Abstract

This dissertation investigates how *type-specific generation*, *multi-hop reasoning*, and *explainability* can be jointly operationalized to make educational question generation (QG) more controllable, pedagogically aligned, and auditable. First, we reformulate QG as a structure-first task (Chapter 3), aligning question types with instructional taxonomies, elevating multi-hop reasoning to an explicit design prior, and embedding explainability as a cross-stage principle rather than a post-hoc add-on. Second, we examine large language models (LLMs) under type-specific objectives (Chapter 4) and show that prompt-based control, paired with post-generation validation, can steer models toward targeted educational aims while mitigating variance and hallucination. Third, we propose a multi-hop pipeline (Chapter 5) that extracts and canonicalizes relations, constructs chains, and performs *content planning* by selecting a chain and a *target answer A* (a contiguous span of one or more nodes). By varying the position and length of A, the pipeline *controls reasoning depth* (hop distance) and thereby calibrates question difficulty. We operationalize explainability-as-verifiability via chain-based views, and we observe a marked efficiency gain in human validation for chain views (mean 122.9 s vs. 244.0 s; 49.6% faster), alongside an efficiency–trust trade-off in subjective ratings. Overall, the results provide *initial* evidence—under pilot-scale evaluation—that multi-hop modeling plus embedded explainability improves controllability over reasoning depth and evaluation efficiency, with design choices needed to balance transparency and perceived trust. We conclude with directions for scaling, hybrid explanation formats, and classroom-based validation.

Keywords: Educational Question Generation; Large Language Models; Multi-Hop Reasoning; Explainability; Controllability