Uncertainty-Aware Conversations in Stroke Rehabilitation

Valerio Bonsignori and **Fosca Giannotti** *Scuola Normale Superiore,*
{valerio.bonsignori,fosca.giannotti}@sns.it
Carlo Metta and **Salvatore Rinzivillo** *Consiglio Nazionale delle Ricerche,*
{carlo.metta,salvatore.rinzivillo}@cnr.it
Francesca Cecchi, Badia Bahia Hakiki and **Stefano Doronzio**
Università di Firenze, IRCCS Fondazione don Carlo Gnocchi,
{francesca.cecchi,badiabahia.hakiki,stefanogiuseppe.doronzio}@unifi.it
Andrea Mannini *IRCCS Fondazione don Carlo Gnocchi*
{amannini}@dongnocchi.it

Abstract

Machine learning deployment in scientific domains faces a critical bottleneck: domain experts struggle to interpret AI-generated insights. We present an uncertainty-aware conversational AI framework that transforms complex ML explanations into natural language for domain experts. Our approach integrates selective classification with conversational explanation using messaging interfaces. We demonstrate this through a chatbot enabling rehabilitation physicians to analyse stroke patient data, with predictions accompanied by natural language explanations and explicit confidence communication. This addresses the intersection of uncertainty quantification and conversational explanation delivery, offering a generalizable methodology for deploying interpretable AI in expert domains.

Keywords: Explainable Artificial Intelligence, Uncertainty Quantification, Conversational AI, Rehabilitation

1 Introduction

The integration of AI into scientific research faces fundamental communication barriers. While domain scientists possess irreplaceable expertise, they often lack the technical backgrounds necessary to understand algorithmic explanations, as demonstrated by Slack et al. [2023]. Although traditional XAI approaches generate technically accurate explanations, Nguyen et al. [2024] argues that they offer limited guidance on how to communicate these explanations effectively to domain experts. Furthermore, dashboard-based interfaces frequently overwhelm users with technical information and fail to support the iterative, exploratory nature of scientific research, say Ampomah et al. [2022].

Conversational AI presents a promising solution by transforming complex concepts into accessible dialogue, as shown by Klievtsova et al. [2023]. Yang et al. [2024] says that existing conversational AI systems rarely address uncertainty quantification, critical for maintaining expert oversight and preventing over-reliance. Reyes et al. [2025] suggests that uncertainty communication relies on visual representations and does not translate to conversational interfaces.

We introduce uncertainty-aware conversational AI for scientific knowledge discovery, combining selective classification with natural language generation. **Contributions:** (1) a conversational AI framework designed to communicate both ML explanations and prediction uncertainty to domain experts; (2) practical implementation translating XAI outputs into clinically contextualised natural language; (3) generalizable methodology for deploying uncertainty-aware conversational AI beyond healthcare.

2 Related Work

Slack et al. [2023] demonstrated that the TalkToModel interface was preferred 73% of the time among healthcare workers over traditional dashboard interfaces, with significantly faster task completion (76s vs 158s). However, most conversational AI work focuses on explanation delivery without quantification of uncertainty. Klievtsova et al. [2023] shows that conversational process modelling frameworks have potential for empowering domain experts to interact with AI systems without deep machine learning knowledge.

Research by Nguyen et al. [2024] in human-centered explainable AI finds that more than 30% of users are unable to understand standard explanations. Technical explanations often fail to support decision-making in professional contexts. Recent work by Ampomah et al. [2022] emphasises translating ML performance metrics into textual explanations tailored for specific domains. Schoonderwoerd et al. [2021] identifies clinical design patterns that emphasise contextual relevance, progressive disclosure, and workflow integration, but implementation in conversational interfaces remains unexplored.

Reyes et al. [2025] reveals that distance-based uncertainty representations outperform probability scores, with healthcare professionals achieving 8.20% higher decision accuracy. Prabhudesai et al. [2023] shows that uncertainty visualisation can promote analytical thinking and reduce over-reliance on AI predictions. Recent work by Traub et al. [2024], Yang et al. [2024] demonstrates effective uncertainty communication in clinical deployment.

Meta-analysis by Vaccaro et al. [2024] of 106 experimental studies shows human-AI combinations achieve augmentation when properly designed. Sarailidis et al. [2023] demonstrates that interactive systems enabling domain experts to integrate scientific knowledge directly into ML processes show higher interpretability and robustness, emphasising complementarity over replacement.

