Post-doctoral position

- **Title:** Learning to quantify registration uncertainty for safer augmented reality surgical navigation.
- **Research axis of the 3IA: Axis 2**, AI for healthcare.
- **Keywords:** machine learning for surgical guidance, uncertainty quantification, conformal prediction, augmented reality, image registration, computer-assisted surgery.
- **Supervisor: Marc-Olivier Gauci**, MD, PhD, HDR, orthopaedic surgeon, Institut Universitaire Locomoteur et du Sport (iULS), CHU de Nice, and ICARE, Inserm U1091, Université Côte d'Azur, Marc-Olivier.GAUCI@univ-cotedazur.fr
- **Co-supervisor: Marco Winckler**, Professor of Computer Science and Human-Computer Interaction, Université Côte d'Azur, CNRS, I3S, SPARKS team, Marco.Winckler@univ-cotedazur.fr
- **The laboratories:**
  - ICARE, Inserm U1091, Institut de Biologie Valrose, Université Côte d'Azur, with the surgical platform of the iULS, CHU de Nice.
  - I3S, Université Côte d'Azur and CNRS, SPARKS team, Sophia Antipolis.
- **Duration:** 24 months. **Starting date:** to be agreed with the selected candidate.

Apply by sending an email directly to the supervisor, including:

- Curriculum vitae with publication list.
- Motivation letter.
- Academic transcripts of the master's degree and evidence of the doctoral degree.
- A letter of recommendation.

1 Context and objective

Artificial intelligence (AI) has entered the operating room through machine learning models for image segmentation, surgical planning and intraoperative scene analysis. These models do not assess themselves. A model that guides a surgical gesture produces an output of unknown local reliability, and the clinician has no way of telling whether the present case is one the model handles well. This project addresses that question in computer-assisted orthopaedic and craniomaxillofacial surgery, where the output of the system is a geometric alignment between a preoperative plan and the patient, and where the error of that alignment is measurable in millimetres and directly connected to the surgical outcome.

Three-dimensional preoperative planning is now routine in these procedures, and patient-specific guides derived from it improve implant positioning compared with freehand technique [1,2]. The transfer of the plan into the operating room remains unresolved. In shoulder arthroplasty the deviation between the planned and the achieved position of the glenoid component stays of the order of a few millimetres even when dedicated instrumentation is used [3], component malposition is associated with periprosthetic osteolysis and with a higher risk of loosening [4,5], and in mandibular reconstruction a comparable deviation on a cutting guide changes the occlusion of the patient. Augmented reality (AR) addresses this transfer by displaying the plan in the surgeon's field of view, and early clinical work in shoulder arthroplasty shows that a holographic overlay can guide glenoid guidewire placement [6].

The limit to clinical adoption lies elsewhere: the guidance makes no statement about its own reliability. Registration accuracy, expressed as the target registration error (TRE) at the surgical site, is reported as an aggregate over a validation dataset, whereas the quantity that governs the surgical decision is the TRE affecting the current patient at the current moment. That error is not constant: it accumulates from calibration, tracking and hand-eye estimation, and it drifts during the procedure because of patient motion, instrument occlusion, retraction and progressive change of the exposed bone surface. An overlay whose registration is still within the clinical tolerance and one whose registration has silently drifted past it are drawn on the display in exactly the same way. A recent review of evaluation metrics for AR surgical navigation states the problem in explicit terms: there is at present no way to measure the correctness of an augmentation or to predict its error during the intervention [7].

The objective of this project is to learn a model that estimates the registration error of the guidance system while the operation is in progress, and to attach a statistical guarantee to that estimate. The problem is supervised learning with an unusual constraint: the supervision signal exists only outside the operating room. The model observes quantities available at runtime and never the ground truth, and it returns a calibrated interval, not a single number, so that a stated bound holds with a controlled probability. Making a learned prediction carry a guarantee where deployment data differ from training data is the AI contribution of the project, and it converts a black box overlay into a component whose reliability can be declared (Fig. 1).

Two lines of work precede this. The uncertainty of a rigid point-based registration can be propagated analytically to the displayed overlay [8,9], which is exact when the error sources are modelled explicitly, and covers neither markerless alignment on bone nor drift accumulated during the procedure. More recently the error has been learned from data, with regression models trained on completed deformable registrations [10,11] and with deep networks that approach the intraoperative setting [12,13]. Neither line addresses bone-anchored navigation under drift, and none provides a calibration guarantee that survives a change of anatomical district. The methodological question of this project sits there.

