Ontologically Faithful Generation of Non-Player Character Dialogues

Anonymous ACL submission

Abstract

We introduce a language generation task grounded in a popular video game. KNUDGE (KNowledge Constrained User-NPC Dialogue GEneration) requires models to produce trees of dialogue between video game characters that accurately reflect quest and entity specifications stated in natural language. KNUDGE is constructed from side quest dialogues drawn directly from game data of Obsidian Entertainment’s The Outer Worlds, leading to real-world complexities in generation: (2) utterances must remain faithful to the game lore, including character personas and backstories; (2) a dialogue must accurately reveal new quest details to the human player; and (3) dialogues are large trees as opposed to linear chains of utterances. We report results for a set of neural generation models using supervised and in-context learning techniques; we find competent performance but room for future work addressing the challenges of creating realistic, game-quality dialogues.

1 Introduction

Player interactions with non-player characters (NPCs) in role-playing games (RPGs) often serve to flesh out backstories while allowing the player to progress through engaging quest storylines (Onuczko et al., 2007). Figure 1 shows a dialogue turn, taken from The Outer Worlds (Obsidian Entertainment, 2019),1 an action RPG renowned for its writing. A key challenge in creating NPC dialogues is that they should serve coherent narratives; utterances must faithfully reflect quest structure and game lore—characters, histories, and entity relationships. Additionally, NPC interactions often take the form of complex trees that can have dozens of nodes. Creating these branching structures according to the many specifications and ontological constraints of dialogue authoring can be time-consuming for game designers (Caropreso et al., 2012) and cost companies millions of dollars (see §A). This motivates the pursuit of tools for automatically generating dialogue trees.

However, there is a lack of realistic benchmarks to train and evaluate models for this purpose. van Stegeren and Theune (2020) highlight that game text corpora should come from real, professionally written games; most research that explores game dialogue relies on crowdsourced or academically-curated text, which is not representative of the highly game- and context-sensitive text of real dialogues. Moreover, related work on game dialogue (Urbanek et al., 2019; van Stegeren and Mysliwiec, 2021), story generation (Akoury et al., 2020; Chen and Gimpel, 2021), and knowledge conditioning for task-oriented dialogue agents (Choi et al., 2018; Mazaré et al., 2018; Feng et al., 2020) does not address complex dialogue trees and interweaving narratives found in deployed RPGs.

---

1https://en.wikipedia.org/wiki/The_Outer_Worlds

** KNUDGE will be publicly available upon publication.
To address this dearth of realistic benchmarks for game dialogue authoring, we introduce **KNUDGE:** KNowledge constrained User-NPC Dialogue GEneration, a set of dialogue trees (in English) extracted from *The Outer Worlds*, an existing video game, and paired with granular ontological constraints. **KNUDGE** contains 159 dialogues from all 45 side quests in *The Outer Worlds*. It contains 4.7K utterances and an average of 1.3K input constraint tokens per tree. For each quest, we annotate each turn in the dialogues with relevant grounding information—quest- and lore-related natural language (NL) support facts pulled from fan-written wikis. Such fine-grained support facts annotations are useful for training and evaluating models to generate game dialogues grounded in quest specifications and game lore. To the best of our knowledge, ours is the first dataset that consists of a set of real game-quality NPC dialogues paired with granular quest and biographical specifications consistent with a well-formed game ontology.

Using **KNUDGE** as a test bed, we devise the task of knowledge-constrained NPC dialogue generation (shown in Figure 2 and described in §2). The complex input specifications and limited training data target a realistic development scenario in which a designer is working on a new, partially-written game. For this task, we introduce a model class, termed DialogueWriters, that leverage in-context learning with LMs such as GPT-3 (Brown et al., 2020) to generate continuations of partial dialogue trees given input constraints. To address the challenges of long constraining specification passages and branching tree structures that are too large to fit in GPT3’s context window, we introduce techniques for prompt construction, retrieved exemplar munging, and transforming tree structures into maximal-coverage, prompt-friendly representations. To encourage the use of game lore to produce engaging dialogue, we experiment with adding mechanisms to subselect relevant facts before generating each utterance. These methods take advantage of the rich, node-level annotations of **KNUDGE** and show improved text overlap with the underlying ontological specifications.

