Human-human collaboration via robot-mediated physical interaction: Coupling machine-learning and control

I. Proposal Information

- Title: Human-human collaboration via robot-mediated physical interaction: Coupling machine-learning and control.
- Acronym: **COACT** (Human-human CollaboratiOn via robot-mediated physical interAction: Coupling machine-learning and conTrol).
- Name(s) of PhD Advisor(s): Marco Cognetti (website, GScholar)
- Host Laboratory: LAAS-CNRS, RIS team

I.1 Short abstract. This PhD investigates how two remotely located humans can physically collaborate through the mediation of a mobile robot. When a skilled expert is unavailable on-site, current remote tools limit the distant partner to an advisory role. We propose instead to deploy a robotic agent alongside the on-site worker so that the remote human can physically participate in the task (e.g., holding, aligning, or applying forces through the robot). A key challenge is that the robot must exhibit autonomous behavior to compensate for communication latency and ensure safe interaction, relying on neural-network estimators to predict the on-site human's actions. Naively coupling such learned models with a physical controller creates compounding stability risks. This project addresses this fundamental problem through three contributions: (i) neural network architectures with formal output-boundedness and uncertainty quantification guarantees; (ii) an adaptive shared-autonomy controller that modulates the robot's behavior based on the confidence of its learned estimates, with a passivity-based stability proof of the full coupled system; and (iii) human-subject experiments comparing co-located, remote-guided, and robot-mediated collaboration to assess how closely the proposed system can approach the coordination quality of direct human-human teamwork.

I.2 Short description of hosting research group / lab. LAAS-CNRS (Laboratoire d'Analyse et d'Architecture des Systèmes) is a CNRS research laboratory in Toulouse, France, with a long-standing tradition in robotics, automatic control, and artificial intelligence. The RIS (Robotics and InteractionS) team, headed by Dr. F.F. Ingrand and Dr. A. Clodic, develops research on autonomous machines that integrate perception, reasoning, learning, action, and reaction capabilities. Its research themes span decisional architectures, motion planning, planning and learning, human-robot interaction and cooperation, aerial robotics, and heterogeneous multi-robot systems. The PhD advisor, M. Cognetti, leads the Aerial Robotics research direction within RIS, focusing on the design, modeling, and control of robots that can physically interact with the environment and with humans, with specific expertise in human modeling during physical interaction, shared-autonomy control, and active perception. The proposed PhD builds directly on this expertise and on the team's broader strengths in human-robot interaction management, which addresses knowledge-based reasoning, situation assessment, task supervision, and quality of interaction across one-to-one and multi-party settings — providing both the robotic platforms and the interdisciplinary collaboration science environment necessary for this project. The PhD student will have access to the team's experimental facilities, including a large motion-capture space equipped with multiple robotic platforms (mobile manipulators, aerial robots) and a variety of sensors (force/torque, vision, haptic suite) for human-robot interaction studies. The student will also benefit from the team's strong connections with the broader robotics and human-robot interaction community, including collaborations with other research labs such as the Rainbow team in Rennes (headed by Dr. P. Robuffo Giordano, expert in shared control), ENAC (Anke Brock) and the Department of Computer Science at University College London (Dr. V. Modugno, expert in machine-learning), industry partners, and access to international conferences and workshops for dissemination and networking.

II. Description of the PhD proposal (3 pages max)

II.1 Context. Across many professional domains, critical tasks require the joint physical effort of two or more humans working in close coordination. In industrial maintenance, assembling or repairing tasks often demands one person to stabilize, hold, or position components while a second performs a complementary action (e.g., tightening or welding). In construction, two workers should coordinate to lift, align, and secure structural elements that neither could handle alone. In healthcare, surgical procedures frequently involve a primary surgeon and an assistant who simultaneously retracts tissue or applies suction. These tasks share a defining characteristic: they are inherently dyadic, i.e., they require at least two agents to act physically, simultaneously, and in tight spatial and temporal coordination.

