

Explainable AI Mediation through Knowledge Graph Design for Psychoeducational Intervention Co-Creation

Triet Bui¹, Jia Rong^{1,2} and Jue Xie^{1,*}

¹Action Lab, Monash University, Melbourne, Australia

²Centre for Learning Analytics Monash, Monash University, Melbourne, Australia

Abstract

This paper investigates how knowledge graph enhanced retrieval-augmented generation can support the orchestration of human and AI agency in the co-creation of parent psychoeducational interventions. We conceptualise intervention development as a high-stakes learning design problem, in which clinicians and domain experts must translate evidence into resources that enable parents to learn, reflect, and adapt strategies across diverse family contexts. We compare chunk-based and proposition-based knowledge graph representations to examine how representational choices influence evidence-grounded retrieval, traceability, and expert control during content curation and adaptation. Rather than positioning AI as an autonomous content generator, we frame it as an explainable mediator that surfaces relevant evidence while preserving professional judgement and accountability. Our findings highlight how knowledge graph design shapes transparency and collaboration in human-AI workflows. We discuss implications for orchestrating human-AI collaboration, evaluating co-creative intervention design processes, and governing AI use in safety-critical psychoeducational settings.

Keywords

Knowledge Graph, Retrieval-Augmented Generation, Explainable AI, Human-AI Collaboration, Psychoeducational interventions, Knowledge Representation

1. Introduction

Collaborative mental healthcare often involves ongoing negotiations between parents, young people, and clinicians, where differences in knowledge, expectations, and priorities shape how care is understood and enacted. These differences also influence the way learning supports are designed, delivered and interpreted. In this context, parenting interventions play a critical role in supporting young people's mental health while also fostering parents' ongoing learning, reflection, self-regulation, and co-regulation with their child. This highlights parenting interventions as not only a clinical effort but a learning design problem. Thus, their effectiveness depends not only on the quality of evidence, but also on how that evidence is translated into guidance that is accessible, consistent, and engaging for diverse families over time, as parents actively adapt and make sense of strategies within their evolving contexts.

Web-based psychoeducational programs such as Partners in Parenting (PiP) [1] aim to bridge this gap by providing personalised, evidence-based resources for parents. Yet, curating and maintaining high-quality intervention content remains a complex and resource-intensive process. Experts must synthesise research findings, ensure consistency in tone and structure, adapt content for different contexts, and maintain alignment with evolving clinical guidelines.

Recent advances in Generative AI offer new opportunities to support this process [2, 3]. However, most AI systems in healthcare focus on information retrieval or content generation without explicitly supporting reflective review, expert judgement, and coordinated decision-making. This limits their ability to function as mediating agents that can assist experts in making informed, context-sensitive decisions during content development.

HAI-Agency: Workshop on Orchestrating Human and AI Agency for Proactive and Reflective Learning, co-located with the 27th International Conference on Artificial Intelligence in Education (AIED 2026), Seoul, Korea, June 27 - July 3, 2026.

*Corresponding author.

✉ triet.bui@monash.edu (T. Bui); jiarong@acm.org (J. Rong); jue.xie@monash.edu (J. Xie)

ORCID 0009-0003-6140-4471 (T. Bui); 0000-0002-9462-3924 (J. Rong); 0000-0003-4549-1627 (J. Xie)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

To address this gap, this paper explores how Knowledge Graph-based **Retrieval-Augmented Generation (RAG)** can support the curation and adaptation of parenting intervention content. Our work focuses on enabling transparent, traceable, and human-steerable AI assistance, shifting from reactive tools toward more proactive, workflow-integrated systems while preserving human agency. By structuring domain knowledge into interconnected entities and relationships, **Knowledge Graphs (KGs)** enable more context-aware and inspectable retrieval [4]. However, rather than assuming a uniform benefit, we examine how different graph representations shape the trade-offs between traceability, contextual alignment, and generative flexibility. Our findings show that graph structure influences not only what information is retrieved, but how effectively it supports expert tasks such as verification, synthesis, and content adaptation. Building on these findings, we discuss implications for designing explainable, reliable, and human-centred AI systems that better support expert workflows in high-stakes, learning-oriented interventions.

2. Modelling Methodology

To support AI-mediated content curation and adaptation, we develop a Knowledge Graph-based RAG system that structures intervention knowledge for more transparent and context-sensitive use. Instead of treating knowledge graphs as purely a retrieval optimisation technique, we conceptualise them as representations of task context, evidence structure, and human-AI coordination needs. In this section, we outline the methodology for modelling psychoeducational information into two complementary graph forms—Chunk-Based and Proposition-Based, each embodying different assumptions about how information is organised, interpreted, and utilised in expert workflows.