Finally, we prescribe evaluation protocols that test models’ capacity to reflect game ontology constraints in addition to generating fluent and coherent dialogue. We conduct automatic and human evaluations of utterances and multi-turn trees generated from specifications for game quests, as well as for a pair of never-before-seen quests written by a professional game designer. Our experiments reveal further room for improvement in aspects such as ontology usage and maintaining coherence.

### 2 Task Definition

As communities seek to develop AI-based writing tools such as Ubisoft’s GhostWriter (Barth, 2023), there arises a need to reconcile the challenge of dialogue authoring with current NLG techniques. An ideal writing tool might allow for a designer to provide the input types depicted in Figure 1 (granular quest information and bios from a game’s lore) and receive a set of generated utterances, or entire trees, to aid in drafting content similarly to GitHub Copilot.\(^2\) As a game ontology can be large, so too might the amount of user-provided knowledge specifications. We structure our new task to target this scenario, for which no such dataset exists to study how to optimize models.

We define the task of knowledge-constrained

\(^2\)https://github.com/features/copilot/
NPC dialogue generation as the mapping from a set of quest constraint statements $Q$, a set of biographical constraint statements $B$, and a list of participant names $P$ to a dialogue tree $D$. We consider two task scenarios: next utterance prediction from a partial (gold) tree, and full dialogue generation of trees with multiple candidates at each turn.

As depicted in Figure 2 (upper), $Q$ comprises statements $[q_1, \ldots, q_m]$ about currently active objectives upon entering the dialogue, about what should occur during it (e.g., pieces of information the NPC should mention), and about new active objectives upon leaving it. $B$ comprises background statements $[b_1, \ldots, b_n]$ about game entities that the dialogue must reflect (Figure 2, middle). The participant list $P$ contains the player character and one or more NPCs who have corresponding facts in $B$. We design these inputs to reflect the kinds of specifications that a developer would provide to the generator during their quest writing process.

Dialogue $D$ is a directed graph $(N, E)$; each utterance node $n \in N$ has a speaker $s \in P$. Branches occur due to the multiple dialogue options for players (Figure 3, right). $D$ has one start node, but can have multiple exit nodes and can contain cycles.3

To simulate a realistic writing scenario, we provide models with a small set of training dialogues from the particular game domain, some $(Q_1, B_1, P_1, D_1), \ldots, (Q_t, B_t, P_t, D_t), t \approx 100$. While this is not enough to perform SGD-based fine-tuning, models such as those described in §4 can leverage it for in-context learning.

3.1 Game Ontology

We acquired dialogue files from the Outer Worlds creators along with permission to release them publicly; we use quest data and game lore from fan wikis, where a quest’s page lists the in-game objectives and journal logs (though the framework allows for using data from official channels instead of fan-made). §B contains further source details.

Quest Information A quest in the Outer Worlds appears in the player’s journal with a high-level synopsis and a sequence of objectives, each of which contains game logs providing additional details. Active objectives are completed, and new ones introduced, during an NPC dialogue. We associate with each objective a walkthrough passage which includes details on the topics, player utterance options, and quest information that the NPC needs to say by the dialogue’s end. A detailed quest anatomy and examples of $Q$ can be found in §C.

Biographical Information We associate with each quest, and in turn each dialogue, a set $B$ of biographical passages about entities appearing or referenced during the quest. We extract passages from entities’ fan wiki pages. While some are only a few sentences, others can be much longer (up to 27), posing a challenge to generation models; often only part of a long biography might be relevant to a given quest. Examples are shown in §D.

3.2 Dialogue Trees

Dialogue trees in The Outer Worlds are complex directed graphs, containing many conditional utterance options depending on the state of the game—e.g. whether the player is of high enough level at...
To extract a more tractable, quest-related subgraph, we 1) identified the nodes that start and end the interaction using online playthrough videos as reference, and then 2) traversed the graph from the start node, following only edges without state-related conditions. Conditional edges were then added manually depending on quest relevance. §E depicts example trees.

