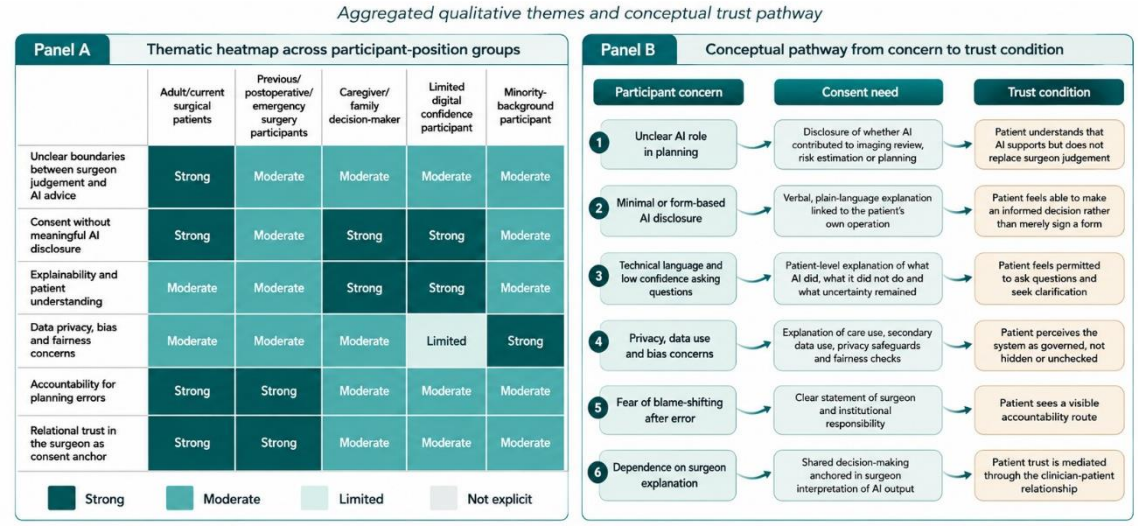

Figure 1. Panelled qualitative visualization of trust and consent in AI-supported surgical planning

Figure 1 shows that participants' trust in AI-supported surgical planning was shaped less by the presence of AI itself than by how AI was disclosed, explained and governed within the consent process. Panel A indicates that concerns about surgeon responsibility, meaningful disclosure and accountability were strongly represented across several participant-position groups, while privacy, bias and fairness were especially visible in the account of the minority-background participant. Panel B shows the implied consent pathway: participants moved from uncertainty about AI influence, technical language, data use and responsibility toward trust when AI was explained in plain language, placed under visible surgeon judgement and supported by institutional accountability. The figure should therefore be interpreted as a qualitative synthesis of thematic patterns, not as a statistical comparison between participant groups.

Panel A. Thematic heatmap: themes across participant-position groups

Theme	Adult/current surgical patients	Previous/postoperative/ emergency surgery participants	Caregiver/family decision-maker	Limited digital confidence participant	Minority-background participant
Unclear boundaries between surgeon judgement and AI advice	Strong	Moderate	Moderate	Moderate	Moderate
Consent without meaningful AI disclosure	Strong	Moderate	Strong	Strong	Moderate
Explainability and patient understanding	Moderate	Moderate	Strong	Strong	Moderate
Data privacy, bias and fairness concerns	Moderate	Moderate	Moderate	Limited	Strong
Accountability for planning errors	Strong	Strong	Moderate	Moderate	Moderate
Relational trust in the surgeon as consent anchor	Strong	Strong	Moderate	Moderate	Moderate