2.1. Parent Intervention Content Used for Modelling

The resulting knowledge graphs were constructed from 241 existing intervention content pages derived from the original PiP program [1], an award-winning, web-based, self-directed health intervention designed to empower parents. Specifically, the selected content modules were drawn from the COVID-19 adaptation of the PiP program (PiP-Plus, <https://partnersinparenting.com.au/>), which supports adult caregivers in developing parenting skills to reduce the risk of adolescent depression and anxiety during the pandemic, and the PiP Kids Autism program (PiP-KA, <https://pka.partnersinparenting.com.au/>), which was designed to help parents of autistic children reduce their child's risk of anxiety and depression. The structured format of these programs—organising evidence-based strategies into clear thematic modules and sections—provides not only a rich evidence base but also an implicit representation of how knowledge is organised and communicated to end users. Additionally, preliminary analysis across existing PiP programs reveals recurring section patterns. This consistency highlights opportunities for an AI system built on top of these interventions to support the development of new interventions. The content is licensed under Creative Commons License (CC BY-SA 4.0), allowing for full utilisation without copyright concerns.

2.2. Knowledge Graph Construction Methods

A KG is defined as a graph $G = (V, E)$ consisting of a set of vertexes (or nodes) V and a set of edges E , representing entities and relationships derived from the knowledge sources, respectively. In this work, the sources are unstructured markdown files containing raw text, markdown syntax, and HTML tags. Extracting entities, relationships, and attributes from this data is a necessary step prior to graph construction. While both graph types share a common extraction pipeline, they differ in how they encode evidence structure and support human-AI interaction. These differences reflect alternative design choices about how context is preserved, how evidence is surfaced, and how experts can trace, interpret, and make decisions based on AI outputs.

2.2.1. Extraction Process

The shared pipeline begins with a data loader that extracts text from PDF files (converted from raw markdown files), followed by a text splitter that segments the content into 500-character chunks with an overlap of approximately 100 characters. Each chunk is then embedded into a 756-dimensional vector space. Guided by a predefined node and edge schema, we employ an LLM-based extraction approach using **Gemini-2.5-Flash**. During this process, entities are identified and assigned generic labels based on the context of the input text. Relationships are then extracted between the identified entities, linking source and target nodes, and are summarised using high-level keywords that capture broader concepts or themes rather than fine-grained details.

As documents are processed sequentially and independently, duplicate entities may emerge across instances. To better control duplication within the resulting knowledge graphs, we apply a rule-based entity deduplication stage. In this step, entities are merged into a single node when their names satisfy predefined similarity conditions. These conditions include: **Perfect Match Resolution**, which merges entities with identical keys, and **Semantic Similarity Resolution**, which measures semantic similarity using cosine similarity computed from vector representations. If the similarity exceeds a defined threshold, the entities are considered synonymous and merged accordingly. Finally, the extracted knowledge is passed to the respective graph builders, and the resulting structure is written to the graph database. The overall extraction and graph construction process is illustrated in Figure 1.

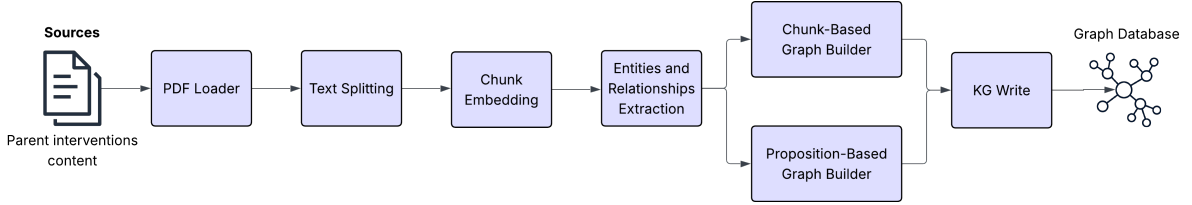


Figure 1: Shared extraction pipeline

2.2.2. Chunk-Based Graph

The Chunk-Based graph preserves the original structure and flow of intervention content, representing it as *Chunk* nodes that store the **exact text segments extracted from source documents**. *Document* nodes capture source-level context (e.g., filenames), while NEXT_CHUNK and FROM_DOCUMENT relationships maintain sequential and hierarchical organisation. This design directly supports tasks where experts need to trace outputs back to original materials or assess how specific information is framed within its broader context. Although chunking improves granularity, it can fragment related information; to mitigate this, *Entity* nodes are linked across chunks via semantic relations, constructing cross-cutting concepts. As such, this representation emphasises traceability, interpretability, and alignment with how content is experienced and reviewed by human experts.

2.2.3. Proposition-Based Graph

In contrast, the Proposition-Based graph restructures content into atomic units of meaning. Proposition statements capturing single facts—have been shown to improve retrieval effectiveness due to their high density of query-relevant information [5]. Furthermore, such a technique achieves competitive results in medical-based contextual summarisation and creative generation during comparison with alternative Graph-based methodologies[6]. Accordingly, this representation encodes knowledge as **propositions rather than raw text**, retaining only *Entity* nodes and their relationships.

Within this representation, both nodes and edges contain descriptive proposition fields that capture either entity attributes or the semantics of relationships. Each entity and relationship is represented as a key-value pair, where the *Key* corresponds to a concept or short phrase (e.g., activities, symptoms, or behaviours), while the *Value* consists of an LLM-generated summary of the relevant contextual

information derived during graph construction. To preserve traceability despite the removal of *Chunk* and *Document* nodes, each node and edge maintains aggregated references to its contributing source documents.

By abstracting information into propositions, this representation shifts the focus from document structure toward conceptual relationships. As a result, it better supports tasks requiring knowledge adaptation, while prioritising conceptual clarity and flexibility for cross-concept reasoning and generative tasks.