3 Methods

Framework Architecture: our uncertainty-aware conversational AI framework combines: (1) a selective classification pipeline generating predictions with confidence estimates, (2) a multi-modal explanation generator producing SHAP

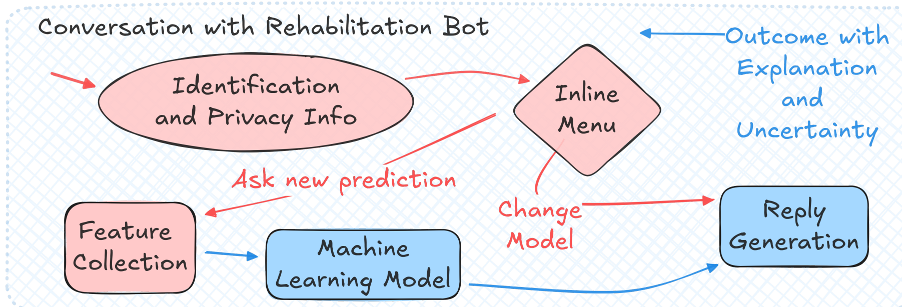

Figure 1: The main loop of the conversational bot. The explanation generator is disjoint from the ML Model, as explanations are model-agnostic.

values, counterfactuals, and decision rules, and (3) a natural language interface translating technical outputs into domain conversations, see Figure 1.

We implement various competitive models, each of which generates predictions along with confidence estimates, allowing for abstention on uncertain predictions and varying confidence levels to be communicated to users. **Uncertainty Communication Design** Our conversational uncertainty communication combines textual and visual elements within messaging interfaces. Prediction confidence uses dynamic emoji selection: high confidence (≥ 0.8) uses 😎, moderate confidence ($\geq$0.7) uses 😐, and low confidence uses 🤔. This provides immediate uncertainty awareness while maintaining conversational naturalness. The system explicitly recommends acceptance or rejection based on confidence thresholds, supporting calibrated trust development.

Explanation Translation Pipeline We developed a template-based natural language generator system converting XAI outputs into contextual explanations. For feature importance, we employ Kernel SHAP, a feature importance explainer theorized by Lundberg et al. [2017], a model-agnostic approach that quantifies each feature's contribution to individual predictions. Counterfactual explanations are generated using LoRE (Local Rule-based Explanations), introduced by Guidotti et al. [2024]. These technical outputs are translated into natural language statements tailored to the rehabilitation domain.

Implementation We instantiated this framework for stroke rehabilitation outcome prediction using clinical assessment data from Finocchi et al. [2024]. The dataset includes stroke patients who were able to walk (ambulatory) prior to their stroke event. Patient descriptors include: Functional Ambulation Categories scale (FAC), Mini-Mental State Examination cognitive assessment (MMSE), functional independence measures (modified Barthel Index), area of lesion, and demographic factors (age, gender, education, cohabitation status).

The chatbot is deployed via Telegram Bot API[1], enabling rehabilitation physicians to analyse patient cases through natural conversation while accessing sophisticated ML explanations. Figure 2 illustrates the chatbot's output.

[1]A known limitation is the API does not support end-to-end encryption.

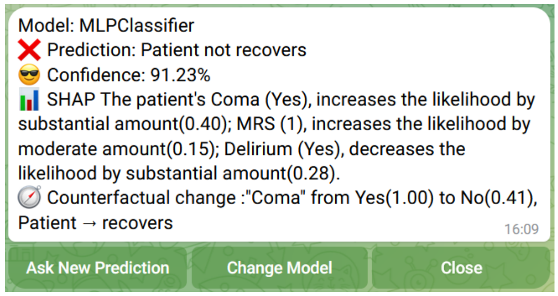

Figure 2: An example of the chatbot's output for the Multi Layer Perceptron. The confidence of the model on the sample is shown just after the outcome. The explanations that follow include SHAP-based feature importance, counterfactuals.

The physician inputs patient features using an inline menu interface, and the system responds with predictions from multiple models. Each outcome is accompanied by confidence scores and detailed explanations including SHAP values, counterfactuals, and decision rules (for tree models). The explanation generator produces technical XAI artifacts, which the chatbot enriches with human-understandable, domain-expert-contextualized natural language.

Privacy and Data Protection Our system processes sensitive health data under GDPR Article 9(2)(j) for scientific research purposes. A Data Protection Impact Assessment (DPIA) was conducted to identify and mitigate privacy risks. Data minimization principles are applied: the system collects only essential clinical features without direct patient identifiers (names, ID numbers). All data is pseudo-anonymized at source. Security measures include TLS encryption for data in transit, password-protected access restricted to authorized medical staff, and role-based access controls. Query data is not persisted; only minimal anonymized logs are retained for system monitoring.

4 Results and Discussion

Framework Demonstration Our implementation integrates uncertainty quantification with conversational explanation delivery. The system processes tabular clinical data through multiple ML models, generates comprehensive explanations using known XAI techniques, and translates outputs into natural language, preserving both technical accuracy and domain relevance.

The conversational interface enables iterative exploration of model predictions, allowing users to inspect different explanation types, compare model outputs, and explore alternative scenarios through counterfactual reasoning. Uncertainty communication through textual and visual cues provides a transparent indication of prediction reliability without disrupting conversational flow.

Addressing the Knowledge Transfer Gap: Traditional XAI approaches fail to bridge the gap between algorithmic explanations and domain expert understanding. Our framework addresses this through: (1) natural language translation of technical explanations, (2) domain-contextualised interpretation of ML outputs, and (3) integrated uncertainty communication supporting appropriate trust calibration. Laranjo et al. [2018] shows that the messaging platform deployment strategy reduces adoption barriers by leveraging familiar interfaces rather than requiring specialised software, aligning with findings that interface accessibility significantly impacts expert acceptance.