Annotating Utterance Nodes with Support Facts

We coordinated with English-speaking professional data annotators to label trees nodes with support facts from the associated quest and lore constraints. We provided a set of instructions shown in §I, which follow a heuristic based on counterfactuals: had a given fact not been included in the constraining knowledge, would the utterance be much less likely to occur? An example of this procedure is depicted in Figure 3. We compute an average exact match (EM) score of 0.52 and Jaccard overlap score of 0.62. These scores represent a high agreement on a subset selection problem where the total set size is often of the order of hundreds.

3.3 Dataset Analysis

Table 1 describes statistics for the extracted dialogue trees and annotations. Notably, 47% of all nodes (57% of NPC nodes) in the dataset are annotated with at least one fact. NPC utterances, for which support knowledge is more common, have an average of 1.0 facts. Dialogues have on average 9.9 quest facts and 73.3 biographical facts that models must consider when producing utterances.

Comparison with Related Datasets

KNUDGE is the first dataset to contain dialogue trees from an actual RPG annotated with game quest and biography specifications. Table 2 compares our dataset to contemporaries with comparable input specifications and generation targets. None contain all the components to target the complexity and granular specificity required to generate quest dialogues of the type found in The Outer Worlds. §F provides an in-depth comparison with select works.

Challenges

Generating a 30-node (or larger) NPC dialogue tree while simultaneously taking into account all of the NPC desiderata described above is a very difficult task, particularly given the shape of the branching, cycle-heavy tree structure. The average of 1321 constraining tokens and 406 utterance tokens, which are both longer and more complexly structured than in existing datasets, poses a challenge to current NLP models, taking up e.g. half of the 4000-token context window of GPT-3 before factoring in other pieces of context.

4 DialogueWriter Methods

We introduce a set of methods called DialogueWriters for generating candidate utterances given the ontological specifications (Q, B, and P) from §2 and a partial subtree S for a dialogue in KNUDGE. DialogueWriters propose utterances at a specified new location branching off S. Given some “most recent” node \( n \in S \), the method maps inputs \((Q, B, P, S, n)\) to a list of candidates \([c_1, \ldots, c_n]\) such that there is a directed edge \( n \to c_i \).
We thus devise a traversal mechanism that, at inference time, converts a current dialogue subtree into a maximal coverage linear history. For “most recent” node $n$, we identify the longest possible path from the start node to $n$, including cycles but only following a given edge once. This produces utterance history $H = [u_1, \ldots, u_n]$. Exploring other tree encoding mechanisms, e.g. via graph encoders (Banerjee and Khapra, 2019; Ouyang et al., 2021) is left for future work. We feed the traversed history to a next utterance generator learnt via supervised or in-context learning described below.

### 4.1 Supervised Learning (SL) Models

We fine-tune a T5-large model (Raffel et al., 2020) via supervised learning (SL) to generate $c_i$ given the concatenation $[B, Q, P, H]$. We truncate context from the left of the when required given T5’s 1024-token window. §G shows example inputs.

#### Supervised Knowledge Selection (KS) Model

We also train a version of the SL baseline that learns to decode support knowledge facts before conditionally generating the utterance $c_i$. This factorizes the next utterance generation into a two-step decision process: first selecting one or more facts from the provided knowledge constraints ($Q \cup B$), and second generating the utterance to reflect the selected facts. We thus make use of our node-level annotations to train the model to generate the concatenated sequence $[j_1^{(1)}, \ldots, j_m^{(1)}]$, s.t. $j_j^{(i)} \in Q \cup B$.

### 4.2 In-Context Learning (ICL) Models

As there is little training data, fine-tuning might not be effective at learning the difficult generation task. As such, we experiment with methods for in-context learning (ICL) with OpenAI’s text-davinci-003 GPT-3 model (Brown et al., 2020). We inject $B, Q, P,$ and $H$ into a formatted prompt that naturally elicits the next utterance as a continuation of $H$. Figure 5 depicts this process; full prompts are shown in §G. This creates a zero-shot prompt. When this does not fill out GPT-3’s 4000-token window, we construct a few-shot prompt by adding dialogs from training quests as exemplars, simulating a scenario in which a developer has written a partial set of quests and is working on a new one. We retrieve exemplars using Okapi-BM25 (Jones et al., 2000) with $[B, Q, P]$ as the query string.

