

Project Proposal

Stochastic Interactive Storytelling via Commonsense-Guided Reinforcement Learning

Team Members

- Ninad Deo (UIN: 535005965)
- Dhruvraj Singh Rathore (UIN: 735007897)

Motivation

Storytelling is inherently uncertain: even if a hero *chooses* to “fight the dragon,” the outcome might be victory, injury, escape, or death. To date, many narrative generation or continuation systems are deterministic or based on supervised learning. We propose to treat **interactive storytelling** as a **stochastic Markov Decision Process**, where the RL agent selects narrative branches but the environment responds probabilistically. By combining **commonsense knowledge** (from datasets like ATOMIC / Story Commonsense) with RL, we aim to produce more robust, coherent, and engaging stories that can adapt to uncertainty.

Why the Project is Interesting & Non-Trivial

- We combine **reinforcement learning** with **natural language and commonsense reasoning**.
- Designing a **stochastic narrative environment** with meaningful reward is quite challenging.
- It requires integrating NLP tools (embeddings, commonsense data) *and* RL architecture (stochastic transitions, policy learning).
- The project is both **creative** (you get to produce stories) and **technical**.

Sequential Nature

The problem is sequential because each narrative decision (at time t) affects all future possibilities and the final outcome. Good intermediate choices may lead to

richer future branches; poor choices may constrain the story irreversibly. The agent must balance short-term coherence with long-term payoff (ending quality).

Components

State Space

Each state represents the current point in the story. It includes:

- The sentence embedding of the current line (using DistilBERT).
- Aggregated emotional and motivational features of the characters from the *Story Commonsense* dataset (e.g. dominant emotions like *joy* or *confusion*, and motivations such as *to understand* or *to have fun*).
- A scene index indicating the position in the narrative.

Action Space

At each state, the agent selects one of several possible story continuations:

- The **true next line** from the dataset, and
- A few **alternative continuations** generated using the ATOMIC commonsense relations (e.g., “xEffect,” “xWant”) or a small language model..

Each action represents a narrative choice that determines how the story will progress.

Transition Function

The environment is **stochastic**: choosing the same action can lead to different next states.

For example, selecting “Nana asks about the socks” might lead to different emotional outcomes such as *curiosity* (60%), *frustration* (30%), or *amusement* (10%).

This uncertainty is modeled using fixed or learned transition probabilities derived from story patterns and ATOMIC relations.

Reward Function

The reward measures how coherent and emotionally consistent the story remains after each action:

- **+1** for coherent transitions (embedding similarity with the true continuation).
- **+1** for consistent or natural emotional change.
- **+5** for reaching a satisfying or joyful ending.
- **-1** for incoherent or abrupt transitions.

Expected First Result

By the end of Phase 1, we expect the DQN agent in a toy environment to reliably reach the “good ending” > 80% of the time (versus $\sim\frac{1}{3}$ for random). That will already demonstrate the feasibility of our approach.

RL Algorithm

We will use a **Deep Q-Network (DQN)** since the action space is discrete, mapping each story state to Q-values for possible continuations. An ϵ -greedy strategy and experience replay will support stable, diverse learning. If instability arises due to stochastic transitions, we will extend to **Proximal Policy Optimization (PPO)** for better handling of uncertainty.

Evaluation & Baselines

- Baseline 1 (Random): choose a random action among candidates at each step.
- Baseline 2 (Oracle): always pick the dataset’s true next continuation when available.
- Metrics / Evaluation:
 - Fraction of times agent picks true next continuation
 - Embedding coherence score (cosine similarity)
 - Diversity / novelty of generated stories
 - Human evaluation (if possible): rating story quality, coherence, enjoyability

Stretch / Bonus Goals

- Use pretrained language models to propose continuations and transition probabilities (making the environment more realistic).
- Develop a simple UI to visualize branching story paths, character emotion flows, and agent choices over time for qualitative evaluation.