In practice, the second skilled human is not always available on-site. The specialist may be located at a distant facility, the expert may serve multiple sites across a territory, or access to the workspace may be restricted by safety regulations, contamination, or geography. Today, when the required collaborator cannot travel, three options exist, all unsatisfactory: (i) postpone the task until co-presence is possible, at significant operational cost; (ii) replace the absent expert with a less qualified on-site worker, degrading quality and safety; or (iii) rely on remote guidance tools (e.g., video calls, voice instructions) in which the remote human can only advise but never physically act.

A fundamentally different approach is to restore the physical presence of the remote human through a robotic agent deployed alongside the on-site worker. In this paradigm, the remote human does not merely observe and instruct but they physically participate in the task by acting through the robot. The robot becomes the remote human's hands in the workspace: it can hold a component, apply a force, position a tool, or stabilize an object, under the guidance and intent of the remote operator.

II.2 Problem and Objective. When two co-located humans collaborate on a physical task, their coordination relies on a rich set of implicit and explicit cues: haptic feedback through the shared object (feeling the partner's forces and intentions), peripheral vision of the partner's posture and gaze, verbal and gestural communication, and a shared mental model of the task built through common sensory experience of the same environment. Introducing a robot as the physical embodiment of one partner fundamentally disrupts all of these channels simultaneously.

Introducing a robot as the physical embodiment of the remote partner inevitably transforms these coordination channels. While force coupling becomes indirect and mutual awareness is mediated by the robot's sensors, this mediation bridges an otherwise impossible physical gap. However, without the robot, the remote human possesses zero physical agency: no force coupling, no shared manipulation, no embodied co-action. Furthermore, pure teleoperation fundamentally is inadequate for this scenario. In fact, the remote human simply cannot react fast enough, or with enough information, to maintain the tight coordination that dyadic physical tasks demand. The robot must therefore exhibit a degree of autonomous behavior: it needs to locally compensate for delays, ensure safe interaction forces, and adapt to the on-site human's actions without waiting for the remote operator's commands. To do this, the robot must continuously estimate critical features of the collaboration — the on-site human's motion intentions, the state of the shared task, the interaction forces, the geometry and constraints of the workspace — and use these estimates to modulate its own behavior in real time.

Modern machine-learning approaches, and in particular neural networks, offer powerful tools for such perception and estimation tasks. Beyond instantaneous estimation, these techniques also enable predicting the future evolution of the collaboration state (e.g., the on-site human's forthcoming motions, the expected task progression, the anticipated interaction forces) over a finite time horizon. This predictive capability opens the door to proactive control strategies: rather than merely reacting to the current state, the robot can generate anticipatory actions that account for how the collaboration is expected to unfold, improving coordination smoothness and compensating for communication latency before its effects are felt. However, naively coupling a learned estimator — whether reactive or predictive — with a robot controller creates a fundamental stability problem that is largely overlooked in the literature. In a standard control pipeline, sensory inputs are assumed to be reliable: the controller treats them as ground-truth measurements and computes actions accordingly. When the input instead comes from a neural network, two main compounding risks arise: (i) misplaced trust: the controller has no mechanism to distinguish a confident, well-calibrated network output from an unreliable extrapolation. It treats every estimate as an "oracle", regardless of quality; (ii) instability through positive feedback: when the network produces an erroneous estimate in an out-of-distribution region, the controller reacts with a corrective action proportional to the perceived error. This strong action pushes the system further from the training distribution, producing an even worse estimate from the network, which provokes an even stronger control response — a positive feedback loop that can rapidly drive the system into instability. In a human-robot collaboration context, this instability directly translates into unpredictable, potentially dangerous physical behavior of the robot toward the on-site human.

