C

Public 510(k) Clarity Audit - K252665

brAIn™ Shoulder Positioning (Cloud-Based AI/ML SaMD for Preoperative Surgical Planning)

Applicant: Avatar Medical

FDA Decision Date : October 20, 2025.

Audit Date : May 11, 2026.

Audit #1 of the Public Clarity Audit series

Source document : FDA 510(k) database, K252665. Available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K252665

Independent review notice. This audit relies exclusively on the public 510(k) Summary documents released by the FDA for K252665. Martinez Clarity has no affiliation with Avatar Medical. This audit is published for educational and methodology demonstration purposes only. The audited submission was successfully cleared by the FDA in October 2025. The findings below represent observations of documentary clarity. They are not regulatory deficiencies, and they do not call into question the cleared status of the device.

Martinez Clarity | martinezclarity.com

c

Executive Summary

This audit applies the Martinez Clarity 4-Axis methodology to the public 510(k) Summary of brAIn™ Shoulder Positioning (K252665), a cloud-based AI/ML software intended to assist preoperative surgical planning for primary total shoulder replacement, manufactured by Avatar Medical and cleared by the FDA on October 20, 2025.

10 findings were identified:

- 4 Red: Likely to trigger a Refuse to Accept (RTA) hold or an Additional Information (AI) request letter under current 2026 review conditions.
- 6 Yellow: Likely to slow review under current reviewer bandwidth constraints.
- 0 Green: Cosmetic, no impact on review.

This submission was cleared by the FDA in October 2025. Under the review conditions of that period, none of these gaps were blocking. Today, however, with the tightening of acceptance criteria, the workforce reduction at CDRH, the mandatory SBOM requirement effective June 2025, and the stricter application of the September 2023 cybersecurity guidance, similar documentary gaps systematically trigger AI request letters on submissions currently under review. The typical added delay is 6 to 10 weeks per AI request.

The most striking pattern in this submission is a cross-reference inconsistency on the K-number itself : the Indications for Use document references submission number **K243292** while the cleared submission is **K252665**. This is the type of finding that the FDA RTA Acceptance Checklist (Appendix A) explicitly verifies at the 15-day acceptance review. Under 2026 conditions, this gap alone would have been flagged before substantive review began.

Axis	**Findings**	**Highest severity**
1. Terminology Consistency	0	n/a

C

Axis	Findings	Highest severity
2. Requirements to Tests Traceability	2	Red
3. RTA Acceptance Checklist Alignment	1	Red
4. Device-Specific Compliance	7	Red

c

Methodology

This audit follows the Martinez Clarity 4-Axis methodology.

Axis 1. Terminology Consistency. Verification of device names, predicate names, components, and technical terms across all sections. Drift signals either a rushed submission or a multi-author draft without a final harmonization pass.

Axis 2. Requirements to Tests Traceability. Verification that every performance claim and risk statement is anchored to a named test protocol, a recognized standard, and a quantified result. Unanchored claims invite reviewer clarification requests.

Axis 3. RTA Acceptance Checklist Alignment. Verification of the public-visible elements of the FDA Refuse to Accept Policy for 510(k)s: complete predicate identification, alignment of indications for use, and the rigor of the substantial equivalence conclusion.

Axis 4. Device-Specific Compliance (Cloud-Based AI/ML SaMD, 2026 context): For software with AI/ML components : Documentation Level per FDA June 2023 guidance, IEC 62304 software safety class, training/validation/test data characterization, demographic diversity, multi-site validation, Predetermined Change Control Plan (PCCP), cybersecurity per FDA September 2023 Premarket Guidance, SBOM (mandatory since June 2025), ISO 14971 risk management, IEC 62366 usability.