2.3. Evaluation Query Types and Retrieval Strategy

To support the following Exploratory Workshop, we employ two complementary query types: **Referencing queries** and **Creative queries**. **Referencing queries** reflect tasks that engage directly with existing PiP content, such as locating prior definitions or verifying how concepts were previously expressed, supporting consistency and traceability. In contrast, **Creative queries** probe higher-level reasoning and adaptation, requiring the system to generate new content grounded in PiP principles—for example, localising examples for different cultural contexts or reframing guidance for new audiences.

These query types correspond to different modes of human–AI coordination: one emphasising verification and alignment with existing evidence, and the other supporting interpretation, transformation, and design. Each query represents a plausible expert task, capturing situations where evidence, contextual understanding, and stakeholder needs must be negotiated. Representative queries and outputs used in the evaluation are included in the Appendices for reference.

To execute these tasks within the retrieval-generation pipeline, both graph representations follow a shared and lightweight retrieval procedure. Retrieval begins with embedding the input query and performing vector-based search over either document chunks (Chunk-Based Graph) or entity propositions (Proposition-Based Graph). The retrieved nodes then serve as seed nodes for a one-hop graph expansion process, allowing additional neighbouring entities and relationships to be incorporated as into the context. Generation is then performed using **Gemini-2.5-Flash**, which uses the retrieved context to answer the input queries. To isolate the effect of graph structure on retrieval quality and generation outcomes, the same generation prompt and model configuration were used across both graph representations.

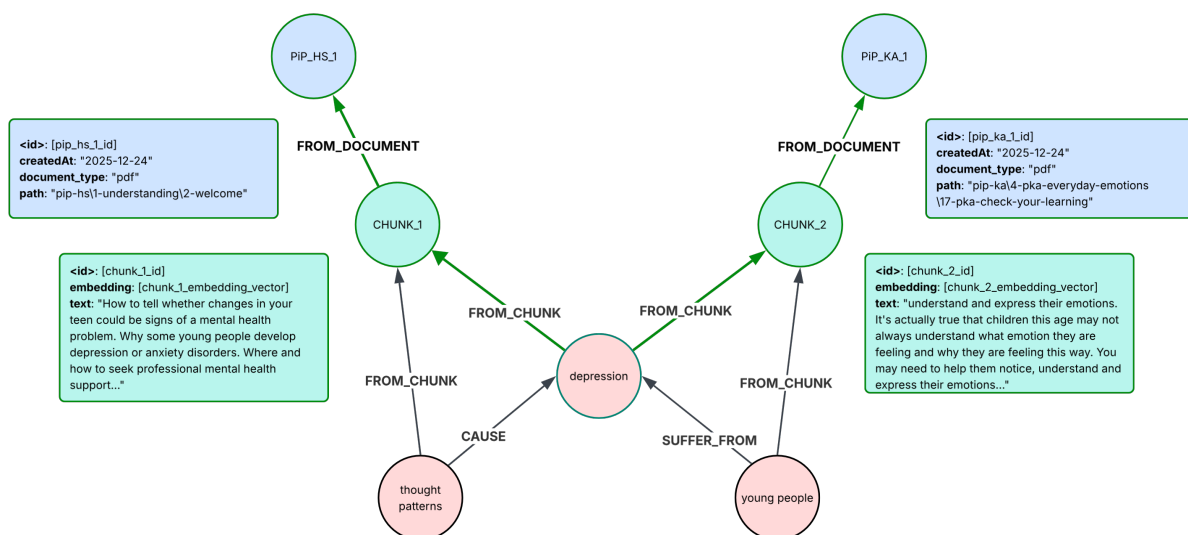


Figure 2: An subgraph of Chunk-Based Graph that allows for traversal between PiP-Plus and PiP-KA

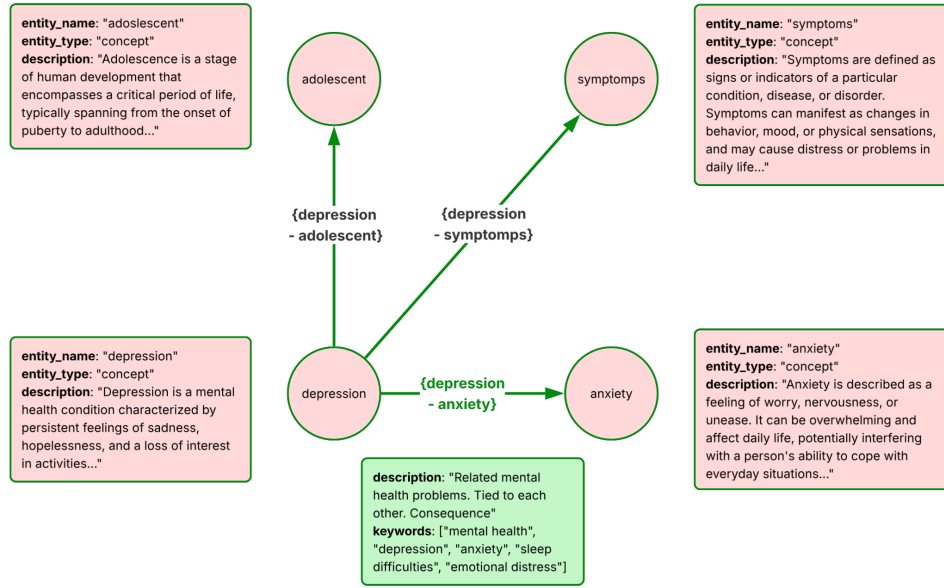


Figure 3: An sub-graph of Proposition-Based Graph that connects depression-oriented concepts