The objectives of this proposal are:

- Breaking the black-box coupling between learning and control. Instead of treating the neural network as an "opaque" module that feeds inputs to a controller, we propose to develop architectures in which the controller is explicitly aware of the reliability of its learned inputs. Two complementary strategies will be pursued:
  - *Intrinsically stable or bounded-output networks*: design or adopt neural network architectures that are either proven to be Lyapunov-stable in closed loop with the controller, or that provide formal guarantees on the boundedness of their output regardless of the input. This prevents the network from producing arbitrarily large erroneous estimates that could destabilize the system.
  - *Uncertainty-aware control*: equip the network with uncertainty quantification (e.g., predictive variance, epistemic uncertainty estimates) and design the controller to modulate its reliance on the network output as a function of this uncertainty. When the network reports low variance, the controller trusts the estimate and acts accordingly; when variance is high the controller falls back to a conservative, safe behavior.
- Designing adaptive shared-autonomy strategies. Building on the uncertainty-aware control framework, propose strategies in which the robot dynamically adjusts its level of autonomy based on both the state of the collaboration and the confidence of its perceptual estimates (see Sec. II.5 for details).
- Experimental evaluation of collaboration quality. Human-subject experiments comparing co-located, remote-guided, and robot-mediated collaboration on dyadic physical tasks, measuring task performance, coordination quality, and subjective experience (see Sec. II.5 for details).

II.3 Brief overview of the state of the art. The state of the art surrounding this proposal sits at the intersection of robot-mediated physical interaction, neural network guarantees, and shared autonomy.

Robot-Mediated Physical Interaction: Remote collaboration has historically relied on audio-visual telepresence and shared control [4], but recent work has shifted toward physical mediation, with shared autonomy frameworks enabling dexterous manipulation under severe latency constraints[3]. *Proposed Contribution:* The proposed approach advances this paradigm by restoring true physical agency to remote experts in industrial and collaborative settings. It moves beyond traditional audio-visual telepresence into genuine, asymmetric robot-mediated physical human-human interaction, retaining the natural dyadic structure of complex tasks.

Stability-Guaranteed Neural Networks: To resolve the positive feedback instability caused by naively coupling learned estimators with physical controllers, the literature emphasizes mathematically bounded network architectures [1, 8, 9]. *Proposed Contribution:* The novelty of this research lies in the combination of machine-learning and control, analyzing the stability of the overall closed-loop system and enforcing mathematical bounds (e.g., Lipschitz continuity) at the architectural level of the neural network to explicitly cap the maximum error injected into the physical system.

Uncertainty-Aware Shared Autonomy: Uncertainty quantification has been integrated into model predictive control for dynamic robotics [5], and variable impedance control with energy-tank passivation ensures stability during autonomy transitions [2], though transparent transitions remain a challenge for user trust [6, 7]. *Proposed Contribution:* Unlike prior approaches that apply uncertainty quantification to single-agent MPC without formal ties to the learned model, our uncertainty signal originates from architecturally-bounded networks with formal output guarantees, enabling a passivity-based stability proof of the full coupled system; moreover, the confidence metric governs a three-way authority allocation (on-site human, remote human, robot autonomy) absent from existing shared-autonomy work.

II.4 Research questions.

1. **RQ1**: Which neural network architectures can provide formal guarantees on output boundedness or closed-loop stability when coupled with a robot controller operating in a human-interactive setting?
2. **RQ2**: How does the robot's level of autonomy and the transparency of its autonomy transitions affect the quality, efficiency, and subjective experience of the dyadic collaboration, both for the on-site human (who physically interacts with the robot) and the remote human (who acts through it)?
3. **RQ3**: Can robot-mediated physical collaboration, with appropriate uncertainty-aware autonomy, approach the coordination quality of co-located human-human collaboration on dyadic physical tasks?

II.5 Approach and methods. The proposed approach is structured in three phases, each building on the previous one.