Severity scale :

- Red: High likelihood of RTA hold or immediate AI request under current FDA review conditions.
- Yellow: Likely to slow review, generate secondary clarification questions.
- Green: Cosmetic gap, no impact on the decision.

c

Findings

AXIS 3 - RTA ACCEPTANCE CHECKLIST ALIGNMENT

Finding 1, K-number cross-reference inconsistency between IFU and submission [Red]

Axis : RTA Acceptance Checklist Alignment Page reference : Indications for Use document, page 9 of 34

Cited text :

"Please type in the marketing application/submission number, if it is known. This textbox will be left blank for original applications/submissions: K243292"

Risk analysis. The Indications for Use document declares the marketing/submission number as K243292, while the cleared submission referenced in the FDA clearance letter is K252665. This is a cross-reference inconsistency at the most basic acceptance level. The FDA Refuse to Accept Policy for 510(k)s explicitly verifies cover letter, IFU, and 510(k) Summary alignment at the 15-day acceptance review before substantive review begins.

Under the current 2026 review climate, with CDRH bandwidth reduced and automated cross-reference checks more strictly applied, RTA holds for submission-number inconsistencies are issued within the 15-day window. The typical impact is a minimum 30-day delay for re-preparation, plus potential reset of timeline expectations and reviewer first-impression credibility.

Recommended fix. Replace K243292 with K252665 in the Indications for Use form's "marketing application/submission number" field. Verify consistency across cover letter, IFU document, Form 3514, and all 510(k) Summary trade name references before any future amendment or related submission.

C

C

AXIS 2 — REQUIREMENTS TO TESTS TRACEABILITY

Finding 2 — Measurement Accuracy claim declared without quantified results [Red]

Axis : Requirements to Tests Traceability Page reference : 510(k) Summary, Performance Data section, pages 7 to 8

Cited text :

"All performance tests for Measurement Accuracy Validation were successfully completed with no deviations, confirming compliance with the required performance standards. The acceptance criterion is set at 1° for angle measurement, 1 mm for distance measurement and 1% for 3D subluxation."

Risk analysis. The submission declares acceptance criteria (**1°** angle, **1 mm** distance, **1%** 3D subluxation) but does not report quantified results from the validation runs. The phrasing "successfully completed with no deviations" lacks numerical evidence: no mean error, no standard deviation, no sample size, no statistical confidence intervals.

For an AI/ML SaMD assisting preoperative shoulder surgical planning, measurement accuracy is the core performance claim. Under the current 2026 review climate, FDA reviewers systematically convert such unanchored claims into AI requests asking for complete performance summary tables with statistical analysis. The typical impact is 6 to 10 weeks per AI request, plus potential need for additional validation data if methodology gaps are identified during the response cycle.

Recommended fix. Restructure the Measurement Accuracy section with quantified performance reporting. Example structure: "Measurement Accuracy validation (n=[sample size] test cases) : angle error mean = [X]° ± [Y]°, distance error mean = [A] mm ± [B] mm, 3D subluxation error mean = [C]% ± [D]%, all values meeting acceptance criteria of 1° / 1 mm / 1%. Statistical analysis per [referenced methodology]."

C

Finding 3 - Streaming Stability claim without quantified metrics [Yellow]

Axis 2: Requirements to Tests Traceability Page reference : 510(k) Summary, Performance Data section, page 8

Cited text :

"Validation demonstrated that the system can handle simultaneous usage by multiple users without performance degradation regarding the following metrics : frames per seconds, jitter and packet loss."

Risk analysis. The submission identifies the relevant performance metrics (frames per second, jitter, packet loss) but provides no quantified test results : no test conditions (number of concurrent users tested, network bandwidth, geographic distribution), no acceptance thresholds, no measured values.

For a cloud-based AI/ML device intended for surgical planning, streaming stability under load is a critical operational claim. While less directly safety-relevant than measurement accuracy, the absence of quantification under current 2026 reviewer bandwidth conditions is likely to generate an AI request asking for the complete streaming performance protocol and results.

Recommended fix. Provide quantified streaming stability validation : "Streaming stability tested with n=[X] concurrent users on [Y] Mbps network conditions. Results : frames per second = [Z] (acceptance : ≥[W]), jitter = [A] ms (acceptance : ≤[B]), packet loss = [C]% (acceptance : ≤[D]%). All metrics within acceptance criteria."