Algorithm 1 One-hop expansion retrieval

Require: $seedIds$ – identifiers of seed nodes in graph G

Ensure: $R, |R| \leq k$

- 1: $S \leftarrow$ nodes in G whose $id \in seedIds$ ▷ seed set
 - 2: $N \leftarrow$ neighbours of S , reachable by one hop, excluding S ▷ expanded neighbours
 - 3: $score(u) \leftarrow$ number of edges between $u \in N$ and any node in S ▷ connection strength
 - 4: **return** top- k nodes in N ordered by score descending, with their metadata
-

2.4. Constructed Graphs Complexity Evaluation

Considering entity-level structures, both constructed graphs are comparable in terms of nodes and edges count, with the Chunk-Based Graph containing 2555 entity nodes and 1520 entity relationships, compared to 2535 nodes and 2586 relationships in the Proposition-Based Graph. This suggests that both approaches extract a similar number of semantic entities, while the Proposition-Based Graph captures a denser set of conceptual relationships.

The Chunk-Based Graph exhibits a higher average degree than the Proposition-Based Graph (2.59 vs. 1.02), indicating broader connectivity between entities and potentially supporting wider contextual traversal. In contrast, the Proposition-Based Graph demonstrates a higher average clustering coefficient (0.075 vs. 0.042), suggesting tighter local conceptual groupings where neighbouring entities are more likely to also connect with one another [6, 7]. This behaviour aligns with the proposition-oriented abstraction process, where semantically related concepts are consolidated into focused factual relationships.

Overall, both graphs remain relatively sparse, as reflected by their low clustering coefficients and modest average degrees. Structurally, this suggests a graph topology in which entities tend to connect around broader central concepts rather than forming densely interconnected semantic communities.

3. Exploratory Workshop Methodology

A qualitative workshop was conducted with two psychoeducational experts (E1 and E2), each with over 10 years of experience developing PiP content. E1 focuses on the foundational PiP-Plus program, while E2 specialises in PiP adaptations for children with autism (PiP-KA), providing complementary

Table 1
Graph Complexity Metrics

Metric	Chunk-Based Graph	Proposition-Based Graph
Entity Nodes Count	2555	2535
Entity-to-entity Edges Count	1520	2586
Avg Degree	2.59	1.02
Avg Cluster Coefficient	0.042	0.075

perspectives on both core content development and contextual adaptation.

The workshop explored the potential integration of an AI system into content development workflows, particularly through the use of existing interventions as a knowledge base. The discussion focused on two areas: **(1) identifying challenges and opportunities to streamline content development**, and **(2) understanding how experts search for, evaluate, and utilise information when creating and adapting content**.

The session lasted 90 minutes and contained hands-on activities with associating in-depth discussion topics. The resulting data were analysed using thematic analysis following [8], with findings presented in Section 4.

3.1. Activity 1 – Workflow Exploratory Activity

An initial flowchart, developed based on the research team’s understanding, was used as a probe to obtain experts’ insights into their content development workflow through iterative feedback.

The flowchart was structured into three key phases: **Defining Audience** (identifying the target audience for a new module and the factors influencing these decisions), **Drafting** (developing content through synthesising information from multiple sources, including the reuse of existing PiP materials) and **Review and Refining** (iteratively revising content to ensure alignment with existing PiP modules and addressing challenges encountered during refinement). In each phase, a researcher facilitated discussions to assess how closely the proposed flowchart reflected actual practices, encouraging experts to validate or challenge the assumptions.

3.2. Activity 2 – Retrieval and Generation Evaluation Activity

This activity adopts a scenario-based approach in order to ground discussion in practical tasks, enabling experts to reflect on their behaviours in context rather than in abstraction [9]. The primary goal was to examine how experts formulate search queries, evaluate AI-generated outputs, and make decisions about their potential use within the workflow. To support this, 10 triplets of the form: (*Query – Chunk-Based Answer – Proposition-Based Answer*) were generated following the Retrieval and Generation Pipeline described in Section 2.3. Then these triplets were presented to experts for comparative evaluation.

The activity was structured around two key components:

- **Query Reflection:** Experts were first asked to evaluate the usefulness of the queries, including whether they reflected questions naturally asked during their workflow, what modifications they would make (e.g., level of specificity or contextual detail), and how they would typically formulate such queries.
- **Comparison and Evaluation of Generated Answers:** The experts were then asked to review and compare the two responses generated for each query. This stage aimed to uncover the criteria experts use to evaluate response quality. Discussions also explored the practical usefulness of the responses within real workflows (such as whether experts would use the outputs directly, adapt them as a drafting starting point, or avoid using them altogether). In addition, experts were encouraged to articulate what constitutes a “good” response, what factors influenced their preference between alternatives, and what would increase their confidence in adopting such AI-assisted tools in practice.

