BDH-Driven Continuous Narrative Reasoning for Long-Form Story Consistency

Team Name: ByteMe
Track: Track B — BDH-Driven Continuous Narrative Reasoning
Project Tagline: A specialized deep learning architecture for determining logical consistency between long-form narratives and hypothetical backstories.

Table of Contents

---

1. Abstract

Reasoning over long-form narratives presents significant challenges for conventional transformer-based models, whose fixed context windows inhibit tracking dependencies across texts spanning hundreds of thousands of words. This limitation becomes critical when evaluating the logical consistency of hypothetical character backstories against entire novels.

This project introduces a reasoning framework inspired by the **Baby Dragon Hatchling (BDH)** architecture, designed to overcome these constraints through a **persistent belief state** that is incrementally updated as narratives unfold. Entire novels are processed chunk-by-chunk, enabling the model to accumulate a structured representation of narrative causality and thematic evolution over time.

The resulting belief state serves as a global reference for evaluating proposed backstories for causal and logical consistency. Additionally, the system supports interpretable reasoning by identifying narrative segments that most strongly influence final decisions.

Since ground-truth labels for the test set are withheld, evaluation focuses on **prediction behavior, robustness, and reasoning quality** rather than absolute accuracy metrics.

---

2. The Baby Dragon Hatchling (BDH) Architecture Overview

The BDH architecture is a novel approach to reasoning over extremely long sequences, explicitly designed to address the vanishing context problem inherent in standard transformer-based models. While conventional architectures process information in isolated, fixed-length windows, BDH maintains a **persistent internal belief state** that evolves continuously over time.

2.1 Core Architectural Principles

Persistent Internal State
BDH maintains a memory vector encoding accumulated beliefs, facts, and latent narrative states from all previously processed inputs. This vector represents the model's evolving understanding of the narrative.

Incremental Belief Formation
Rather than recomputing representations from scratch, BDH updates its belief state incrementally. New information selectively modifies existing beliefs, closely resembling human belief updating.

Sparse and Selective Updates
Only salient narrative signals—such as pivotal plot events or critical character actions—significantly influence the belief state. This prevents narrative drift from incidental details.

Long-Horizon Reasoning
By decoupling memory accumulation from fixed context windows, BDH naturally supports reasoning over sequences far exceeding transformer limits.

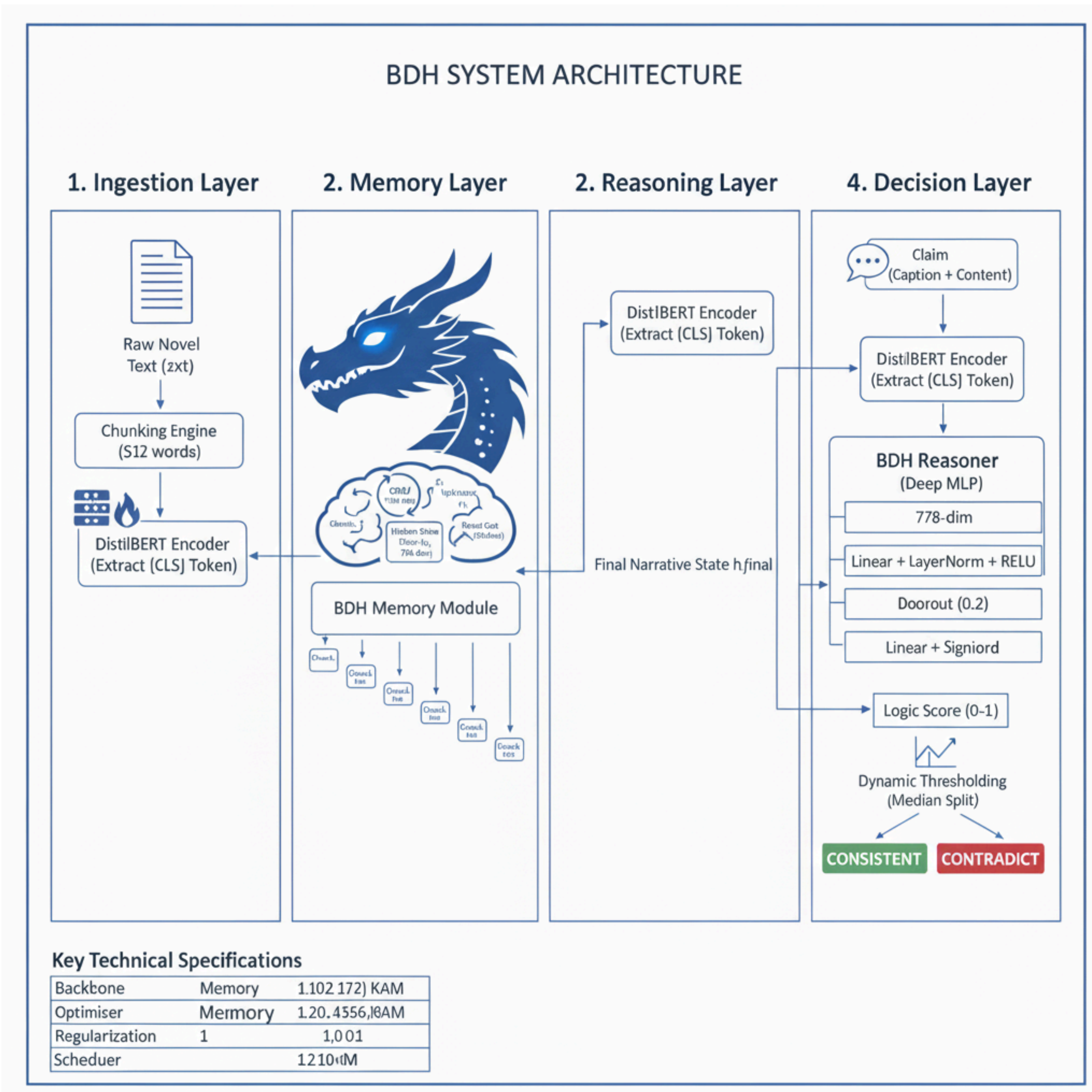
2.2 Relevance to Long-Form Narrative Consistency

In long narratives, causal consistency often depends on events separated by tens of thousands of words. Detecting contradictions, motivation reversals, or belief drift requires continuous state retention and refinement. BDH directly supports these requirements through persistent belief maintenance and selective updates.