1. Phase 1 - Stability-guaranteed neural networks for collaborative robotics.: The first phase addresses *RQ1* by investigating two complementary directions: (i) how to guarantee the boundedness and/or the stability of the output of transformed-based neural networks (e.g., Lipschitz-constrained layers, spectral normalization, or output saturation layers); (ii) quantify the uncertainty of the network's estimates through techniques such as Bayesian neural networks or Monte Carlo dropout, and validate the calibration of these uncertainty estimates.
2. Phase 2 - Adaptive shared-autonomy control: The second phase addresses *RQ2* and integrates the learning techniques from Phase 1 into the robot controller. We will design an optimization-based variable-impedance shared-autonomy controller whose behavior is explicitly in function of the remote human's commands and the local estimates of the on-site human's state and intentions (produced by the neural network, with associated uncertainty). The controller will operate along a continuum between two modes: (i) high-confidence regime: when the network's uncertainty is below a task-specific threshold, the robot actively assists the collaboration by anticipating the on-site human's movements; (ii) low-confidence regime: when uncertainty rises (e.g., the system has drifted into an under-represented state) the robot smoothly transitions to a conservative behavior. Concretely, it increases its mechanical compliance (becoming lighter and more backdrivable for the on-site human), reduces its autonomous contributions, and increases the weight given to the remote operator's direct commands. In the limit, it reverts to a purely passive compliant mode that guarantees physical safety regardless of the quality of the learned estimates. The transition policy between high- and low-confidence regimes will be designed to be smooth and transparent to the humans. Importantly, a central requirement in shared autonomy is the bidirectional, multimodal communication that connects both humans through the robot. The remote operator is immersed in the task via a VR headset, a haptic device rendering the robot's end-effector forces in real time, and spatial audio relaying the on-site environment. In return, the remote human's head and hand motions are mapped to the robot's camera and end-effector. For the on-site human, the robot itself is the interface: its compliance, motion, and physical cues convey the remote partner's intent. In particular, we want to investigate how different feedback channels (audio, video, and haptic) can be combined to provide a coherent, non-invasive communication of the remote human's intent and the robot's autonomy state to the on-site human.

   A formal stability analysis of the coupled system (network + controller + human interactions) will be conducted. The key insight is that the boundedness guarantees from Phase 1 provide the missing piece for classical stability proofs: because the network output is bounded by construction, and the controller's gain scales inversely with uncertainty, the total energy injected into the system by erroneous estimates can be upper-bounded. This enables a passivity-based proof that the robot's behavior remains stable (i.e., the interaction forces and velocities remain bounded) regardless of the accuracy of the learned human model.
3. Phase 3 — Experimental evaluation of robot-mediated dyadic collaboration. The third phase addresses *RQ2* and *RQ3* through human-subject experiments. As a representative scenario, consider a maintenance operation in a confined industrial facility: two agents must collaboratively mount a panel onto a frame equipped with alignment pins. One agent holds and adjusts the panel's orientation against gravity while the other guides it onto the pins and secures the fasteners. This task demands continuous, simultaneous force coordination and cannot be performed sequentially or by a single operator. This task will be performed under four conditions: (i) *co-located baseline*: both humans are physically present and collaborate directly. This is the "ground-truth"; (ii) *remote voice/video*: one human is remote and can only guide the on-site human verbally and visually; (iii) *robot-mediated, fixed teleoperation*: the remote human controls the robot with standard bilateral teleoperation; and (iv) *robot-mediated, uncertainty-aware autonomy*: the full system proposed in this project. For each condition, we will measure: (i) objective task performance (e.g., task completion time, positioning accuracy); (ii) coordination quality (e.g., synchronization between the two agents); and (iii) subjective experience through standardized questionnaires assessing cognitive load (NASA-TLX), asking for trust in the robot, sense of partnership with the other human, perceived agency, and collaboration fluency. Particular attention will be paid to comparing conditions 3 and 4, to isolate the contribution of the uncertainty-aware autonomy to collaboration quality, and to comparing condition 4 with condition 1, to assess how close robot-mediated collaboration can approach co-located performance.

The contributions of this PhD thesis will be on: foundational learning (year 1), control integration (year 2), experimental validation (year 3), with publications possible after each phase.