C

AXIS 4 — DEVICE-SPECIFIC COMPLIANCE (2026 CONTEXT)

Finding 4 — Cybersecurity not substantively addressed for cloud-based device [Red]

Axis 4: Device-Specific Compliance Page reference : 510(k) Summary, page 3 (cloud-based declaration) and page 6 (guidance reference)

Cited text :

"The brAIn™ Shoulder Positioning software is a cloud-based application intended for shoulder surgeons... the software validation activities were performed in accordance with... 'Content of Premarket Submission for Management of Cybersecurity in Medical Devices.'"

Risk analysis. The device is explicitly declared as cloud-based, which makes the FDA Cybersecurity Premarket Guidance (September 2023) fully applicable. The Summary references the guidance by name in section 2.8 but provides no **substantive cybersecurity content**: no threat modeling framework (STRIDE, PASTA), no authentication mechanism described, no encryption standards (TLS, AES), no audit logging procedures, no patch management plan, no vulnerability disclosure process, no SBOM reference.

Under the current 2026 review climate, reviewers systematically issue RTA holds or extended AI requests for cloud-based software submissions with insufficient cybersecurity documentation. The typical impact is 8 to 12 weeks for cybersecurity remediation requests alone, plus reputational signaling that the manufacturer's cybersecurity maturity may be insufficient.

Recommended fix. Add a dedicated Cybersecurity section in the Summary covering : (1) threat modeling per chosen framework, (2) authentication mechanisms (OAuth 2.0, MFA), (3) encryption (TLS 1.3 in-transit, AES-256 at-rest), (4) audit logging per defined standard, (5) patch management plan, (6) vulnerability disclosure process, (7) SBOM reference in standardized format. Cross-reference detailed cybersecurity documentation in submission appendix.

C

Finding 5 — SBOM not referenced despite post-June 2025 submission [Red]

Axis : Device-Specific Compliance Page reference : 510(k) Summary, entire document (absence finding)

Cited text. The 510(k) Summary contains no reference to Software Bill of Materials (**SBOM**) in any format.

Risk analysis. The Software Bill of Materials (SBOM) became mandatory for all 510(k) submissions containing software effective June 2025. This submission was received by FDA on August 25, 2025, two months after the SBOM mandate became effective. The Summary contains no reference to SBOM in any standardized format (SPDX, CycloneDX) or as an appendix reference.

For a cloud-based AI/ML device with multiple software dependencies (ML frameworks, cloud SDK, DICOM libraries, web interface frameworks), absence of SBOM is an acceptance check trigger under 2026 review context. The typical impact is 4 to 6 weeks for SBOM remediation requests alone.

Recommended fix. Add an SBOM reference in the Performance Data section or as Appendix: "Software Bill of Materials provided in [SPDX/CycloneDX] format in Appendix [X]. Components listed with version, license, and known vulnerability status. Update mechanism described for post-deployment SBOM maintenance."

Finding 6 — Shoulder Side Detection ground truth methodology inadequately documented [Yellow]

Axis : Device-Specific Compliance (AI/ML methodology) Page reference : 510(k) Summary, Performance Data section, page 7

C

Cited text :

"Validation compared the correct shoulder side (right or left) in DICOM images to a manual assessment performed by a Clinical Solutions Specialist."

Risk analysis. The ground truth for Shoulder Side Detection validation was established by a single "Clinical Solutions Specialist" whose qualifications and independence are not characterized. The submission does not describe the specialist's credentials: (board-certified radiologist, orthopedic surgeon, AI/ML labeling expert), does not establish inter-rater reliability against a second expert, and does not indicate blinding procedures during the validation assessment.

For an AI/ML device intended to support clinical decisions, FDA Good Machine Learning Practice (GMLP, October 2021) expects multi-rater ground truth with documented reliability metrics. Under the current 2026 review climate, reviewers increasingly request these methodology details for AI/ML SaMD submissions.

Recommended fix. Document the ground truth methodology : "Shoulder side ground truth established by [N] independent board-certified [specialty] readers, blinded to algorithm output. Inter-rater reliability : Cohen's kappa = [X] (95% CI : [Y] to [Z]). Cases with disagreement resolved by [adjudication procedure]."

