



The Tenth International Conference on Informatics and Assistive Technologies
for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2025)



School Health Dialogue A Prompt-Expansion and Response-Visualization Framework

Hayato Tomisu^{1, 2}, Kazue Yamamura³,
Junya Ueda⁴, Tsukasa Yamanaka¹

¹Ritsumeikan University ²Shiga University ³Ritsumeikan Moriyama Junior and Senior High School ⁴ImpactLab
tomisu@fc.ritsumei.ac.jp (Hayato Tomisu)



Hayato Tomisu

from
Shiga, Japan

Affiliation

Ritsumeikan University

Ritsumeikan Global Innovation Research Organization

Visiting Researcher

Shiga University

Graduate School of Data Science

Ph.D. Student

Research Topic



Data Science



IoT



Education



Digital Twin

Our research group project

3

Mental health counseling sessions increase during COVID-19.



2021's

Creation of a Digital Health Room on the Metaverse

2023's



2025's

2024's

Expansion of the Health Room, ICT Utilization

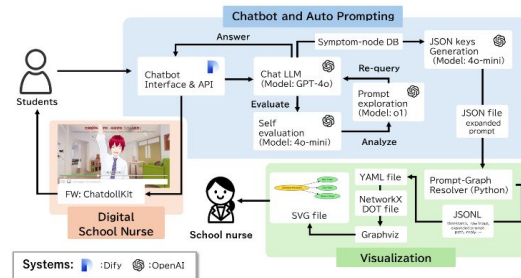


Creation of a Digital School Nurse Using Generative AI



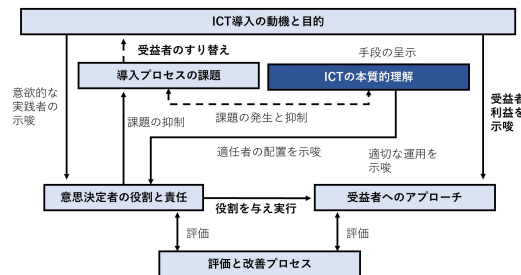
More personal conversations with AI avatars and transparent visualization of conversation contexts

1



Modeling Methods for Establishing ICT in Educational Settings

2



1. Background
2. Proposed Method
3. Evaluation and Results
4. Conclusion

Background

+52.6%

Increase in infirmary visits
(2019-2021)

[Yamamura, 2023]



Key Problem

- Adolescents express symptoms vaguely
- High cognitive burden on school nurses
- Rapid severity decisions from incomplete information

I don't feel
good...



I'm actually being
bullied.

Prompt extension techniques are being explored as a derivative of Chain-of-Thought. [Zhang et al., 2023; Zhou et al., 2023; Wei et al., 2022; Wang et al., 2023]

Research is being conducted on techniques that use graphs to visualize reasoning paths, thereby organizing thoughts in an easily understandable manner. [Li et al., 2025; Cao, 2024; Yao et al., 2023]

Gap: Existing work evaluated mainly on synthetic Q&A or programming tasks

- No work targets the ambiguity of adolescent health communication
- No integration of safeguarding needs and workflow constraints



【Objective】

Develop an automated AI chatbot that **transforms ambiguous student utterances into structured, explainable dialogue flows** while reducing cognitive burden on school nurses.