#### ICL Knowledge Selection (KS) Model

As with the SL framework, we also devise a version of the ICL DialogWriter that first decodes one or more support facts before generating an utterance. We elicit this behavior from GPT-3 by augmenting all utterances in the dialogue history with support fact nodes, provided they exist (see Figure 15). Further model training details are found in §H.

### 5 Experiments

We split KNUDGE into train, development, and test splits on the basis of quests (28/5/12), such that at test time all input components will be unseen (test set $B$’s contain a combination of previously seen and totally novel entities). For each test dialogue, we evaluate predicted utterances at each $n$th node in
the tree conditioned on a subtree composed of the nodes and edges between the first \( n-1 \) nodes. We measure 1) human judgments of NPC desiderata and 2) utterance overlap against gold items and quest and bio knowledge (§5.3). We then run a case study to collect human judgements of the skeletons of entire dialogues generated end-to-end (§5.5).

### 5.1 Baseline Models

To measure the effect of node-level knowledge selection (KS), we compare against an ICL model that selects only one statement instead of many. We randomly sample gold facts to construct its prompt. We also compare against an oracle KS ICL model, which conditions on the full gold knowledge annotations for the reference utterance. We maximize the number of in-context examples for all ICL ablations; e.g. the no knowledge model’s prompt can have dozens of such examples, as they are quite short. These ablations thus explore the tradeoff between the impact of the number of in-context examples and the presence of ontological statements.

To measure the effect of KS, we compare against SL and ICL no KS models. To measure the effect of conditioning on \( Q \) and \( B \), we compare against ablations to the non-KS ICL model: a no knowledge model that conditions only on the participants \( P \) and utterance history \( H \), and a quest only model that conditions on \( P, H, \) and \( Q \), but not \( B \).

### 5.2 Next Utterance Prediction (NUP)

**Human Evaluation** In coordination with a data specialist, we conduct human evaluation to examine models’ qualitative NUP performance. Generated utterances over 100 test items were judged on a 4-point Likert scale for each of four criteria: 1. **Coherence**: does the utterance follow naturally from the utterances in the history? 2. **(Non-)Violation**: does the utterance create contradictions with any of the sentences in the biographical or quest passages? 3. **Biography Usage**: does the utterance make use of the input biographical passages in \( B \)? 4. **Quest Usage**: does the utterance progress the dialogue according to the quest sentences in \( Q \)? We provide the full set of annotator instructions in §1. Results were verified via bootstrap testing.

**Automatic Evaluation** We use reference-based metrics BLEU-4 (Papineni et al., 2002) and rescaled BERTScore-F1 (Zhang et al., 2020). We evaluate generated next utterance against the following single- and multi-reference sets: 1) the gold utterance \( n_{1} \), 2) the quest statements in \( Q \), and 3) the biography statements in \( B \). BERTScore against \( Q \) and \( B \) is a neural analog to Knowledge-F1 (KF1; Shuster et al., 2021). We evaluate the gold utterance to show a performance upper bound.

### 5.3 Results

Next utterance prediction results under human and automatic metrics are shown in Table 3 and Table 4. Table 3 shows that no model reaches gold coherence nor gold use of \( B \) under human evaluation, suggesting room for improvement on both.

**Impact of Knowledge Selection:** Table 4 depicts the trend that KS variants of the ICL model score a point or two higher than non-KS on overlap with \( B \) and \( Q \), reflecting that KS effectively selects and cues the infusion of specific facts into generations. Oracle KS improves BLEU score with \( B \) but not \( Q \), while one-fact KS has the opposite effect. The ICL ablations of \( B \) and of \( \{ B, Q \} \) have according drops in overlap with both fact sets.

All \( B \) and \( Q \)-conditioned ICL models perform equivalently under all human metrics except (non)violations, at which KS models perform a decimal point worse. We can conclude that KS improves the capacity of ICL writers to directly reflect knowledge passages (i.e. by copying spans), at the expense of a slightly higher likelihood of contradictions. This characterization can be appealing to a game developer; they might prefer for the automatic writer to use their own provided wordings of various facts when generating candidates.

**Interpretation of Automated Metrics:** We note that automatic metrics that check for overlap with reference text will, if used in isolation, give only a partial picture for evaluating generations in
Table 4: NUP BLEU and BertScore for models against gold utterances and statements in $B$ and $Q$. Results for the latter two shown beside the gold utterance’s score.