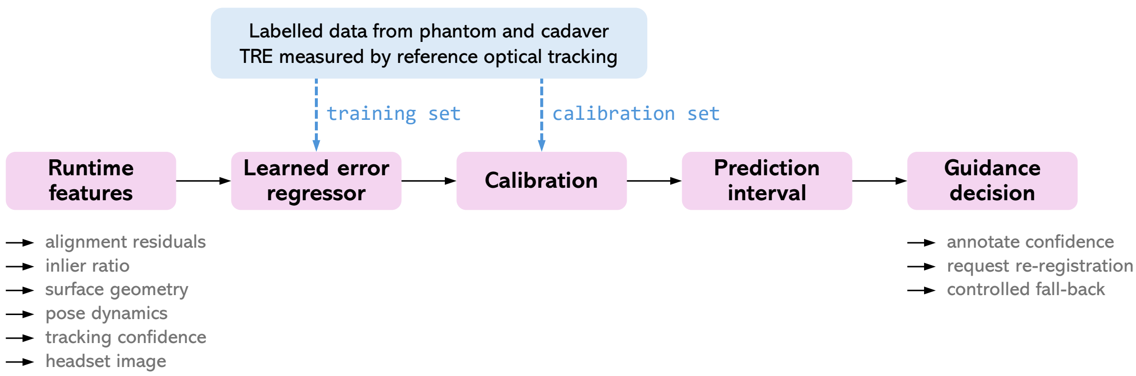

Figure 1: Inference pipeline. Labels exist only offline; the model never sees the TRE at inference time.

2 Research programme

The project is organised in four work packages (WP).

WP1. Controlled generation of a labelled dataset (M1–M8). No public dataset associates the runtime state of a surgical navigation system with the registration error that was actually present. Building one is the first task. Optical tracking of a reference rigid body provides the TRE label, while the navigation system is driven through controlled perturbations that reproduce the intraoperative failure modes: partial occlusion, degenerate surface coverage, specimen motion, tracking dropout, progressive removal of bone. Acquisitions are performed on anatomical phantoms of the scapula and of the mandible and extended to cadaveric specimens. An established metrological protocol for the assessment of AR tooltip accuracy provides the measurement basis [14].

WP2. Learning to predict the error (M6–M14). Each frame exposes a feature vector: alignment residuals and inlier ratio, geometric descriptors of the acquired surface such as coverage, curvature and conditioning, temporal statistics of the pose, tracking confidence, and the image seen by the head-mounted display. The work package establishes interpretable baselines with gradient boosting on engineered features, then moves to deep models acting directly on the point cloud and on the image stream, with aleatoric and epistemic uncertainty modelled separately [15]. Comparing the two is part of the result: an interpretable estimator is easier to defend in a clinical risk assessment.

WP3. Calibration under distribution shift (M10–M20). A predicted error is useful only if the confidence attached to it is trustworthy. Conformal prediction is a distribution-free method that yields intervals with guaranteed coverage under exchangeability [16], and exchangeability fails here in a specific way: consecutive frames are correlated, and the anatomy of the test case may differ from the training distribution. The contribution is a calibration scheme that keeps its coverage when the estimator moves from phantom to cadaver and from the shoulder to the mandible.

WP4. Integration and clinical assessment (M16–M24). The calibrated estimate becomes a runtime safety signal that gates the guidance (Fig. 2): the display is annotated with the current confidence, re-registration is requested when the upper bound exceeds the clinical tolerance, and the system degrades to conventional technique when the bound cannot be met. How the estimate is presented decides whether it is used at all: a confidence value that is misread, or that competes for attention at the moment of the gesture, degrades the decision it was meant to support. The evaluation of uncertainty displays is an open problem in visualisation research [17], with a taxonomy of its own in the medical case [18]. The interface and its assessment with users are designed with the SPARKS team of I3S, whose work concerns the engineering and evaluation of interactive systems, and the clinical assessment is run with the surgical team of the iULS at CHU de Nice.

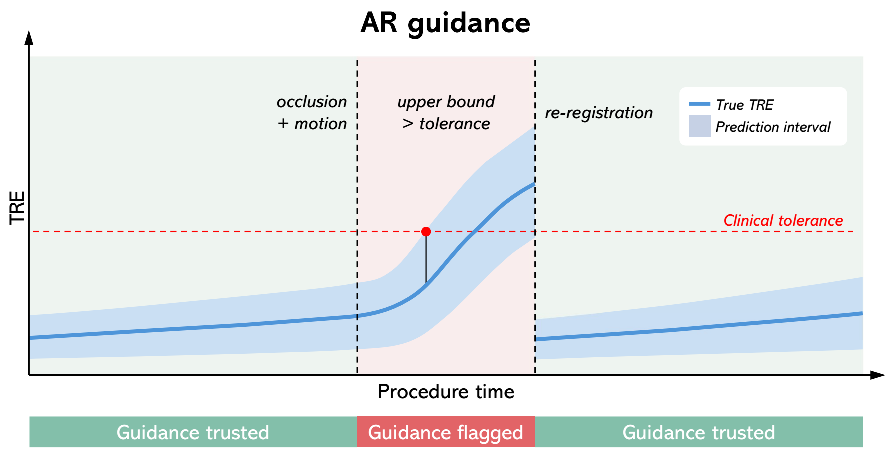

Figure 2: Expected behaviour over one procedure, drawn schematically. The decision is taken on the upper bound of the predicted interval, the only quantity observable in the operating room.