III. Nature of digital collaboration (1 page max)

III.1 Function (communication, sharing, coordination, other) The collaboration targeted by this PhD is primarily physical coordination: two humans must exert simultaneous, coupled forces on a shared object to achieve a common manipulation goal. This coordination is mediated by a robotic agent that transmits forces, motions, and intentions between the partners. Communication (verbal, visual, and haptic) and sharing (of workspace perception, task state, and confidence information) are embedded within this coordination loop but are instrumental to it rather than goals in themselves.

III.2 Type (synchronous, asynchronous) The collaboration is strictly synchronous. Dyadic physical tasks (e.g., holding, aligning, inserting) require both agents to act simultaneously and to adapt to each other's forces and motions in real time (sub-second loop). Asynchronous contributions (e.g., one partner preparing while the other is absent) fall outside the scope of this work.

III.3 Time scale (second, hours, months, years,..) The relevant time scales span from milliseconds to minutes. At the lowest level, force and impedance control loops operate at 1 kHz (milliseconds). Human motor coordination and intention estimation unfold over hundreds of milliseconds. A single collaborative subtask (e.g., aligning and fastening a panel) lasts seconds to minutes. The full experimental session per dyad is on the order of one hour. Longer-horizon collaboration dynamics (learning across days, team formation) are not addressed.

III.4 Group size The collaboration involves exactly two humans: one on-site, one remote, with a robot acting as the physical mediator. The triad (on-site human, robot, remote human) is the fundamental unit studied. At the beginning of the project, the focus will be on two humans collaboration to establish the core principles and validate the approach. In the last year, the framework could be extended to a small group of three or four humans (e.g., two on-site workers collaborating with two remote experts through two robots) to explore how the uncertainty-aware shared-autonomy framework scales with additional collaborators and how the robot can mediate more complex multi-party interactions.

III.5 Space (co-located, remote, hybrid) The collaboration is inherently hybrid: one human is co-located with the robot and the task, while the other is remote. The on-site human interacts through direct physical contact; the remote human interacts through the robot's teleoperation interface. The research explicitly studies how this spatial asymmetry affects coordination quality compared to fully co-located and fully remote (voice/video only) baselines.

III.6 Other Two additional dimensions are relevant. First, symmetry of roles: unlike most remote collaboration tools where the remote party is limited to an advisory role, the proposed system restores a physically active role to the remote human, creating a more symmetric partnership. Second, autonomy as a collaboration variable: the robot is not a transparent channel but an active third agent whose level of autonomous intervention varies dynamically, making the nature of the collaboration itself adaptive.

IV. Contribution to digital collaboration: Expected results and Impact (1 page max)

IV.1 What type of contribution(s) is the PhD expected to make? The PhD is expected to make contributions across four dimensions: theoretical, methodological, technical, and empirical.

IV.1.1 Theoretical The PhD will deliver a formal stability analysis of the closed-loop system coupling a neural-network with a shared-autonomy controller and two human partners. By combining architectural output bounds (Lipschitz constraints) with uncertainty-modulated control gains and energy-tank passivation, the analysis will yield a passivity-based proof guaranteeing bounded interaction forces and velocities regardless of the accuracy of the learned model. This constitutes a theoretical contribution at the intersection of machine learning and control theory, applicable beyond the specific collaboration scenario. For the digital collaboration community, this provides the first formal safety guarantee for a collaboration technology in which a computational agent actively mediates physical interaction between humans, ensuring that the technology cannot degrade collaboration below a provable safety floor, a property absent from existing collaboration tools.

IV.1.2 Methodological The project will produce a design methodology for safely integrating learned perception modules into physical human-robot interaction controllers. Concretely, this includes: (i) guidelines for selecting and enforcing architectural constraints on neural networks destined for closed-loop control; (ii) a principled procedure for calibrating uncertainty estimates and mapping them to controller authority levels; and (iii) a three-way authority allocation scheme (on-site human, remote human, robot) with transparent transition policies. Beyond robotics, this methodology offers a template for any digitally-mediated collaboration system in which an AI agent must decide when to intervene, when to defer to one user, and when to defer to another — a problem that generalizes to collaborative editing, shared decision-making, and mixed-initiative interaction design.

