A Multi-Stage AI-Assisted Framework for Formalizing Narrative Clinical Pathways into FHIR PlanDefinition

Giovanni Donato Gallo*, Patrick-Benjamin Bök†, Daniela Micucci*
* *University of Milano-Bicocca*, *Department of Informatics, Systems, and Communication*, Milan, Italy
{giovanni.gallo, daniela.micucci}@unimib.it
† *Rhine-Waal University of Applied Sciences*, *Faculty of Communication and Environment*, Kamp-Lintfort, Germany
patrick-benjamin.boek@hochschule-rhein-waal.de

***Abstract*—The digital transformation of smart cities increasingly includes healthcare as a core domain, where the adoption of interoperable systems and the automation of clinical processes are essential to support healthcare professionals in delivering coordinated and efficient care. However, clinical pathways are still predominantly expressed as free-text narratives, which limits their integration into such interoperable and automated digital health ecosystems. This paper presents `SHAPE`, a multi-stage pipeline for converting unstructured clinical pathways into HL7 FHIR R5 `PlanDefinition` resources. The approach combines LLM-based reasoning, Retrieval-Augmented Generation (RAG), and deterministic composition to produce structured and standards-compliant representations. The pipeline enables action extraction, relationship modeling, and semantic grounding via SNOMED CT, supporting interoperable care pathways for coordination, automation, and clinical decision support in smart healthcare infrastructures.**

***Index Terms*—Healthcare Interoperability, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Semantic Enrichment**

I. Introduction

The increasing adoption of digital health technologies has highlighted a critical gap in healthcare interoperability: while Electronic Health Records (EHRs) are now widespread, clinical pathways are still predominantly documented as free-text narratives. In this work, clinical pathways refer to textual descriptions of coordinated clinical activities, including treatments, monitoring procedures, follow-up actions, and their temporal and logical dependencies. Although suitable for human interpretation, these textual descriptions cannot be directly integrated into digital health platforms for workflow automation, remote patient monitoring, care coordination, or clinical decision support [1].

This limitation has significant implications for smart cities investing in integrated health platforms, continuity of care, and digital chronic disease management programs, which are unable to fully exploit the clinical knowledge embedded in narrative clinical pathways unless it is translated into standardized, machine-readable, and executable representations

Recent advances in Large Language Models (LLMs) have demonstrated promising capabilities in extracting structured information from clinical text [2]. However, existing approaches mainly focus on converting clinical notes into patient-specific FHIR resources, such as observations, diagnoses, or medications [3], [4]. The automatic generation of higher-level *knowledge artifacts*, capable of representing prospective care pathways and reusable treatment protocols as computable structures, remains largely unexplored.

To address these challenges, we present `SHAPE`, a framework for transforming unstructured clinical pathways into standardized HL7 FHIR R5 `PlanDefinition` resources. The proposed approach integrates LLM-based reasoning, retrieval-augmented generation (RAG) for semantic grounding, and deterministic composition to enable the formalization of care pathways while ensuring clinical validity and interoperability.

II. Methodology

Figure 1 presents the `SHAPE` multi-stage architecture for transforming free-text clinical pathways into FHIR `PlanDefinition` resources. The approach adopts a separation-of-concerns strategy that combines three complementary computational paradigms. *LLM-based reasoning* is employed for tasks requiring semantic interpretation of natural languange clinical text, while *RAG-based grounding* constrains generation through retrieval from validated knowledge sources, reducing hallucination risks. *Deterministic procedural construction* is applied to enforce FHIR structural constraints and guarantee resource consistency. A Human-in-the-Loop (HITL) layer enables expert review, error correction, and systematic evaluation of intermediate and final outputs.

The first stage transforms the unstructured input text into a structured representation of the clinical pathway as a directed graph of clinical actions and their dependencies. This stage is implemented as a single LLM invocation guided by explicit chain-of-thought reasoning instructions. The model is prompted to identify atomic clinical actions within the text and to infer their temporal and logical relationships before generating any structured output. This design encourages a global understanding of the clinical pathway and reduces the likelihood of premature classification errors.