Finding 7 — Ruler Performance Testing limited to synthetic inputs [Yellow]

Axis : Device-Specific Compliance (AI/ML methodology) Page reference : 510(k) Summary, Performance Data section, page 8

Cited text :

C

"Validation assessed the measurement precision of the ruler tool by comparing its accuracy between anatomical landmarks to theoretical distances calculated from their spatial coordinates."

Risk analysis. The Ruler Performance Testing methodology compares the tool's measurements to "theoretical distances calculated from spatial coordinates", meaning the inputs are synthetic ideal coordinates rather than clinically realistic measurements. This validates the mathematical correctness of the ruler algorithm but does not address performance under real-world conditions : CT acquisition artifacts, anatomical variations (osteoarthritis, prior surgery), or boundary cases of the segmentation output.

For a device intended for clinical use, validation on synthetic inputs alone is methodologically incomplete. Under the current 2026 review climate, this gap is likely to generate AI requests asking for clinical-scenario robustness data.

Recommended fix. Supplement synthetic validation with clinical scenario testing : "Ruler tool also validated on n=[X] representative clinical cases including pathological anatomies (osteoarthritis, post-traumatic deformity, prior surgical intervention). Measurement error in clinical scenarios : mean = [Y] mm ± [Z] mm, with documentation of edge-case behavior."

Finding 8 — Training and testing data demographic imbalance and arithmetic inconsistency [Yellow]

Axis : Device-Specific Compliance (AI/ML, GMLP) Page reference : 510(k) Summary, Performance Data, Segmentation Performance Testing, page 7

Cited text :

"Testing Set... Geographical Origin: 58.2% Europe (46), 41.8% USA (33) for left ; 56.4% Europe (53), 43.6% USA (41) for right"

C

Risk analysis. Two related issues are bundled in this finding.

First, demographic imbalance toward the European population. The testing set is 56 to 58% European and 42 to 44% US origin. For a device intended for commercialization in the US market, FDA Good Machine Learning Practice (GMLP, October 2021) explicitly recommends that training and testing populations reflect the intended use demographic distribution. Disbalance toward a non-US population without documented subgroup analysis raises generalization concerns.

Second, arithmetic inconsistency in the same demographic section. The Summary reports for the training set on the left side : 46.9% Europe (n=68) and 56.4% USA (n=77), which sums to 103.3% rather than 100%. This is either a typographical error or imprecision in the demographic accounting itself.

Both observations together are likely to trigger an AI request for clarification and potentially additional subgroup performance analysis. Under the current 2026 review climate, arithmetic errors in performance data are read as a signal of imprecise data governance.

Recommended fix. Correct the arithmetic inconsistency in training set demographics. Provide subgroup performance analysis by geographic origin demonstrating non-inferior performance in the US population versus European population. If subgroup analysis reveals disparity, augment training data or document the limitation as a labeling consideration.

Finding 9 — IEC 62304 software safety class not declared [Yellow]

Axis : Device-Specific Compliance Page reference : 510(k) Summary, Performance Data, Software Verification and Validation Tests, page 6

Cited text :

C

"the software validation activities were performed in accordance with ANSI AAMI IEC 62304:2006/A1:2016 — Medical device software — Software life cycle processes"

Risk analysis. The submission references IEC 62304 as the software lifecycle standard but does not declare the software safety class (Class A: no injury possible; Class B: non-serious injury possible; Class C: serious injury or death possible). For a Class II SaMD with AI/ML components that supports surgical planning decisions, Class B would be the minimum expected designation, with potential Class C consideration depending on the risk analysis outcome.

Absence of explicit class declaration prevents the reviewer from validating that the documentation depth and verification activities match the appropriate class requirements. Under current 2026 review, this gap typically generates an AI request asking the manufacturer to declare the class and confirm conformance to corresponding documentation requirements.

Recommended fix. Declare explicitly : "Software developed in conformance with IEC 62304:2006/A1:2016. Software Safety Class: [B or C] based on hazard analysis per ISO 14971:2019. Documentation level and verification activities adapted to the declared class per IEC 62304 sections 5.5 through 5.8."

Finding 10 — Predetermined Change Control Plan (**PCCP**) not mentioned for AI/ML device [Yellow]

Axis : Device-Specific Compliance (AI/ML lifecycle) Page reference : 510(k) Summary, entire document (absence finding)

Cited text :

The 510(k) Summary contains no reference to a Predetermined Change Control Plan (PCCP) or post-deployment performance monitoring procedure.