Panel B. Conceptual alluvial pathway: concern → consent need → trust condition

Participant concern	Consent need	Trust condition
Unclear AI role in planning	Disclosure of whether AI contributed to imaging review, risk estimation or planning	Patient understands that AI supports but does not replace surgeon judgement
Minimal or form-based AI disclosure	Verbal, plain-language explanation linked to the patient's own operation	Patient feels able to make an informed decision rather than merely sign a form
Technical language and low confidence asking questions	Patient-level explanation of what AI did, what it did not do and what uncertainty remained	Patient feels permitted to ask questions and seek clarification
Privacy, data use and bias concerns	Explanation of care use, secondary data use, privacy safeguards and fairness checks	Patient perceives the system as governed, not hidden or unchecked
Fear of blame-shifting after error	Clear statement of surgeon and institutional responsibility	Patient sees a visible accountability route
Dependence on surgeon explanation	Shared decision-making anchored in surgeon interpretation of AI output	Patient trust is mediated through the clinician-patient relationship

DISCUSSION

This qualitative study explored how participants understood trust and consent in the context of AI-supported surgical planning. The findings show that participants were not opposed to AI-assisted planning in principle, but their acceptance was conditional on disclosure, plain-language explanation, visible surgeon responsibility, data transparency, fairness safeguards and clear accountability. Participants distinguished between AI as a bounded clinical support tool and AI as an invisible or poorly explained authority within the planning process. This distinction is important because consent confidence was shaped less by whether AI was used and more by whether participants understood how it contributed to the plan, how the surgeon interpreted its output and who remained responsible for the final recommendation.

A central finding was the perceived boundary between surgeon judgement and AI advice. Participants were more comfortable when AI was presented as an aid to image interpretation, anatomical modelling or risk estimation, but they were less comfortable when its influence appeared hidden or when responsibility seemed to shift away from the surgeon. This finding supports emerging evidence that patients want to be informed when AI contributes to healthcare decisions and that disclosure should be related to the clinical significance of AI involvement rather than presented as a generic technical statement (25). In surgical planning, this means that a proportionate disclosure model is needed. Minor background AI functions may require brief explanation, whereas AI outputs that materially influence risk estimation, operative approach, implant planning, route selection or the recommendation to proceed should be discussed directly during consent.

The findings also show that meaningful disclosure requires more than the presence of the term "AI" on a consent form. Participants wanted verbal explanation in language that connected AI use to their own operation. This supports the distinction between technical explainability and patient-centred explainability. Technical explanations may focus on model architecture, variables or validation, but patients usually need to understand what the tool was used for, what it did not do, what uncertainty remained and how the surgeon used or challenged the output. This interpretation is consistent with prior work warning that explainability in healthcare AI can create false reassurance if it does not produce real clinical understanding (12,13). The implication is that consent conversations should avoid both extremes: they should not overwhelm patients with mathematical detail, but they should also not reduce AI disclosure to a vague reassurance that the system is "safe" or "approved."

Accountability emerged as another major condition for trust. Participants were concerned that if AI-supported planning contributed to harm, responsibility might be displaced onto the system. This concern aligns with broader discussions of accountability and safety in healthcare AI, where responsibility may be distributed across clinicians, institutions, developers and regulators (27). In the surgical consent context, accountability should therefore be addressed before harm occurs, not only after an adverse event. Patients should be told that AI output is reviewed by the surgeon, that the surgeon remains responsible for the final clinical recommendation and that the institution has procedures for validation, monitoring, documentation and incident review. Such communication may reduce fear that AI creates an accountability gap.

Privacy, data use and fairness were also closely linked to trust. Participants did not treat data governance as a distant technical issue; they connected it to whether AI-supported planning could be trusted for patients like themselves. Concerns about representativeness are consistent with wider evidence that algorithmic systems can reproduce or worsen health disparities when data sources, proxy measures or implementation pathways are poorly governed (14,28). Consent materials should therefore separate immediate clinical use of data from secondary uses such as audit, quality improvement, model refinement, research or commercial development. Where relevant, patients should also be informed that AI tools require ongoing evaluation across different patient groups and that fairness concerns are part of institutional governance, not merely developer responsibility.