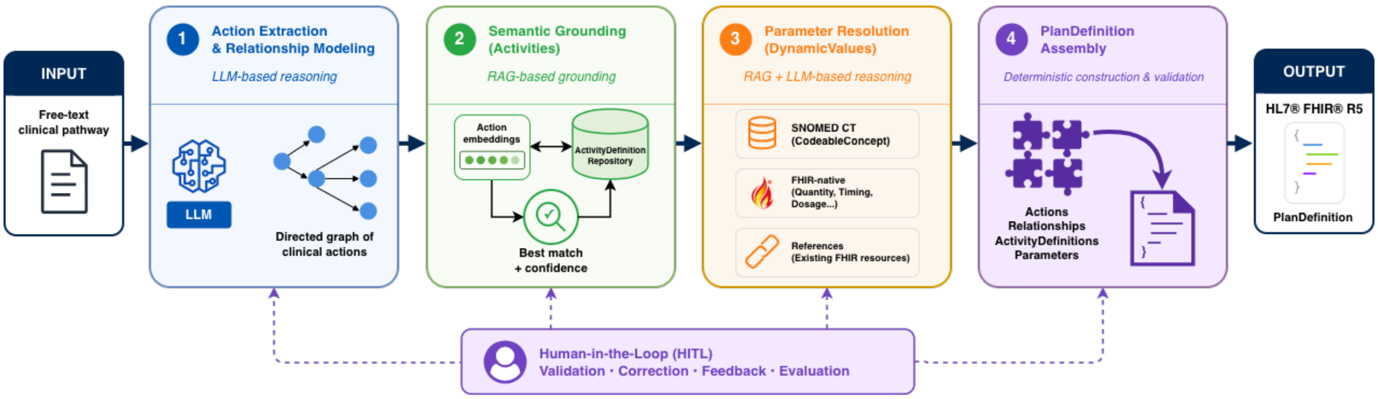

Fig. 1. Conceptual overview of the `SHAPE` pipeline architecture.

The second stage maps identified clinical actions to a curated repository of FHIR `ActivityDefinition` resources, which provides reusable and standards-compliant representations of clinical activities and their parameters. The matching process adopts a retrieve-then-select strategy: candidate resources are first retrieved through dense embedding similarity in a vector index, and subsequently ranked by an LLM according to semantic relevance, clinical context, and specificity. The model also estimates a confidence level to support HITL-based validation when required. The repository is maintained through clinical and technical expert curation, ensuring validity, FHIR compliance, and extensibility without requiring modifications to the pipeline architecture.

Each `ActivityDefinition` exposes a set of configurable parameters through `dynamicValue` elements. The third stage identifies parameter values in the action text and converts them into FHIR-compliant representations using a metadata-driven resolution framework. `CodeableConcept` parameters are grounded to SNOMED CT concepts through retrieval-augmented semantic matching, while FHIR-native types (e.g., `Quantity`, `Dosage`, and `Timing`) are generated using structure-guided prompts derived from FHIR definitions. Temporal constraints extracted from the clinical pathway are likewise encoded as FHIR timing structures. Finally, reference parameters are linked to existing FHIR resources through a cascading resolution strategy, with unresolved cases escalated to HITL review.

The final stage assembles the outputs produced by the previous stages into a structurally valid FHIR R5 `PlanDefinition` resource. Once action extraction, semantic grounding, and parameter resolution have been completed, resource generation becomes a fully deterministic process. The resulting `PlanDefinition` is subsequently validated against the official FHIR specifications, ensuring structural compliance while minimizing the risk of syntactic inconsistencies and hallucinated representations.

III. Digital Health and Well-being Implications

`SHAPE` addresses the interoperability gap between narrative clinical knowledge and computable care pathways by enabling the transformation of clinical pathways into executable, standardized structures. The framework supports care coordination across heterogeneous healthcare environments, facilitating workflow automation, remote patient monitoring, and continuity of care. By formalizing clinical pathways, it provides a foundation for personalized management of chronic conditions and scalable monitoring through consistent, structured processes. The proposed privacy-preserving design prevents the exposure of patient-level data to LLMs, limiting processing to abstracted representations of care logic. The combination of FHIR standards, clinical ontologies (e.g., SNOMED CT and LOINC), open-source LLMs, and Human-in-the-Loop supervision promotes interoperability, transparency, and safety, positioning AI as an enabling infrastructure that operationalizes clinical expertise rather than replacing it.

IV. Future Work

Future work will focus on extending the formalization of conditional action logic into fully FHIR-compliant executable conditions. In addition, a comprehensive evaluation framework will be developed to quantitatively assess pipeline performance, including a comparative analysis of different open-source LLMs against a single-shot `PlanDefinition` generation baseline. The experimental evaluation will also investigate the impact of HITL supervision, comparing step-by-step validation against final-stage-only review in terms of quality, efficiency, and clinical reliability of the generated artifacts.

References

[1] "A corpus-driven standardization framework for encoding clinical problems with hl7 fhir," *Journal of Biomedical Informatics*, vol. 110, p. 103541, 2020.

[2] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, no. 7972, pp. 172–180, 2023.

[3] Y. Li, H. Wang, H. Z. Yerebakan, Y. Shinagawa, and Y. Luo, "Fhir-gpt enhances health interoperability with large language models," *NEJM AI*, vol. 1, no. 8, p. AIcs2300301, 2024.

[4] T. A. Miller, A. J. McMurry, J. Jones, D. Gottlieb, and K. D. Mandl, "The SMART Text2FHIR pipeline," *AMIA Annual Symposium proceedings*, vol. 2023, pp. 514–520, Jan. 2024.