4. Exploratory Workshop Findings

4.1. Potential AI-support use cases

Experts' feedback revealed important gaps between the research team's initial assumptions and actual practice, while also highlighting opportunities for AI support. In practice, the workflow includes an additional stage between **Defining Audience** and **Drafting**, where emergent needs are identified through both empirical and literature-based research. Specifically, two primary sources inform this process: **co-design workshops with relevant groups** (parents, young people, and education professionals) and **evidence from literature reviews** (Delphi studies and meta-analyses). Together, they guide the identification of topics for new modules or content expansions.

Once topics are identified, experts refine goals through ongoing consultation with existing PiP content and its underlying evidence-base. This helps ensure alignment with established modules while supporting targeted adaptations rather than replacing core content. However, experts also actively identify gaps in the current evidence-base and supplement these through additional literature. As noted by E2, when available evidence is limited or inconclusive, the process becomes highly iterative, requiring repeated comparison between new findings and existing evidence to assess alignment. This back-and-forth highlights the challenge of reconciling multiple sources of evidence during content development.

Additionally, **Drafting** and **Review and Refining** processes are shaped by the experts' deep familiarity with existing PiP modules. As core contributors to the original content, experts are often able to efficiently recall and retrieve relevant materials. However, this reliance on experience and memory introduces limitations, as such familiarity is not easily transferable to newer team members, who may struggle to locate or reuse existing content effectively (*"For a new PhD student or a new research team member who doesn't know PiP so well, then it may not be so easy."* – E1).

Consistency checking also emerged as a critical but resource-intensive task. E2 described recurring activities such as ensuring inclusive language, maintaining alignment across modules, and avoiding unnecessary duplication of concepts. These checks require manual comparison across multiple sources and are *"quite easy to miss"*. More broadly, experts described drafting and review as one of the most time-consuming aspects of the workflow. As noted by E1, this stage can involve extensive collaboration with senior academics and stakeholders, often spanning several workshops and requiring tens of hours over a period of 6 to 12 months. Ensuring fidelity requires detailed verification at the sentence level against multiple sources, including the evidence-base, co-design findings, and existing modules. This emphasises not only the complexity of the task but also the high cognitive and temporal demands placed on experts.

These challenges point to several opportunities for RAG-based systems across different stages of the workflow. During **Drafting** and **Review and Refine**, such systems could support more targeted retrieval of relevant PiP content, enabling experts to efficiently locate and reuse existing materials to maintain cross-module consistency. In the **Co-design** stage, RAG-based approaches could further assist in coordinating between existing evidence and new information, helping to surface gaps across materials and support more informed content development. .

4.2. Trust factors in AI usage

According to the experts, AI is not currently used in PiP content development due to concerns about hallucinations and unreliable outputs. Despite these reservations, the experts expressed interest in the deliberate use of AI systems to support and streamline the more tedious and manually intensive aspects of the workflow (discussed in Section 4.1). However, their willingness to adopt such systems depends on establishing sufficient trust and confidence across three key dimensions:

1. **Human oversight** retained as a core component of quality assurance, as highlighted by E1: *"But there's still one of us on the end, you know, confirming whatever it generates. Because I wouldn't be trusting it totally"*.

2. **Traceability within of AI-generated content** including clear references to source materials as well as integration of only reliable sources into the knowledge base.
3. **A protected environment** to safeguard unpublished content and intellectual property, alongside strict control over data inputs and outputs ("*We want to make sure everything stays in a protected environment and not fit into any kind of machine learning process outside the system.*" - E2)

4.3. Evaluating KG Representations for Referencing and Creative Tasks

The comparison between Chunk-Based and Proposition-Based graph representations reveals distinct strengths and limitations across different types of tasks.

For **referencing queries**, the Chunk-Based approach was consistently preferred. Experts found its responses more accurate and closely aligned with existing PiP content, largely due to its preservation of original text and contextual framing. As a result, its outputs were considered a strong and reliable starting point for further expansion. In contrast, the Proposition-Based approach often produced more general or, at times, inaccurate outputs, such as failing to capture key concepts. In one instance, it incorrectly suggested that certain concepts (e.g., child emotional distress) were not present in PiP, despite expert confirmation that they are. However, the Chunk-Based approach was not without limitations: in some cases, its responses were perceived as incomplete or insufficiently synthesised, requiring prior familiarity with the content to be fully understood.

For **creative queries**, overall performance was weaker across both methods, emphasising the challenges of supporting content adaptation tasks. In some cases, both approaches generated outputs that were perceived as overly "textbook-like" and not reflective of how parents or young people would realistically communicate. Even when responses were technically correct, issues with tone and naturalness reduced their practical usefulness.