Minimal Cognitive Load

Students use brief, natural language



Human Transparency

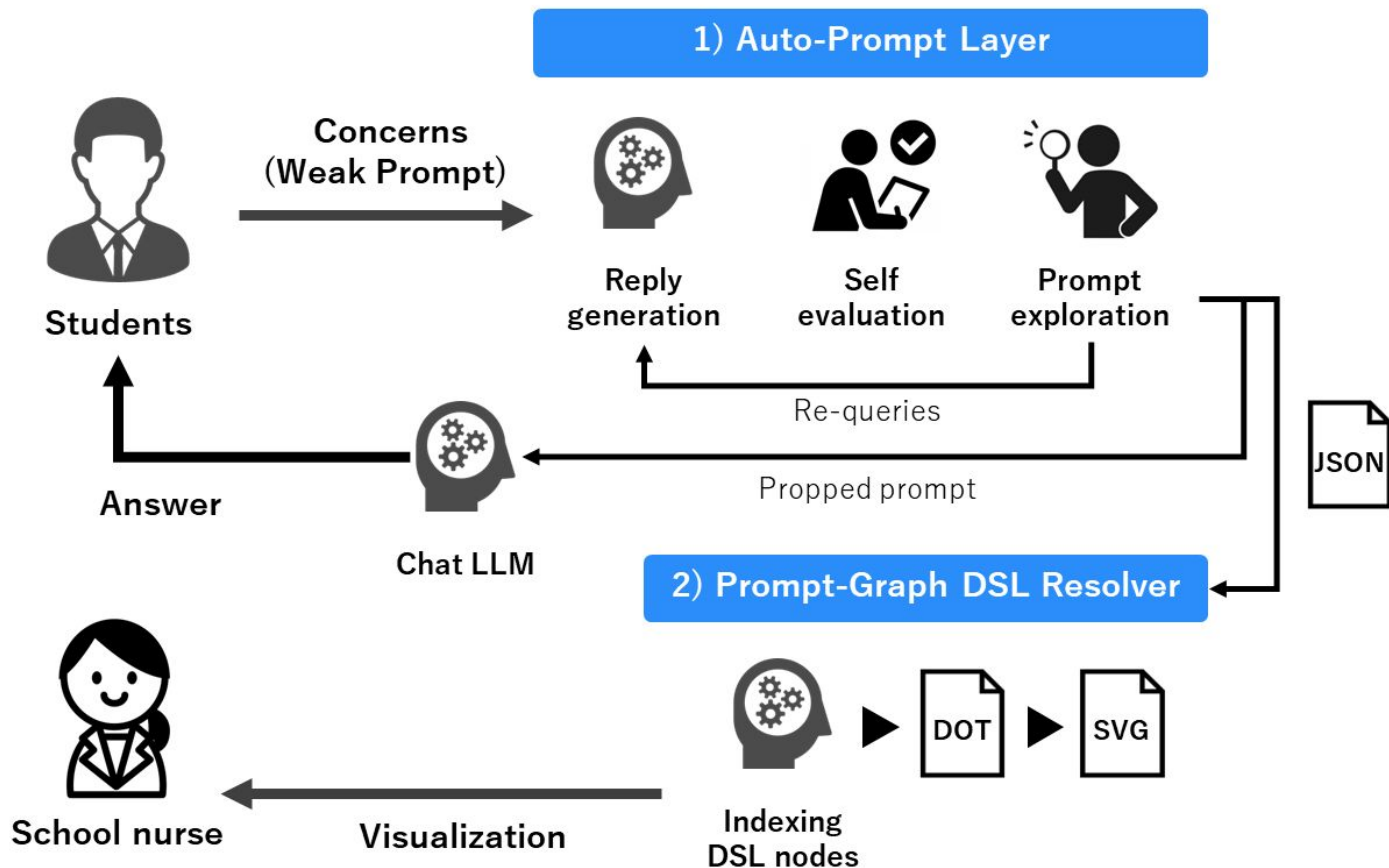
Nurses understand reasoning

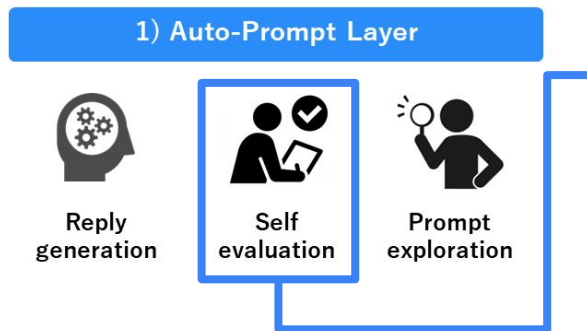


Auditability

Reproducible reasoning steps

Proposed Mehod





Self Evaluation Scoring

$$\text{Score} = w_c C + w_e E + w_s S$$

C (Coverage): Fraction of filled slots

$$C := \text{coverage}(J) = \frac{\# \text{ filled slots}}{4}$$

State $s = \langle u, j \rangle$: Current prompt u + slot vector j

E (Empathy): Cosine similarity to empathy prototype vector

S (Safety): Binary check: 1 if safe, 0 if unsafe

Acceptance criterion: $\text{Score} \geq \tau$ (threshold)

Original Student Utterance: "My head feels heavy and a little queasy"