Implications for Scientific AI: Our work demonstrates the feasibility of uncertainty-aware conversational AI for scientific knowledge discovery, with implications extending beyond clinical applications. Framework principles, selective classification, multi-modal explanation generation, and natural language translation are generalizable to domains where experts need to understand and act with AI-generated insights. The emphasis on maintaining expert agency while providing AI assistance aligns with successful human-AI collaboration patterns identified by Vaccaro et al. [2024]. By enabling experts to control exploration while accessing sophisticated algorithmic reasoning, our framework supports complementarity rather than replacement.

Limitations and Future Work Current limitations include: (i) reliance on template-based natural language generation, which may not capture all nuances of domain-specific communication; (ii) the Telegram Bot API's lack of end-to-end encryption presents a transport security consideration for handling sensitive health data in production deployments; (iii) the evaluation through formal user studies with rehabilitation physicians remains as planned future work. We envision exploring large language model integration for more flexible explanation generation. Controlled user studies with domain experts will assess explanation effectiveness, trust calibration, and decision-making support in practice. For clinical deployment beyond research contexts, migration to a GDPR-compliant infrastructure with end-to-end encryption would be implemented.

5 Conclusion

We present an uncertainty-aware conversational AI framework specifically designed for scientific knowledge discovery, addressing critical deployment gaps in expert domains. Our approach successfully integrates selective classification with natural language explanation delivery, enabling domain experts to access sophisticated ML reasoning through familiar conversational interfaces while maintaining appropriate awareness of prediction uncertainty. The rehabilitation medicine demonstration shows feasibility with broader scientific potential.

By preserving expert agency while providing AI assistance, our framework supports human-AI collaboration rather than replacement. This work shows research directions at the intersection of uncertainty quantification, interactive AI, and scientific knowledge discovery, showing how the translation of AI advances can be made practical as tools that amplify expertise across different domains.

References

I. Ampomah, J. Burton, A. Enshaei, and N. Al Moubayed. Generating textual explanations for machine learning models performance: A table-to-text task. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3542–3551, 2022.

A. Finocchi, S. Campagnini, A. Mannini, S. Doronzio, M. Baccini, B. Hakiki, D. Bardi, A. Grippo, C. Macchi, J. Navarro Solano, et al. Multiple imputation integrated to machine learning: predicting post-stroke recovery of ambulation after intensive inpatient rehabilitation. *Scientific Reports*, 14(1):25188, 2024.

R. Guidotti, A. Monreale, S. Ruggieri, F. Naretto, F. Turini, D. Pedreschi, and F. Giannotti. Stable and actionable explanations of black-box models through factual and counterfactual rules. *Data Min. Knowl. Discov.*, 38(5):2825–2862, 2024.

N. Klievtsova, J.-V. Benzin, J. Mangler, and S. Rinderle-Ma. Conversational process modeling: Can generative ai empower domain experts in creating and redesigning process models? *arXiv preprint arXiv:2304.11065v2*, 2023.

L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9), 2018.

S. M. Lundberg et al. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 2017.

T. Nguyen, Q.-T. Le, D. Phung, et al. How human-centered explainable ai interface are designed and evaluated: A systematic survey. *arXiv preprint arXiv:2403.14496*, 2024.

S. Prabhudesai, L. Yang, S. Asthana, X. Huan, Q. V. Liao, and N. Banovic. Understanding uncertainty: how lay decision-makers perceive and interpret uncertainty in human-ai decision making. In *Proceedings of the 28th international conference on intelligent user interfaces*, pages 379–396, 2023.

J. Reyes, A. U. Batmaz, and M. Kersten-Oertel. Trusting ai: does uncertainty visualization affect decision-making? *Frontiers in Computer Science*, 7:1464348, 2025.

G. Sarailidis, T. Wagener, and F. Pianosi. Integrating scientific knowledge into machine learning using interactive decision trees. *Computers & Geosciences*, 170:105248, 2023.

T. Schoonderwoerd, W. Jorritsma, M. Neerincx, and K. Bosch. Human-centered xai: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies*, 154, 2021. doi: 10.1016/j.ijhcs.2021.102684.

D. Slack, S. Krishna, H. Lakkaraju, and S. Singh. Explaining machine learning models with interactive natural language conversations using talktomodel. *Nature Machine Intelligence*, 5:873–883, 2023.

J. Traub, T. J. Bungert, C. T. Lüth, M. Baumgartner, K. H. Maier-Hein, L. Maier-Hein, and P. F. Jäger. Overcoming common flaws in the evaluation of selective classification systems. *Advances in Neural Information Processing Systems*, 37:2323–2347, 2024.

M. Vaccaro, A. Almaatouq, and T. Malone. When combinations of humans and ai are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8:2293–2303, 2024.

M. Yang, H. Chen, W. Hu, M. Mischi, C. Shan, J. Li, X. Long, and C. Liu. Development and validation of an interpretable conformal predictor to predict sepsis mortality risk: retrospective cohort study. *Journal of Medical Internet Research*, 26, 2024.