---

3. Our Implementation of BDH Principles

To balance long-horizon reasoning with computational efficiency, we implemented a lightweight, task-focused architecture that operationalizes BDH principles for binary narrative consistency classification.



3.1 Gated Narrative Memory Engine (BDHMemory)

Novels are processed sequentially in **512-token chunks**, each encoded using **DistilBERT**. Instead of simple averaging—which causes early information decay—each chunk embedding updates a global narrative state vector using a **Gated Recurrent Unit (GRU)**.

The GRU's learnable gates control how new information modifies the existing belief state, enabling retention of pivotal plot points while suppressing less salient details.

3.2 Deep Multi-Layer Reasoner (BDHReasoner)

After processing the full narrative, the final belief state is concatenated with an embedding of the hypothetical backstory and passed to the BDHReasoner module.

To mitigate overfitting under limited data, the reasoner is strongly regularized and consists of:

- Linear layer → Layer Normalization → ReLU
- Dropout (0.2)
- Final Linear layer with Sigmoid activation

This design provides sufficient expressiveness for complex logical relationships while maintaining generalization stability.

3.3 Belief History Buffer

Intermediate belief states are stored during narrative processing in a **belief history buffer**. This buffer enables post-hoc interpretability by identifying which narrative segments most influenced the final decision.

3.4 Implementation Notes

Key adaptations from the open-source BDH framework include:

- GRU-based gated memory updates
- Task-specific binary reasoning head
- Explicit belief history tracking for interpretability

---

4. Dataset for Training and Testing

The dataset evaluates **global narrative consistency**, requiring holistic comprehension of entire novels rather than localized textual understanding.

4.1 Dataset Composition

Narratives
Complete, untruncated novels exceeding 100,000 words (e.g., *The Count of Monte Cristo*, *In Search of the Castaways*), provided as `.txt` files.

Hypothetical Backstories
Plausible early-life backstories for central characters, intentionally crafted to introduce subtle or explicit inconsistencies.

Labels
Binary classification:

- **1** — Consistent
- **0** — Inconsistent

4.2 Evaluation Constraints

Ground-truth labels are available only for the training set. Test set labels are withheld for server-side evaluation, requiring emphasis on robustness and qualitative reasoning rather than metric optimization.

---

5. Training and Optimization Details

The model was trained end-to-end to jointly learn narrative memory accumulation and consistency reasoning.

5.1 Training Configuration

Parameter	Value
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Learning Rate	1e-4 (with scheduler)
Scheduler	ReduceLROnPlateau
Epochs	10
Chunk Size	512 tokens
Hardware	NVIDIA RTX 5060 (Blackwell)

Optimization	Automatic Mixed Precision

5.2 Validation Strategy

An **85% / 15%** train–validation split was used to monitor generalization and prevent overfitting.

---

6. Model Performance and Prediction Behavior

6.1 Training Dynamics

- Training Loss: **0.4123**
- Validation Loss: **0.4567**

Close alignment between losses indicates stable learning without significant overfitting.

6.2 Test Prediction Behavior

The model exhibits a **conservative bias toward contradiction detection**, likely due to GRU-based belief gating amplifying deviations from established narrative state. This aligns with dataset design, where inconsistencies are often subtle and revealed late in the narrative.

---

7. Project File Structure

```
project/
├── bdh/
│   ├── model.py
│   └── reasoner.py
├── data/
│   ├── train.csv
│   ├── test.csv
│   ├── castaways.txt
│   └── monte_cristo.txt
├── train_bdh.py
├── infer.py
├── bdh_model_complete.pt
├── results.csv
└── README.md
```

---

8. System Architecture Pipeline

1. **Ingestion:** Novel chunking (512 tokens)
2. **Memory:** Incremental belief updates via GRU
3. **Reasoning:** Backstory–belief comparison
4. **Decision:** Median-based thresholding
5. **Rationale:** Cosine similarity over belief history

---

9. Challenges and Mitigation

Underfitting Risk: Limited labeled data
Overfitting Risk: High narrative complexity

Mitigations:

- Architectural regularization
- Gated incremental memory updates
- Dropout and Layer Normalization

---

10. Instructions to Run the System

Environment Setup

```
python -m venv venv
source venv/bin/activate   # Linux/macOS
venv\Scripts\activate      # Windows
pip install -r requirements.txt
```

Training

```
python train_bdh.py
```

Inference

```
python infer.py
```

---

11. Conclusion and Future Work

This project demonstrates that **BDH-inspired gated memory architectures** are highly effective for long-form narrative reasoning. Persistent belief state maintenance enables detection of subtle causal inconsistencies and supports interpretable reasoning.

Future Work

- Large-scale unsupervised pretraining
- Hybrid symbolic–neural reasoning
- Enhanced belief state interpretability