<table>
<thead>
<tr>
<th></th>
<th>Gold Reference</th>
<th>SL-KS</th>
<th>ICL-KS</th>
<th>KS Variants:</th>
<th>Non-KS Baselines:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL  BS</td>
<td>BL BS</td>
<td></td>
<td>ICL-KS-One</td>
<td>SL (no KS)</td>
</tr>
<tr>
<td>Gold Reference</td>
<td>–</td>
<td>4.9</td>
<td>20.8</td>
<td>2.2</td>
<td>17.8</td>
</tr>
<tr>
<td>SL-KS</td>
<td>2.6</td>
<td>21.3</td>
<td>24.8</td>
<td>26.6</td>
<td>9.3</td>
</tr>
<tr>
<td>ICL-KS</td>
<td>7.1</td>
<td>25.1</td>
<td>8.3</td>
<td>24.3</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Table 5: Pairwise comparison (% head-to-head wins) between generated trees from ICL DialogueWriters. For example, 81.2% of the time ICL outputs were preferred over a competing approach with respect to coherence.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ICL-KS</td>
<td>50.0</td>
<td>68.8</td>
<td>68.8</td>
<td>75.0</td>
<td>43.8</td>
</tr>
<tr>
<td>ICL</td>
<td>81.2</td>
<td>75.0</td>
<td>81.2</td>
<td>75.0</td>
<td>75.0</td>
</tr>
<tr>
<td>ICL-Quest</td>
<td>56.2</td>
<td>56.2</td>
<td>43.8</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>ICL-No Know.</td>
<td>12.5</td>
<td>0.0</td>
<td>6.2</td>
<td>0.0</td>
<td>31.2</td>
</tr>
</tbody>
</table>

5.5 Full Dialogue Generation Case Study

We run a case study for a full tree generation scenario. We task models with generating 10 rounds of dialogue given just the specifications $B$, $P$, $Q$, and one starting utterance. At each turn, we generate three candidate nodes using the writer, then randomly “commit” one to the linear history. This creates a 31-node tree (example in Figure 16) that can serve as a ‘spine’ of proposed content to be fleshed out by a developer into a more complicated tree. We selected 8 test dialogues from the game with varying quest roles, e.g. starting vs continuing quests. Additionally, we constructed 8 more test items from 2 totally novel quests, written for us by a professional game designer, that occur in the Outer Worlds universe and contain entities from the original game. We showed trees generated by DialogueWriters to data specialists for evaluation following the ACUTE-Eval (Li et al., 2019) protocol of pairwise comparison. Annotators selected which of two trees were preferred for the following criteria: coherence, nonviolation, biography and quest usage analogous to §5.2, and additionally 5. Content Suggestion: Do the multiple candidates at each turn propose interesting subtrees? 6. Engagingness: does the tree hold your attention and make you want to hear more from the NPC?

Results are provided in Table 5. Performance is measured as the rate at which annotators selected a model’s tree in a pairwise comparison under the 6 criteria listed in §5.5. We find that annotators preferred the trees of the ICL writer most frequently compared to the other models under all criteria except $Q$ usage. We also find that the ICL-KS method is more frequently preferred for the never-before-seen quests written by the professional designer.
while it is less frequently preferred for dialogues from the actual game. The opposite holds for ICL-Quest only. A possible explanation is that the designer provided many biographical specifications for his quests (filling the GPT3 context window), which the Quest only model does not see.

6 Related Work

Si et al. (2021) focus on the task of story continuation through dialogue between multiple characters while modeling the inter-character relations. However, such past work does not concern with the notion of grounding knowledge or quest objectives to be covered in the generated dialog. van Stegeren and Theune (2020) propose three sources for building NPC dialogue corpora. However, their proposed datasets do not contain any grounding annotation and are not accompanied by explicit descriptions of entities and characters. Callison-Burch et al. (2022) explore automatic generation of conversational turns by players of the tabletop RPG, Dungeons and Dragons, in which NPCs serve a very different role in the gameplay. Scheherazade’s Tavern (Aljammaz et al., 2020) augments a pattern-matching-based NPC interaction system with facts the character knows about the game world. More broadly, past work has explored applications of text generation in various gaming applications such as quest description generation (van Stegeren and Mysliwiec, 2021), dialogue generation (Si et al., 2021), persona-specific agents in text environments (Urbaneck et al., 2019), and new text world generation (Fan et al., 2020; Ammanabrolu et al., 2022).