Interestingly, the Proposition-Based approach demonstrated advantages in specific creative scenarios. In one case, the query asked the model to generate a scenario showing how a parent might respond supportively to a teenager's anxiety, its generated output was described as more "*relatable and realistic*", with experts noting it could be used with only minor revisions. This suggests that abstracting away from source text can, in some contexts, enable more flexible and natural content generation. In contrast, the Chunk-Based approach in this scenario retained more technical and instructional language (e.g., "minimise," "validate," "identify"), which, while appropriate for explanation, was less suited to scenario-based generation. This appears to stem from an overextension of PiP's formal style, resulting in outputs that felt rigid or overly prescriptive. This suggests that directly inheriting source content without nuanced adaptation can be problematic. However, these behaviours may not be solely attributable to graph structure or retrieval representation. The generation prompt itself may also influence how strongly the model adheres to formal instructional language versus adapting toward more natural or context-sensitive phrasing. Example queries and corresponding outputs illustrating these differences are provided in the Appendices.

5. Discussion

This work contributes to the design of AI mediators that support alignment and shared decision-making in collaborative healthcare. We examine how knowledge graph-enhanced retrieval-augmented generation can support the co-creation of psychoeducational interventions through transparent, traceable, and human-steerable AI assistance. Rather than acting as an autonomous generator, the AI functions as a mediator within expert workflows, supporting verification, synthesis, and adaptation while preserving human judgement. Findings from participatory workshops show that knowledge graph design influences not only retrieval performance, but also alignment with real-world practices and trust requirements.

From these findings, we derive three key design implications for AI-mediated systems in high-stakes psychoeducational workflows. Through ethics-approved participatory workshops with domain experts, we identified practical use cases for AI-supported systems, requirements for fostering trust between

experts and AI, and considerations for evaluating KG-based generation across different task contexts. Three key themes emerged from these findings, each informing the design of AI-mediated systems in parenting intervention settings:

Effective query design to assess RAG-based systems: Evaluating RAG-based systems in this domain must reflect the complexity of real-world tasks, where information needs are tightly coupled with downstream use (e.g., adaptation, verification, or synthesis). Consequently, the types of queries used directly influence which aspects of system performance are revealed. The findings suggest that effective queries are typically specific and context-rich, clearly defining the target population, context, and intended task rather than remaining broad or generic. The workshops also highlighted challenges in navigating multiple co-existing PiP programs and versions, where maintaining awareness of all available content becomes difficult. As a result, queries requiring synthesis across different PiP domains were viewed as particularly valuable. Evaluating RAG systems on such tasks can better reveal their ability to function as a unified knowledge base that supports cross-module reasoning and more efficient coordination of information during intervention development.

For example, queries such as *“I need to find specific content that may need to be adapted for teenagers who are neurodiverse (e.g., autism and ADHD) across all ‘school refusal’-related modules in PiP”* or *“Can you find all sections of PiP that relate to an autistic student in a school environment and identify where adaptations may be needed?”* were identified as particularly useful. These types of queries support targeted retrieval of relevant sections, which can then serve as a basis for adapting or developing new content.

Evaluation criteria for generated content: While effective query design enables more realistic, task-centric evaluation, it must be complemented by clear criteria for assessing output quality. Based on workshop insights, we identified a set of implicit criteria commonly used to assess generated responses. These criteria reflect not only the correctness of information but also its usability within real workflows. Formalising them provides a foundation for more rigorous evaluation frameworks and heuristics, enabling more meaningful experiments to assess and improve the usability of AI-mediated information retrieval and generation systems.

- **Traceability:** This emerged as the most critical criterion. Because new content is developed through the synthesis of multiple information sources (e.g. existing PiP materials, co-design findings, trial evidence, and broader literature), it is essential to understand how different sources contribute to specific parts of the generated output. Evaluation should therefore consider whether the system correctly attributes information and allows experts to inspect its origins. More broadly, traceability contributes directly to user control and trust in AI-generated content, an aspect discussed further in the next section.
- **Factual Accuracy:** Beyond citation, evaluation also considers whether the system correctly synthesises information. This includes maintaining canonical distinctions—such as differences between modules, target populations (e.g. children vs. teenagers), and conceptual definitions. Errors in synthesis, even when sources are correctly retrieved, can significantly reduce the reliability of the output.
- **Relevancy:** Generated responses do not always fully or directly address the query. While iterative prompting can sometimes improve results, this places an additional burden on users and reduces overall system usability. Evaluation should therefore assess how well outputs align with the intent and scope of the query, while also considering the role of retrieval mechanisms in shaping context.
- **Tone alignment:** PiP content is designed to empower parents and therefore requires careful attention to tone and language. Outputs should be supportive, inclusive, and sensitive to diverse cultural and social contexts. In particular, scenario-based or conversational content should feel natural and relatable, rather than overly technical. While some heuristics exist, this highlights the need for more explicit guidelines and evaluation frameworks for tone and communication style.

Designing for Human Control in AI-Mediated Systems: A central implication of this work is that explainability alone is insufficient. AI systems must be designed to support meaningful human

control across the workflow. AI is consistently positioned not as an autonomous decision-maker, but as a collaborator—*“an agent to do the work, but we direct it”*. This reflects a broader perspective on human–AI orchestration in this domain: rather than replacing human judgement, AI should support it by making processes more efficient, transparent, and controllable under human direction. Importantly, such control is not singular, but spans multiple dimensions.