IV.1.3 Technical The PhD will produce an integrated software and hardware demonstrator: a robotic platform capable of mediating a dyadic physical task between a co-located and a remote human, equipped with stability-guaranteed neural-network estimators and an adaptive shared-autonomy controller. The software stack (encompassing the bounded-output network architectures, the uncertainty-aware control pipeline, and the experimental measurement framework) will be released as open-source to enable reproducibility and adoption by the community.

IV.1.4 Empirical The Phase 3 experiments will produce the first controlled, within-subject comparison of co-located, remote-guided, and robot-mediated dyadic physical collaboration on the same task. The resulting dataset — comprising objective performance metrics (completion time, positioning accuracy, peak forces), force/motion coordination signals (cross-correlation, synchronization indices, coordination breakdowns), and standardized subjective scales (NASA-TLX, trust, collaboration fluency) — will provide quantitative evidence on how much coordination quality a robot mediator can recover relative to co-located interaction, and how uncertainty-aware autonomy improves upon fixed teleoperation. From the digital collaboration perspective, these results will establish, for the first time, where robot-mediated physical co-action sits on the spectrum between purely informational remote collaboration (video/voice) and fully co-located teamwork. This fills a gap in the collaboration science literature, which has extensively studied screen-mediated and telepresence-mediated collaboration but has virtually no data on technology-mediated physical joint action between remote partners.

IV.1.5 Expected impact on digital collaboration research. More broadly, this PhD extends the scope of digital collaboration from the informational domain (sharing documents, screens, audio-video streams) into the physical domain (sharing forces, motions, and manipulation capabilities). It demonstrates that digital collaboration need not be limited to exchanging symbols and signals: with appropriate robotic mediation and AI-driven autonomy, remote partners can become genuine physical co-actors. The three-way authority allocation framework (two humans and one autonomous agent) also provides a concrete instantiation of human-AI-human mediation collaboration pattern, that is increasingly relevant as AI systems take on mediating roles in group work, negotiation, and cooperative problem-solving. The experimental protocol and metrics developed in Phase 3 can be adapted by the collaboration science community to evaluate future systems that blend physical and digital interaction channels.

References

[1] Hongkai Dai et al. "Lyapunov-stable neural-network control". In: *Robotics: Science and Systems XVII*. 2021.

[2] Shail Jadav et al. "Shared Autonomy via Variable Impedance Control and Virtual Potential Fields for Encoding Human Demonstrations". In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2024, pp. 15151–15157.

[3] Amy Phung et al. "A Shared Autonomy System for Precise and Efficient Remote Underwater Manipulation". In: *IEEE Transactions on Robotics* 40 (2024), pp. 4147–4166.

[4] Pragathi Praveena et al. "Periscope: A Robotic Camera System to Support Remote Physical Collaboration". In: *Proceedings of the ACM on Human-Computer Interaction, Issue CSCW*. ACM, 2023.

[5] Alessandro Saviolo et al. "Active Learning of Discrete-Time Dynamics for Uncertainty-Aware Model Predictive Control". In: *IEEE Transactions on Robotics* 40 (2024), pp. 1273–1291.

[6] Svenja Y Sch"ott et al. "A Literature Survey of How to Convey Transparency in Co-Located Human–Robot Interaction". In: *Multimodal Technologies and Interaction* 7.3 (2023), p. 25.

[7] Mario Selvaggio et al. "Autonomy in Physical Human-Robot Interaction: A Brief Survey". In: *IEEE Robotics and Automation Letters* 6.4 (2021), pp. 7989–7996.

[8] Xujie Song et al. "LipsNet: A Smooth and Robust Neural Network with Adaptive Lipschitz Constant for High Accuracy Optimal Control". In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR. 2023, pp. 32253–32272.

[9] Siqi Zhou et al. "Bridging the model-reality gap with Lipschitz network adaptation". In: *IEEE Robotics and Automation Letters* 7.1 (2022), pp. 642–649.