3 Expected outcome

- An open labelled dataset pairing navigation runtime features with measured TRE, on scapula and mandible, which does not currently exist.
- A calibrated machine learning model of the registration error, released as open-source software with a documented interface to standard navigation and AR pipelines.
- An evaluation protocol for the reliability of AI-based surgical guidance, for a field that currently has no reporting standard [7]. It reports empirical coverage, expected calibration error and the correlation between predicted uncertainty and measured error, and it covers the effect of the confidence display on the surgical decision, assessed with the methods established for the evaluation of interactive systems with their users [19].

- Publications in machine learning, medical image analysis and computer-assisted surgery venues, and a quantitative argument for the regulatory qualification of AI-based surgical guidance.

4 Candidate profile

Applicants should hold a PhD in engineering. The project requires a specific research background related to the following topic: optical tracking and marker-based or markerless pose estimation, rigid three-dimensional registration, computer vision algorithms applied to intraoperative data, and the design and execution of accuracy validation protocols, since producing labelled data of measurable quality is half of the work. Prior development of AR guidance systems for surgery and familiarity with head-mounted display platforms are a strong asset, as is experience in evaluating interactive systems with users. Programming in Python is required, and experience with Unity or C++ is useful on the intraoperative side. Knowledge of statistical machine learning or of uncertainty quantification is welcome and will be developed within the project, in the research environment of the 3IA Côte d'Azur institute. An interest in working alongside clinicians in the operating room is necessary, and fluent English is required.

5 References

[1] Gauci MO. Patient-specific guides in orthopedic surgery. *Orthop Traumatol Surg Res* 2022;108(1S):103154.

[2] Gauci MO, Boileau P, Baba M, Chaoui J, Walch G. Patient-specific glenoid guides provide accuracy and reproducibility in total shoulder arthroplasty. *Bone Joint J* 2016;98-B(8):1080–1085.

[3] Jacquot A, Gauci MO, Chaoui J, Baba M, Deransart P, Boileau P, Molé D, Walch G. Proper benefit of a three dimensional pre-operative planning software for glenoid component positioning in total shoulder arthroplasty. *Int Orthop* 2018;42(12):2897–2906.

[4] Ho JC, Sabesan VJ, Iannotti JP. Glenoid component retroversion is associated with osteolysis. *J Bone Joint Surg Am* 2013;95(12):e82.

[5] Farron A, Terrier A, Büchler P. Risks of loosening of a prosthetic glenoid implanted in retroversion. *J Shoulder Elbow Surg* 2006;15(4):521–526.

[6] Italia K, Launay M, Gilliland L, et al. Improving glenoid guidewire placement in shoulder arthroplasty: a comparative study of mixed reality holographic overlay technique with freehand technique and conventional navigation. *JSES Int* 2025;9(3):981–987.

[7] Malhotra S, Halabi O, Dakua SP, et al. Augmented reality in surgical navigation: a review of evaluation and validation metrics. *Appl Sci* 2023;13(3):1629.

[8] Nicolau S, Pennec X, Soler L, Ayache N. An accuracy certified augmented reality system for therapy guidance. *ECCV* 2004, LNCS 3023:79–91.

[9] Fitzpatrick JM, West JB, Maurer CR. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imaging* 1998;17(5):694–702.

[10] Sokooti H, Saygili G, Glocker B, Lelieveldt BPF, Staring M. Quantitative error prediction of medical image registration using regression forests. *Med Image Anal* 2019;56:110–121.

[11] Bierbrier J, Gueziri HE, Collins DL. Estimating medical image registration error and confidence: a taxonomy and scoping review. *Med Image Anal* 2022;81:102531.

[12] Salari S, Rasoulian A, Rivaz H, Xiao Y. FocalErrorNet: uncertainty-aware focal modulation network for inter-modal registration error estimation in ultrasound-guided neurosurgery. *MICCAI* 2023, LNCS 14229.

[13] Gheiji B, Elyassirad D, Vatanparast M, et al. CONReg: uncertainty-aware medical image registration using conformal prediction. *J Imaging Inform Med* 2026.

[14] Salerno F, Ulrich L, Maculotti G, Moos S, Genta G, Vezzetti E, Galetto M. A metrological approach for augmented reality tooltip tracking assessment. *Comput Ind* 2026;175:104430.

[15] Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *NeurIPS* 2017.

[16] Angelopoulos AN, Bates S. Conformal prediction: a gentle introduction. *Found Trends Mach Learn* 2023;16(4):494–591.

[17] Hullman J, Qiao X, Correll M, Kale A, Kay M. In pursuit of error: a survey of uncertainty visualization evaluation. *IEEE Trans Vis Comput Graph* 2019;25(1):903–913.

[18] Ristovski G, Preusser T, Hahn HK, Linsen L. Uncertainty in medical visualization: towards a taxonomy. *Comput Graph* 2014;39:60–73.

[19] Winckler M, Bernhaupt R, Bach C. Identification of UX dimensions for incident reporting systems with mobile applications in urban contexts: a longitudinal study. *Cogn Technol Work* 2016;18(4):673–694.