Past work has pursued dialogue systems that steer the conversation towards a topic (Wu et al., 2019) or a given NL sentence (Sevegnani et al., 2021; Gupta et al., 2022) while conversing with a user. Other work in NLG has explored generating outputs with high-level NL specifications such as string item agendas (Kiddon et al., 2016), sets of facts (Orbach and Goldberg, 2020), or author goals (Riedl, 2009). KNUDGE also comprises NL specifications, though they are comparably richer.

7 Conclusion

Humans play games to be entertained, and they pay money expecting a high-quality experience. When game dialogue is used to advance a carefully crafted storyline, it should be both engaging and consistent with the larger narrative. Language models are increasingly capable of producing engaging dialogue, which researchers have explored how to ensure is consistent with underlying knowledge specifications. To date, this research has focused on datasets developed for the sake of experimentation, rather than actual high-quality game data.

This paper introduces KNUDGE, a dataset of NPC dialogue trees coupled with a relevant game ontology, drawn from the title The Outer Worlds. In contrast to prior work, KNUDGE is based on content created by a high-budget, high-profile game development studio, thereby exemplifying real-world complexities in NPC dialogue authoring. We pose a knowledge-grounded generation task that mirrors a realistic development scenario with limited training data over a complex ontology of quests and lore. We find that LM-based methods are able to generate fluent dialogue that relates to provided specifications, but they do not match the quality of professional writers, particularly in terms of coherence and use of the game lore. We hope that KNUDGE drives the development of new techniques for faithful game dialogue generation.
**Limitations**

We find that the proposed DialogueWriter models leave room for improvement on persona embodiment. Human-quality utterances more seamlessly and dynamically incorporate emotions fitting of characters and situations, while model-generated utterances can be comparatively bland. This work also focuses on side quests whose NPCs are generally not as fleshed out as those in main quests. Generating quests containing major NPCs with long bios and important roles in the main story of a game, e.g. companion characters, is also left for future work.

**Knudge** recasts a set of fan articles about an existing game as specifications to an automatic dialogue tree writer. It therefore assumes that game developer will write structured game lore and high-level quest specifications in a similar manner beforehand when coming up with new content. Future work can look at copilot tools for authoring such high-level quest specifications and design of new characters.

We report results with large pre-trained language models whose training data was not publicly released. It is therefore difficult to know whether the game data used for experimentation is part of the training data for such models, as The Outer Worlds came out in 2019. As such, the results from such large language models should be interpreted with caution. We partially mitigate the issue by having an expert game developer construct a totally new quest specification, and report results on the this previously ‘unseen’ test data.

**Ethics Statement**

We acknowledge that there maybe bias in the data used to train the neural language models considered in this paper (T5 and GPT-3) that would lead to NPC dialogues that are offensive, implicitly or explicitly discriminatory. This poses a potential risk for deployed models, as using the proposed DialogueWriters as content suggestion tools might lead to RPG content that reflects these biases. We hope that professional game developers will have the resources to moderate damaging content before it makes its way into released products.

**References**

Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. 2020. STO-


Angela Fan, Jack Urbanek, Pratik Ringshia, Emily Dinan, Emma Qian, Siddharth Karamcheti, Shrimai Prabhunoye, Douwe Kiela, Tim Rocktäschel, Arthur


Appendix

A Cost of Authoring NPC Dialogues

Outer Worlds has 10 narrative/design credits, which seems to be about average (Fallout has 4-12, while Skyrim has 9). Per salary.com, that’s a position with an average salary of $58k per year (and probably more for AAA titles). Given a 2-3 year average development time for AAA titles, that works out to a conservative ballpark estimate of $1.2m for just this one game.