Relational trust in the surgeon was the strongest anchor of AI acceptability. Participants were more willing to trust AI-supported planning when the surgeon explained what the system showed, how it informed the plan and where the surgeon's own judgement remained central. This finding is consistent with evidence that patients may feel apprehensive about healthcare AI when they perceive reduced human contact, limited transparency or weakened professional responsibility (29). AI communication should therefore be integrated into shared decision-making rather than separated as a technical appendix to consent. The surgeon's role is not to transfer AI output passively to the patient, but to interpret, contextualize and, when necessary, challenge it.

The study has practical implications for implementation. First, institutions using AI-supported surgical planning should develop patient-facing consent scripts that explain AI involvement in plain language. These scripts should clarify whether AI contributed to imaging review, anatomical modelling, risk estimation or planning options. Second, surgeons and surgical teams should receive communication training on how to explain AI without overstating certainty or shifting responsibility. Third, consent forms should include a short AI disclosure section when AI use is material to the plan, but this should supplement rather than replace verbal discussion. Fourth, hospitals should establish documentation practices that record when AI-supported outputs were used, how they were interpreted and whether the final plan differed from the AI suggestion. Fifth, AI governance processes should include validation, monitoring, incident-response pathways and fairness review, so that patient-facing communication is supported by institutional accountability rather than individual reassurance alone.

The findings support a three-tiered practical consent model for AI-supported surgical planning. The first tier is disclosure: patients should be told whether AI contributed materially to the surgical plan. The second tier is explanation: patients should receive a plain-language account of what the AI did, what it did not do and how the surgeon used the output. The third tier is accountability: patients should be reassured that the surgeon and institution remain responsible for clinical decision-making, safety monitoring and response to error. This model does not require a separate consent process for every minor algorithmic function, but it does require explicit discussion when AI involvement is material to the patient's decision.

This study should be interpreted within its qualitative limitations. The sample was purposive and small, which is appropriate for exploratory qualitative insight but does not estimate the prevalence of each concern among surgical patients. Participants' accounts may have been influenced by social desirability, especially if recruitment occurred in a clinical environment where patients perceived the surgical team or institution as authoritative. If participants were recruited through surgical pathways, they may also have felt reluctant to express distrust of clinical technology. The study may be affected by limited specialty diversity, variation in participants' prior understanding of AI and possible differences between hypothetical and actual experience of AI-supported planning. If interviews involved translation, some conceptual nuance may have been lost. Member checking was not reported, and the absence of transcript return or participant validation may limit credibility. These limitations do not invalidate the findings, but they mean the themes should be treated as interpretative evidence for improving consent communication rather than as population-level conclusions.

Future research should test patient-facing AI consent language across surgical specialties, compare patients with actual and hypothetical exposure to AI-supported planning and examine whether structured

disclosure improves understanding, decisional confidence and perceived accountability. Further work should also include surgeons, nurses, hospital administrators, AI developers and ethics/governance stakeholders to understand how patient expectations can be operationalized in real surgical workflows. Importantly, future studies should avoid assuming that AI disclosure alone is sufficient. The present findings suggest that trust depends on the quality of explanation, the visibility of human judgement and the credibility of institutional accountability.

CONCLUSION

AI-supported surgical planning may strengthen clinical preparation, but it creates consent challenges when patients are not clearly informed about its role. In this qualitative study, participants did not reject AI-assisted planning in principle; rather, they accepted it conditionally when it was explained as a bounded support tool, placed under visible surgeon judgement and supported by institutional accountability. The main mechanisms shaping perceived acceptability were transparent disclosure, plain-language explanation, reassurance that AI would not replace professional judgement, clarity about data use and fairness, and a visible route for responsibility if errors occurred. These findings support a proportionate consent approach in which patients are told when AI materially contributes to imaging review, risk estimation, anatomical modelling or operative planning, and are given the opportunity to ask questions before agreeing to surgery. Institutions implementing AI-supported surgical planning should develop consent scripts, clinician communication training, documentation procedures and governance pathways that make AI use understandable and accountable. The findings should not be interpreted as evidence that AI improves surgical outcomes or increases uptake of surgery; rather, they show how patients may define the conditions under which AI-supported planning becomes acceptable and trustworthy.