↓ Expansion

First Pass - Slot Extraction:

symptom: headache ✓	intensity: mild ✓	coSymptom: nausea ✓	duration: ∅ ✗
------------------------	----------------------	------------------------	------------------

Coverage $C = 3/4 = 0.75$

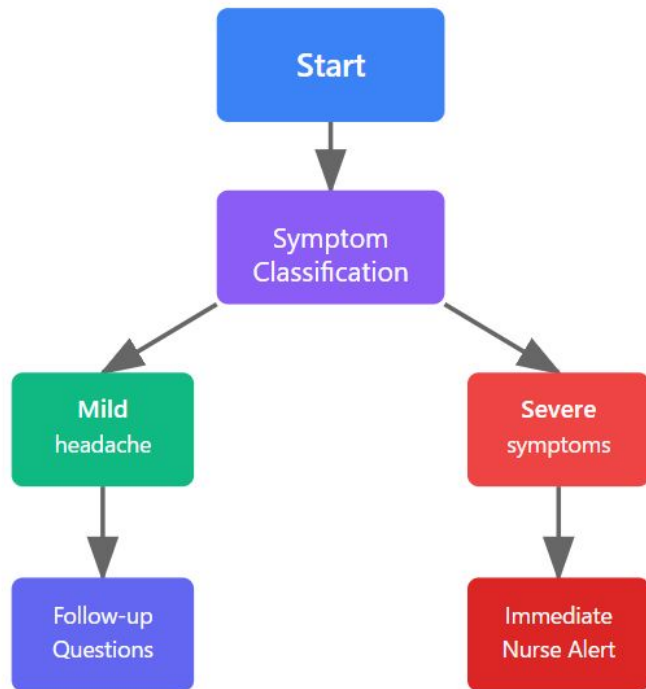
↓ Apply APPEND_ASK and Compete slots

System asks clarification: "How long have you been feeling this way?"

Complete Slot Vector J*: symptom: headache, intensity: mild, duration: 15min, coSymptom: nausea

→ Ready for Layer 2 mapping

Deterministic Decision Graph



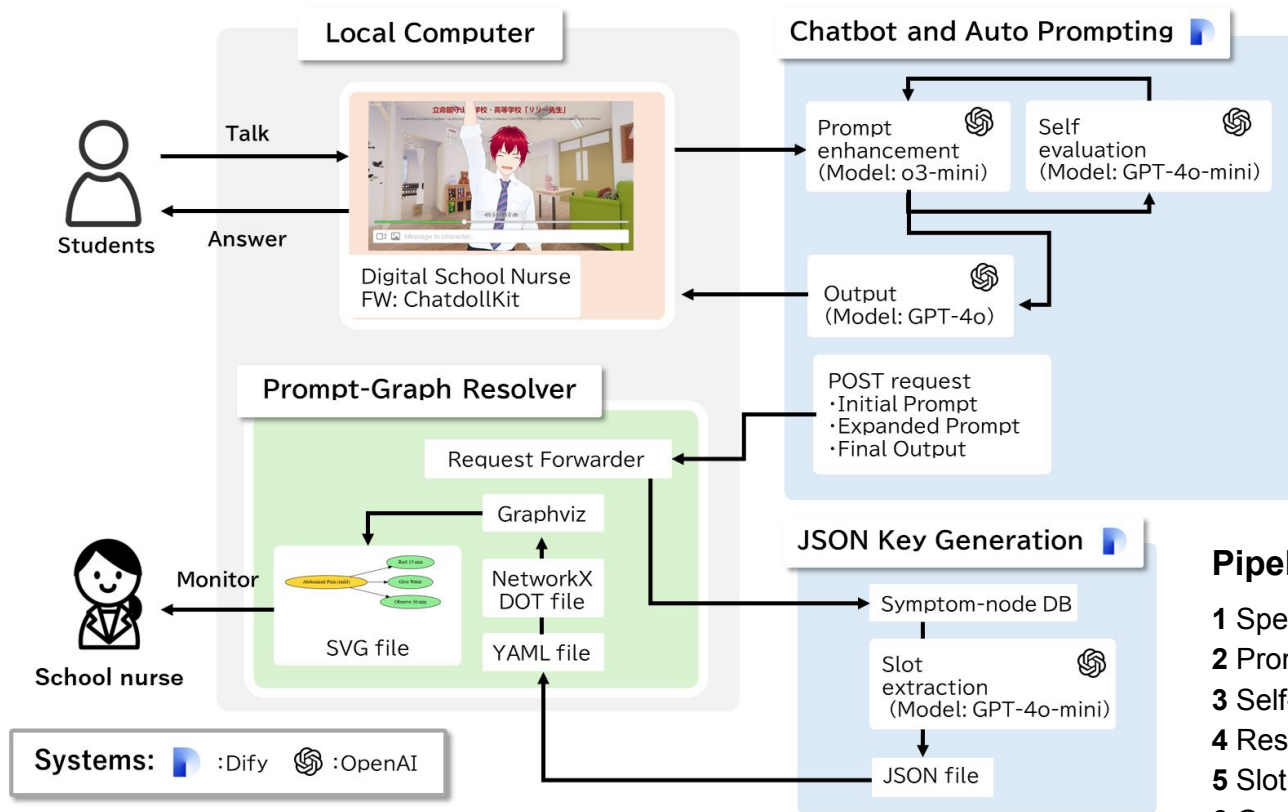
Key Features: Slot vector J^* $\rightarrow$ unique path P

