# 📄 context.md — Knowledge Representation & Reasoning Project

## Project Title

# Ontology-Driven Mental Health Knowledge Graph with Explainable Causal Reasoning

---

## 1. Project Overview

University students increasingly experience emotional and mental-health challenges such as academic stress, anxiety, burnout, sleep disturbance, and social withdrawal. While many AI systems attempt to address these issues using black-box machine learning models, they often lack **transparent reasoning**, **causal explanations**, and **ethical safeguards**.

This project develops a **Knowledge Representation & Reasoning (KRR)–based system** that uses a **Mental Health Ontology and Knowledge Graph (KG)** to support **explainable, non-clinical emotional wellness assistance**.

The system combines:

- **OWL ontologies** for formal knowledge representation
- **RDF knowledge graphs** for structured facts
- **SWRL rules** for explicit symbolic reasoning
- **SPARQL rules** for graph materialization and querying
- **Natural Language Processing (NLP)** for extracting emotions, symptoms, and triggers from user messages
- **Causal explanation generation (WHY-answers)**
- **Safety-first escalation policies**

The final system is a **context-aware conversational assistant** that reasons symbolically over accumulated user evidence and explains *why* certain risk patterns or interventions are suggested.

The system runs **locally** and is intended **strictly for educational and well-being support**, not diagnosis or therapy.

---

# 2. Motivation

Students frequently express concerns such as:

- insomnia
- persistent stress
- academic pressure
- anxiety before exams
- difficulty concentrating
- social withdrawal
- irritability
- feelings of hopelessness

These signals often appear gradually and across multiple interactions. Traditional chatbots treat messages independently and provide surface-level responses.

A **Knowledge Graph + Reasoning Engine** enables:

- structured representation of emotions, symptoms, triggers, and mental states
- explicit modeling of cause–effect relationships
- cumulative reasoning over time (continuous context)
- transparent and inspectable inference
- explainable, ethical decision support

This directly aligns with the core goals of the **Knowledge Representation & Reasoning (KRR)** course.

---

# 3. Project Goals

### ✔ Design a formal Mental Health Ontology

### ✔ Build a Knowledge Graph capturing:

- emotions
- symptoms
- triggers
- mental-health–related risk states
- interventions

### ✔ Implement symbolic reasoning using:

- SWRL rules (primary inference)
- SPARQL rules (graph updates and explanations)

### ✔ Build a continuous, context-aware conversational agent

### ✔ Provide causal "WHY" explanations for system outputs

### ✔ Expose confidence and uncertainty transparently

### ✔ Enforce strict safety and escalation policies

---

### Explicit Non-Goals

❌ No clinical diagnosis
❌ No therapy or medical advice
❌ No autonomous decision-making in high-risk cases

---

## 4. Technologies & Tools

### Ontology & Knowledge Graph

- OWL
- RDF / Turtle
- Protégé

### Reasoning

- SWRL (explicit causal and logical rules)
- OWL reasoner (Pellet or HermiT)
- SPARQL (CONSTRUCT / INSERT for rule materialization and querying)

> Note: SPARQL is used to *apply* rule outcomes and retrieve explanations, not as a standalone reasoning engine.

### Backend

- Python
- RDFLib
- Owlready2
- FastAPI (local API)

### NLP

- NLTK (tokenization, preprocessing)
- scikit-learn (lightweight keyword or pattern extraction)
- sentence-transformers (semantic similarity for concept mapping)

### Frontend (Optional)

- React-based chat interface (already implemented)
- UI serves purely as an interaction layer and does not influence reasoning

Note: Frontend and initial backend infrastructure are fully implemented and are not part of the project's research contribution.

---

## 5. System Architecture Overview

```
User Message
   ↓
NLP Extraction Layer
- emotions
- symptoms
- triggers
- extraction confidence
   ↓
Ontology Concept Mapping
   ↓
Session Knowledge Graph (temporary)
   ↓
SWRL Reasoning Engine
   ↓
Inferred Risk States / Patterns
   ↓
Causal Explanation Generator
   ↓
Confidence & Uncertainty Estimator
```

```
    ↓
Response + Safety Check
```

Each user session maintains its **own temporary subgraph**, enabling continuous reasoning across messages.