C

Risk analysis. The device uses machine learning for automatic segmentation. Summary page 3 states : "automatically segments using machine learning". The FDA Predetermined Change Control Plan (PCCP) framework, finalized in September 2024, allows manufacturers of AI/ML devices to pre-specify the types of modifications they may make to their model without filing new 510(k) submissions.

While PCCP is not strictly mandatory, its absence under current 2026 review climate increasingly signals to reviewers that the manufacturer has not planned for model retraining, drift monitoring, or performance recalibration over the device lifecycle. This may generate an AI request asking how the manufacturer intends to manage model evolution, or may result in the submission being constrained to a frozen-model configuration with limited update flexibility.

Recommended fix. Add a PCCP section in the Performance Data area : "Predetermined Change Control Plan : model retraining triggers ([criteria]), retraining data sources, validation requirements per retraining cycle, post-deployment performance monitoring procedure, notification thresholds for performance drift. PCCP scope : [defined types of model changes that fall under PCCP versus changes requiring new 510(k)]."

c

Conclusion

The audited submission (K252665, brAIn™ Shoulder Positioning) was successfully cleared by the FDA on October 20, 2025.

The documentary analysis identified 10 clarity gaps distributed across the 4 axes:

4 Red findings.

- K-number cross-reference inconsistency between the Indications for Use document (K243292) and the cleared submission (K252665).
- Measurement Accuracy validation claim declared without quantified results, statistical analysis, or sample sizes.
- Cybersecurity guidance referenced by name but not substantively addressed for a cloud-based device.
- Software Bill of Materials (SBOM) not referenced despite submission filed two months after the June 2025 mandate.

6 Yellow findings.

- Streaming stability metrics named but not quantified.
- Shoulder Side Detection ground truth established by a single specialist with undocumented credentials.
- Ruler Performance Testing validated only on synthetic inputs, no clinical scenario robustness.
- Demographic imbalance toward European population, plus arithmetic inconsistency in training set demographics.
- IEC 62304 software safety class not declared.
- Predetermined Change Control Plan (PCCP) not mentioned for an AI/ML device.

C

These gaps did not block clearance under 2025 review conditions. Under the current 2026 review climate, with CDRH workforce reduced, the Q-Sub program curtailed, the stricter application of the September 2023 cybersecurity guidance, and the mandatory SBOM requirement effective June 2025, three categories of gaps systematically increase review timeline risk.

First, the K-number cross-reference inconsistency between the IFU and the cleared submission is the type of finding that the RTA Acceptance Checklist verifies at the 15-day acceptance review. Under 2026 conditions, this gap alone would have been flagged before substantive review began.

Second, the cybersecurity documentation gap is significant for a cloud-based device. The FDA Cybersecurity Premarket Guidance (September 2023) is referenced but not substantively addressed, and the SBOM mandate effective June 2025 is unaddressed in a submission filed two months after the mandate became effective.

Third, the AI/ML methodology specifications, including ground truth, demographic representation, IEC 62304 safety class, and PCCP, are partial. Individually each is a Yellow severity finding. Cumulatively they signal that the AI/ML lifecycle documentation does not meet the 2026 expectation level for SaMD submissions.

For new submissions in 2026, these documentary gaps are addressable in 5 to 10 days of preparation work. Missing them generates AI request cycles of 6 to 10 weeks each, which is the gap that a Martinez Clarity audit is designed to close upstream of submission.

c

Apply for a Free Clarity Audit

Martinez Clarity runs independent clarity audits on 510(k) drafts following the 4-Axis methodology. We deliver a complete documentary risk map within 5 business days, ahead of submission, to protect your time-to-market schedule.

To apply for a free clarity audit, or to review our methodology and our two other public audits (BioButton System K241101, Catalyft LS Expandable Interbody System K241992), visit martinezclarity.com.

Currently offering 3 free audits to QA/RA leads with a 510(k) submission planned in the next 90 days.

Martinez Clarity | martinezclarity.com | remy@martinezclarity.com