First, beyond relying solely on existing PiP content, the system could support dynamic control over which sources inform specific queries. These sources may include summaries of research, workshop notes, or other context-specific materials that are not formally embedded in the core intervention content. This suggests a design direction in which users can explicitly guide the system, for example: *“follow the principles in existing PiP modules, but prioritise this document and apply this specific strategy”*. Enabling this level of control increases transparency and helps address a key limitation observed during evaluation—namely, the system’s inability to capture implicit definitions or distinctions not explicitly stated in the source material. For instance, differences between PiP-KA (targeting children aged 5–11) and the PiP-Plus program for adolescents (aged 12–17) were not reliably surfaced, indicating the need for more extensible integration of knowledge sources.

Second, beyond controlling inputs, the structure of the knowledge itself plays a critical role in supporting informed decision-making. From a knowledge representation perspective, hierarchical relationships exist within PiP concepts under the broader domain of mental health. While the current work incorporates a rudimentary document-based hierarchy—particularly in the Chunk-Based graph—this could be extended to explicitly model conceptual hierarchies. Doing so would allow users to better trace relationships between concepts, navigate content more intuitively, and identify opportunities for reuse or adaptation across contexts. Exposing these structures through interactive interfaces could further enhance usability. Prior work on GraphRAG systems has explored visual and graph-based interfaces to support such interactions, demonstrating improved interpretability and transparency by making relationships between entities and evidence more explicit [10, 11].

Together, these findings suggest that the design of AI-supported systems in this domain must extend beyond improvements in retrieval or generation performance. Instead, they should be conceptualised as collaborative infrastructures that enable meaningful human control, support flexible integration of reliable and context-specific knowledge sources, and align closely with the structured yet evolving nature of real-world workflows. This reframing is critical for supporting trust, accountability, and sustained adoption in safety-critical contexts.

6. Conclusion

This paper examined how knowledge graph design shapes explainable AI mediation in the co-creation of psychoeducational interventions. Framing intervention development as a high-stakes, learning-oriented design process, we showed how knowledge graph-enhanced retrieval-augmented generation can support expert workflows while preserving human agency. Comparing Chunk-Based and Proposition-Based representations revealed key trade-offs between traceability, contextual alignment, and generative flexibility. Our findings position AI not as an autonomous generator but as a mediating component that supports verification, synthesis, and adaptation of evidence. In particular, traceability, task-aligned query design, and mechanisms for human control emerged as critical for trustworthy use. More broadly, this work highlights the importance of designing human-centred AI systems that align with real-world workflows, where knowledge graphs enable both interpretability and flexible use in safety-critical contexts.

References

- [1] M. B. Yap, K. A. Lawrence, R. M. Rapee, M. C. Cardamone-Breen, J. Green, A. F. Jorm, Partners in parenting: A multi-level web-based approach to support parents in prevention and early intervention for adolescent depression and anxiety, *JMIR Mental Health* 4 (2017) e59. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5750418/>. doi:10.2196/MENTAL.8492.
- [2] P. Zhou, Z. Zhang, F. Yang, Knowledge graphs towards ai-assisted smart education, *Proceedings of The 2nd International Conference on Intelligent Education and Computer Technology, IECT 2025* (2025) 866–871. URL: [/doi/pdf/10.1145/3764206.3764342?download=true](https://doi.org/10.1145/3764206.3764342?download=true). doi:10.1145/3764206.3764342.
- [3] Q. Wang, F. Yang, L. Quan, M. Fu, Z. Yang, J. Wang, Knowledge graph and its application in the study of neurological and mental disorders, *Frontiers in Psychiatry* 16 (2025) 1452557. doi:10.3389/fpsyt.2025.1452557.
- [4] Z. Guo, L. Xia, Y. Yu, T. Ao, C. Huang, Lightrag: Simple and fast retrieval-augmented generation (2024). URL: <https://arxiv.org/pdf/2410.05779>.
- [5] T. Chen, H. Wang, S. Chen, W. Yu, K. Ma, X. Zhao, H. Zhang, D. Yu, Dense x retrieval: What retrieval granularity should we use?, *EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* (2023) 15159–15177. URL: <https://arxiv.org/pdf/2312.06648>. doi:10.18653/v1/2024.emnlp-main.845.
- [6] Z. Xiang, C. Wu, Q. Zhang, S. Chen, Z. Hong, X. Huang, J. Su, When to use graphs in rag: A comprehensive analysis for graph retrieval-augmented generation (2025). URL: <https://arxiv.org/pdf/2506.05690v1>.
- [7] A. I. Nesterov, On clustering coefficients in complex networks, 2024. URL: <http://arxiv.org/abs/2401.02999>.
- [8] V. Braun, V. Clarke, Using thematic analysis in psychology, *Qualitative Research in Psychology* 3 (2006) 77–101. URL: <https://www.tandfonline.com/doi/abs/10.1191/1478088706qp0630a>. doi:10.1191/1478088706QP0630A.
- [9] M. B. Rosson, J. M. Carroll, V. Tech, B. Va, *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (2002) 1032–1050.
- [10] T. Wang, J. He, C. Xiong, Ragviz: Diagnose and visualize retrieval-augmented generation, *EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of System Demonstrations* (2024) 320–327. doi:10.18653/v1/2024.EMNLP-DEMO.33.
- [11] V. Firsanova, Y. Khlusova, Aggile: Automated graph generation for inference and language exploration, in: *CEUR Workshop Proceedings*, 2025. URL: <https://www.scopus.com/pages/publications/105008498345>.