---

# 6. Ontology Design

## Core Classes

- Emotion
- Symptom
- Trigger
- MentalState
- RiskLevel
- Intervention
- BehaviorPattern

## Example Subclasses

**Emotion**

- Stress
- Anxiety
- Sadness
- Fear
- Irritability

**Symptom**

- Insomnia
- Fatigue
- RapidHeartRate
- BreathDifficulty
- SocialWithdrawal

**Trigger**

- ExamPressure
- AcademicWorkload
- FamilyPressure
- SocialExposure

**MentalState (Risk Patterns)**

- AcademicStress
- AnxietyRisk
- BurnoutRisk
- PanicRisk
- DepressiveSpectrum

**Intervention**

- BreathingExercise
- GroundingTechnique
- Journaling
- TimeManagement
- TalkToCounselor

---

### Core Object Properties

- hasSymptom
- experiencesEmotion
- triggeredBy
- persistsFor
- leadsTo
- increasesRiskOf
- recommendedIntervention

---

## 7. Knowledge Graph Representation

Example RDF triples:

```
:User1 :hasSymptom :Insomnia
:User1 :experiencesEmotion :Stress
:User1 :triggeredBy :ExamPressure
:Insomnia :associatedWith :Anxiety
:AcademicStress :canLeadTo :Anxiety
:Anxiety :recommendedIntervention :BreathingExercise
```

---

## 8. Reasoning Framework

## 8.1 Symbolic Rule-Based Reasoning (Primary)

All **state inference** is performed using **SWRL rules**.

Example (conceptual):

```
Student(?s) ^
hasSymptom(?s, Insomnia) ^
experiencesEmotion(?s, Stress) ^
triggeredBy(?s, ExamPressure) ^
persistsFor(?s, ?d) ^ swrlb:greaterThan(?d, 7)
→ AtRiskStudent(?s)
```

Rules are deterministic, inspectable, and explainable.

---

## 8.2 SPARQL-Based Graph Materialization

SPARQL `CONSTRUCT` / `INSERT` rules are used to:

- materialize inferred triples
- maintain session graphs
- retrieve explanation paths

SPARQL does **not** decide truth; it operationalizes rule outputs.

---

## 8.3 Dataset Usage and Role

This project does **not train predictive machine-learning models** for mental health diagnosis. Instead, datasets are used in a **knowledge-supporting and validation role**, consistent with KRR principles.

## Datasets Referenced

- **MHP Anxiety–Stress–Depression Dataset (Figshare)**
- **Kaggle: Sentiment Analysis for Mental Health**
- **PMC mental health research datasets (conceptual reference)**

## How Datasets Are Used

Datasets contribute in the following ways:

1. **Ontology Vocabulary Validation**
    - Common symptoms, emotions, and triggers were identified from datasets
    - Ensures ontology concepts reflect real student expressions
2. **NLP Extraction Pattern Design**
    - Frequent phrases and linguistic patterns informed:
        - keyword lists
        - synonym mappings
        - semantic similarity thresholds
3. **Causal Strength Annotation (Optional)**
    - Aggregated statistics from datasets may inform:
        - relative strength of relationships (e.g., Insomnia → Anxiety)
    - Stored as ontology annotations (e.g., `:causalStrength`)
    - **Never used to trigger inference**
4. **Evaluation & Case Study Validation**
    - Synthetic or anonymized samples from datasets are used to:
        - test rule coverage
        - validate explanation coherence

> **Important Constraint:**
> Dataset-derived information **does not replace symbolic rules** and **does not introduce probabilistic inference**.

---

# 9. Continuous Context-Aware Reasoning

Unlike single-turn chatbots, this system:

- accumulates symbolic facts across messages
- reasons over persistence and frequency
- updates beliefs incrementally

Example:

1. "I can't sleep." → Insomnia
2. "Exams are stressing me." → Stress + ExamPressure

3. “I feel anxious in class.” ⟶ Anxiety

⟶ System infers **AnxietyRisk** due to accumulated evidence.