REFERENCES

1. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg. 2018;268(1):70-6. doi:10.1097/SLA.0000000000002693.

2. Maier-Hein L, Eisenmann M, Sarikaya D, März K, Collins T, Malpani A, et al. Surgical data science: enabling next-generation surgery. Nat Biomed Eng. 2022. [Add volume, issue, page range and DOI if available.]

3. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31-8. doi:10.1038/s41591-021-01614-0.

4. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021. Available from: WHO website. Accessed 2026 Jun 14.

5. Cohen IG, Slottje P. Artificial intelligence and informed consent. In: Artificial Intelligence in Health Care. Bethesda: National Center for Biotechnology Information; 2024. Available from: NCBI Bookshelf. Accessed 2026 Jun 14.

6. Pressman SM, Borna S, Gomez-Cabello CA, Haider SA, Haider CR, Forte AJ. Artificial intelligence and ethics in surgery. Surg Open Sci. 2024;18:51-8. doi:10.1016/j.sopen.2024.01.009.

7. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195. doi:10.1186/s12916-019-1426-2.

8. Gille F, Jobin A, Ienca M. What we talk about when we talk about trust: theory of trust for AI in healthcare. Intell Based Med. 2020;1-2:100001. doi:10.1016/j.ibmed.2020.100001.

9. Gerke S, Minssen T, Cohen IG. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artificial Intelligence in Healthcare. London: Academic Press; 2020. doi:10.1016/B978-0-12-818438-7.00012-5.

10. European Commission. Regulation laying down harmonised rules on artificial intelligence: Artificial Intelligence Act. Brussels: European Commission; 2024. Available from: European Commission/AI Act website. Accessed 2026 Jun 14.

11. U.S. Food and Drug Administration. Artificial intelligence and machine learning in software as a medical device. Silver Spring: U.S. Food and Drug Administration; 2025. Available from: FDA website. Accessed 2026 Jun 14.

12. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20:310. doi:10.1186/s12911-020-01332-6.

13. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-50. doi:10.1016/S2589-7500(21)00208-9.

14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342.

15. World Health Organization. Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Available from: WHO website. Accessed 2026 Jun 14.

16. Pruski M. AI-enhanced healthcare: not a new paradigm for informed consent. J Med Ethics. 2024. [Add volume, issue, page/article number and DOI if available.]

17. Braun V, Clarke V. One size fits all? What counts as quality practice in reflexive thematic analysis? Qual Res Psychol. 2021;18(3):328-52. doi:10.1080/14780887.2020.1769238.

18. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach. 2020;42(8):846-54. doi:10.1080/0142159X.2020.1755030.

19. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research. Int J Qual Health Care. 2007;19(6):349-57. doi:10.1093/intqhc/mzm042.

20. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research. Acad Med. 2014;89(9):1245-51. doi:10.1097/ACM.0000000000000388.

21. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res. 2016;26(13):1753-60. doi:10.1177/1049732315617444.

22. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs. 2016;72(12):2954-65. doi:10.1111/jan.13031.

23. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research. Fam Med Community Health. 2019;7(2):e000057. doi:10.1136/fmch-2018-000057.

24. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human participants. Ferney-Voltaire: World Medical Association; 2024. Available from: World Medical Association website. Accessed 2026 Jun 14.

25. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Patient perspectives on informed consent for medical AI: a web-based experiment. J Med Internet Res. 2024. [Add volume, issue, article number and DOI if available.]

26. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework. Gaithersburg: National Institute of Standards and Technology; 2023. Available from: NIST website. Accessed 2026 Jun 14.

27. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ. 2020;98(4):251-6. doi:10.2471/BLT.19.237486.

28. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. Nat Med. 2020;26:16-7. doi:10.1038/s41591-019-0649-2.

29. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med. 2021;4:140. doi:10.1038/s41746-021-00509-1.

30. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med. 2018;15(11):e1002689. doi:10.1371/journal.pmed.1002689.