7. Appendices

Table A.1

Reference and Creative Task Queries

Query's Type	Query
Reference Task Query	What topics are covered in PiP that relate to anxiety in teenagers?
	How does PiP define emotional regulation in adolescents?
	Are there common principles across PiP modules about how parents should respond to children's emotional distress?
	Does PiP provide any frameworks or structured approaches for helping parents support teenagers experiencing anxiety?
	How does PiP typically describe and explain emotional distress in teenagers across different modules?
Creative Task Query	Generate a short scenario showing how a parent might respond supportively when a teenager expresses anxiety about school.
	Create a short example showing the difference between non-effective and effective parental responses to autistic children's communication difficulties and frustration.
	Based on concepts introduced in PiP, generate three practical tips for parents to help teenagers cope with exam-related stress.
	This is a conversation depicting a conflict between parent-teen: Parent: "I think you need to spend at least two more hours studying every night if you want to improve your grades." Teen: "Studying more hours isn't going to help if I'm already exhausted." Parent: "But your grades are slipping, so you need to put in more effort." Teen: "I am putting in effort. I just think you forcing more study time is just making me more stressed." Adapt this example so it is culturally appropriate for new migrant family from East Asia in Australia where open disagreement between parent-child is less common.
	The following PiP content refers to support mental health resources for parents. Replace these with equivalent services appropriate for a Canada context and explain differences for access: "1. If you would like to find a registered psychologist in your area, the Australian Psychological Society Find a Psychologist (https://psychology.org.au/find-a-psychologist) service can assist with this. You can also speak to your GP about getting a referral to a psychologist."

Table A.2

Query 3 and generated answers

Query 3 :Are there common principles across PiP modules about how parents should respond to children's emotional distress?	
Chunk-Based Graph	Proposition-Based Graph
Yes, there are common principles across PiP modules regarding how parents should respond to children's emotional distress. These principles emphasize support, understanding, and teaching coping strategies. Specifically, modules suggest parents should: • Supporting the child when something is bothering them. This involves picking up on a child's emotions while they are mild and talking about them, rather than ignoring or pushing them aside, which can prolong distress and hinder learning. • Help children manage their emotions. This includes teaching strategies to calm down when experiencing anxiety and helping them express their emotions better. • Help children feel a sense of control while also helping them be okay with uncertainty . • Show patience and empathy during difficult emotional times. • Avoid burdening the child with the parent's own problems . • Take care of themselves when experiencing challenging feelings related to their child's behavior or emotions, suggesting reaching out for support or talking to a PiP coach . • The modules also highlight the importance of understanding a child's cues for affection and responding accordingly, and showing interest in the child's experiences and opinions.	The provided context does not contain information about common principles across PiP modules regarding how parents should respond to children's emotional distress. However, it does indicate that parenting programs, such as the PiP program , can be used for the prevention of teen depression and anxiety . Additionally, a supportive family environment is important for nurturing and supporting a child, and children can experience emotional problems . Recognizing and understanding emotions is also a key aspect of a child's development. One module, "Manage Tough Emotions," discusses thinking patterns .

Table A.3

Query 6 and generated answers

Query 6 : Generate a short scenario showing how a parent might respond supportively when a teenager expresses anxiety about school.	
Chunk-Based Graph	Proposition-Based Graph
Teenager (Alex): "Mom, I'm so stressed about my math test tomorrow. I just know I'm going to fail." Parent (Maria): (Calmly, sitting beside Alex) "I can see you're feeling really anxious about this math test, Alex. It's completely understandable to feel that way when you have something challenging coming up; we all feel anxious at times. I remember feeling that kind of worry too." (Maria acknowledges Alex's anxiety, demonstrating acceptance). "But I also know how much effort you've put into studying this week. You're a smart kid, and I have confidence in your ability to tackle this. We can go over a few more practice problems together if you want, or just talk through any specific parts that are making you feel overwhelmed. We're not going to minimize how you're feeling, because I know it feels very real to you right now".	<i>A teenager, Alex, walks into the living room, dropping their backpack with a sigh.</i> <i>Their parent, Sarah, notices Alex's slumped shoulders and worried expression.</i> Sarah: "Hey, you look a bit down. Is everything okay?" Alex: "Not really. I have a huge project due next week, and a test in history, and I just feel so overwhelmed. I don't think I can do it all." Sarah: (Sitting down next to Alex, making eye contact) "It sounds like you're feeling a lot of anxiety about school right now, and that's a really tough feeling to carry. Thanks for telling me what's going on". "I hear how stressed you are. Can you tell me more about what specifically feels overwhelming? I'm here to listen without interrupting, and we can try to understand it together". Alex: "It's just... everything. What if I mess up the project? What if I fail the test?" Sarah: "Those are really valid worries, and it's okay to feel that way. We can talk through what's making you feel this way, and maybe we can break it down into smaller steps, or just talk about how you're feeling".

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o for grammar and spelling checks, as well as for paraphrasing and rewording portions of the text. All generated outputs were subsequently reviewed and edited by the authors as necessary. The authors take full responsibility for the content of this publication.