---

# 10. Causal Explanation Engine (WHY-Answers)

For every inference or recommendation, the system can answer:

- **Why is this risk state inferred?**
- **Why was this intervention suggested?**

Explanation structure:

```
Detected Concepts
→ Rule Triggered
→ Risk State Inferred
→ Intervention Mapped
```

Explanations are generated using **knowledge graph traversal**, not heuristics.

---

Each explanation includes:

- Triggered SWRL rule identifier
- Matched ontology concepts
- Ranked explanation path (if multiple exist)
- Confidence level based on evidence strength

This ensures explanations are:

- traceable
- reproducible
- auditable

---

# 11. Hybrid Causal Backing (Optional Enhancement)

To improve explanation quality (not inference):

- Simple statistical or causal estimates (e.g., conditional probabilities or Bayesian networks) may be learned from small datasets.
- These values are stored as ontology annotations (e.g., `:causalStrength`).

Important constraint:

> Statistical values are used **only to rank explanation steps**, never to infer states.

---

# 12. Ranking, Confidence & Uncertainty

The system computes an **evidence-strength score** using:

- NLP extraction confidence
- number of matched reasoning rules
- persistence of symptoms across time

This score is used to:

- present **Low / Moderate / High confidence**
- communicate uncertainty transparently

Example:

> "I'm moderately confident (70%) because anxiety signals were strong, but duration evidence is limited."

The score **does not control rule firing**.

---

## 12.1 Ranking and Recommendation Algorithms

The system may infer **multiple possible interventions** for a given risk state.
Ranking is applied **after inference**, not during reasoning.

## Ranking Criteria

Interventions are ranked using a deterministic scoring function based on:

1. **Rule Priority**
    - Each SWRL rule is assigned a priority level
    - Higher-priority rules dominate recommendations
2. **Evidence Strength**
    - Number of matched symptoms, emotions, and triggers
    - Persistence of evidence across multiple turns
3. **Causal Strength (Optional Annotation)**
    - Ontology annotation indicating strength of association

    - Used only for ranking explanations, not inference

4. **Safety Constraints**

    - Non-invasive interventions are preferred at lower risk
    - Professional support is prioritized at higher risk

### Ranking Formula (Conceptual)

```
RecommendationScore =
  (RulePriority × EvidenceCount × PersistenceFactor)
  × CausalStrength
```

### Important Constraints

- Ranking does **not affect which rules fire**
- Ranking does **not override safety escalation**
- All rankings are explainable and inspectable

---

## 13. Safety & Escalation Policies (Mandatory)

Safety policies operate **outside** the reasoning engine.

### Hard Escalation Rules

If the user mentions:

- self-harm
- suicidal ideation

Then:

- ontology reasoning is bypassed
- emergency contacts are shown
- professional help is strongly encouraged

### Soft Escalation Rules

If multiple severe symptoms or high-risk patterns are inferred:

- immediate professional support is suggested
- campus counseling resources may be shown

All escalation events are:

- logged
- stored in an audit trail

---

## 14. Ethical Boundaries

This system:

- does **not** diagnose
- does **not** replace mental-health professionals
- provides **non-clinical, educational support only**

The system explicitly communicates its limitations to users.

---

## 15. Deliverables

1. Mental health ontology ( `.owl` , `.ttl` )
2. Base knowledge graph + session graphs
3. SWRL rule set
4. SPARQL rule/query files
5. NLP extraction pipeline
6. Backend service
7. Chat interface (optional)
8. Case studies with explanations
9. Final report (architecture + reasoning)
10. Dataset-informed ontology annotations and validation notes
11. Recommendation ranking specification

---

## 16. Scope Summary

✔ Ontology-driven reasoning
✔ Explainable causal inference
✔ Continuous context
✔ Safety-first design

❌ No deep learning training
❌ No deployment
❌ No authentication

## 17. Final Note

This project demonstrates a **true KRR system** where:

- knowledge is explicit
- reasoning is symbolic
- explanations are causal
- uncertainty is acknowledged
- safety is prioritized

> While a modern chat interface is used for interaction, the project's contribution lies entirely in **ontology-driven reasoning, causal explanation, and ethical KRR system design**.
> The interface serves only as an access point to the reasoning engine.

---

# ✔ This file is complete and ready to use as `context.md`.