Breadth-first predicate evaluation

Visual Subgraph GP: Updates at every turn Shows current node + 2-hop neighbors

Just looking at the AI's response takes time, but seeing this makes it clear at a glance what's wrong and what should be done!





Pipeline (7 steps):

- 1 Speech → Text
- 2 Prompt enhancement (o3-mini)
- 3 Self-evaluation (GPT-4o-mini)
- 4 Response generation (GPT-4o)
- 5 Slot extraction → JSON
- 6 Graph rendering → SVG
- 7 Avatar speaks + nurse dashboard updates

Evaluation and Results

Evaluation 1: Prompt Expansion Quality

Data: 50 Student Complaints

- 20 complaints from school nurse handbooks

- 30 ambiguous examples mined from social media and Q&A platforms

Procedure: Cosine similarity using Sentence-BERT (all-mpnet-base-v2)

Evaluation 2: Answer Quality Assessment

Data: 50 Generated Responses

Procedure: Mean scores computed for each dimension across all 50 responses compared by exemplary nurse responses created by domain expert

Evaluation 3: Slot Extraction Performance

Data: 50 Slot Extractions (200 total slots)

Procedure: Exact string match per slot from generated sentences by confusion matrix

Auto-prompt expansion achieved strong performance, effectively capturing student intent with no hallucinations.

0.82

F1 Score

41

True Positives

9

False Negatives

No False Positives below threshold

Key Findings



High F1 Score (0.82)

Effectively captures student intent



No Hallucinations

No fabricated values below similarity threshold



Safe Failure Mode

Errors from partial omissions, not fabrication



Error Pattern

FN primarily from missing contextual details

The system demonstrated consistently empathetic tone, appropriate for adolescent communication. However, results reveal opportunities for improvement in medical specificity and follow-up guidance.

3.63

/ 5.00

Accuracy

3.71

/ 5.00

Completeness

4.73

/ 5.00

Tone

✓ Strengths

- High tone score (4.73) confirms consistently empathetic responses
- Appropriate for adolescent communication
- Natural, supportive language maintained

⚠ Areas for Improvement

- Lower accuracy (3.63) reveals gaps in medical specificity
- Completeness (3.71) indicates missing follow-up guidance
- Need better clinical depth in responses

Error analysis revealed that lexical variation in student expressions requires controlled vocabularies and synonym normalization for improved performance.

0.43

F1 Score

39

True Pos.

56

False Pos.

47

False Neg.

0.41

Precision

0.45

Recall

Error Analysis

1. Lexical Variation

Students use different words for same concepts

Example: "really bad" ≠ "severe" ≠ "intense"

2. Severity Mapping

Inconsistent mapping to 3-level scale (mild/moderate/severe)

3. Location Terms

High variation in describing body parts and locations

Conclusion

Achieved by This Research

Key Contributions

Layer 1: Auto-Prompt Expansion

Enriches vague inputs (F1 = 0.82)

Layer 2: Prompt-Graph DSL

Transparent visualization for nurses

- Reduces cognitive burden on school nurses
- Maintains human oversight

Future Work

- Controlled vocabularies & synonym normalization
- Post-editing rules & supervised adapters
- Long-term field validation & health outcomes