B Dataset Construction Details

B.1 Data Sources

Quest data and walkthrough passages were pulled from the Outer Worlds wiki of Fextralife, a gamer-focused site containing fan-made walkthroughs for many popular RPGs. Game entity biographies were collected from Fandom. The biography passage for a given entity is the same across all quests in which the entity appears in KNUDGE, and the set of entities is the same for all dialogues in a quest. Passages were segmented into individual sentences via punctuation boundaries. We identified relevant dialogues and their decision points using playthrough videos by the YouTube user, LordMatrim. All wiki articles were written in English by site users.

C Quest Anatomy and Example Items

Figure 6 provides a detailed anatomy of a KNUDGE quest, combining in-game quest data with corresponding passages from the fan walkthrough.

Figure 7 shows example quest items with corresponding game data and walkthrough passages segmented into statements.

In whole, a dialogue’s quest passage set Q contains:

1. The synopsis (1-2 sentences)
2. The in objective(s) active when entering the dialogue (1 sentence), the associated game log (1-2) and walkthrough passage (3-10).
3. The out objective(s) active upon leaving the dialogue, and the associated game log.

D Example Entity Biography Passages

Figure 8 shows example entities from The Outer Worlds with corresponding biographical passages.

E Example Dialogues Items

Figure 9 depicts a full example input item conveying quest, biographical, and participant specifica-

F Comparison with Other Datasets

van Stegeren and Theune (2020); van Stegeren and Mysliwiec (2021) consider datasets of publicly-available side quest data from RPGs such as World of Warcraft. However, their datasets vary in dialogue and quest coverage; for WoW their input is just a quest name and objective, and the generation target is a single-turn, few-sentence quest.

9https://www.youtube.com/@l0rdmatrim
10For the first dialogue in a quest, we associate the walkthrough passage describing how to obtain the quest.
11The dialogue can lead to multiple new active objectives, some optional. If the dialogue concludes the quest, then no leaving objective is associated.
description spoken by an NPC. They collect data for the game *TorchLight II* contains quest datapoints with a limited number of dialog utterances per quest and no multi-turn interactions or trees. Others of their collected datasets contain complex branching trees but without constraining knowledge. The dialogues of LIGHT (Urbanek et al., 2019) are more akin to NPC dialogues, though they comprise few-turn linear chains between two characters in self-contained episodes rather than quest-grounded interactions between a player and an NPC serving multiple game purposes. The size of constraining passages on the LIGHT dialogues are also a scale smaller than those of KNUDGE. The biographical constraints of KNUDGE are most similar to that of TVSTORYGEN (Chen and Gimpel, 2021), who also pull articles from fandom wiki pages. However, theirs is a story generation dataset where the target is a longform article describing a TV episode.

### H Model Training

To construct training items, we iterate through the nodes of each gold dialogue tree in a canonical order \([n_1, \ldots, n_t]\), where \(n_1\) is the tree’s start node. We create a separate item with each \(n_i\) as the generation target. We construct the subtree \(S(i)\) comprised of all nodes \([n_1, \ldots, n_{i-1}]\) and all edges between them. We then construct the input/output pair \((Q, B, P, S(i)) \rightarrow n_i\).

### Supervised Learning

To train SL DialogueWriter models, for every target node in the training quest dialogues, we construct 5 training examples using different random paths to the node. We train the model for 3 epochs using the default arguments from Hugging Face’s example summarization model training script.\(^{14}\) T5 models were trained with a batch size of 1 across 8 Quadro RTX 6000 for an average of 5 hours.

### In-Context Learning

Given a test item, we construct a BM25 index over the training dialogues and use it to construct an \(n\)-shot ICL prompt where \(n\) depends on the remaining space available in the context window. Few-shot examples are linearized dialogues containing the most possible nodes from the gold tree. Contexts are left-truncated and can start with partial examples.

### I Human Evaluation Directions

Below, we enumerate the instructions shown to annotators during human evaluation:

- **Coherence**: does the utterance follow naturally from the utterances in the history? (1) Utterance

\(^{13}\)Table 2 describes statistics for the 82 TorchLight II quests that contain both objective annotations and dialogue lines.

\(^{14}\)https://github.com/huggingface/transformers/blob/main/examples/pytorch/summarization/run_summarization.py
Figure 6: Anatomy of a KNUDGE quest. At any time, an *Outer Worlds* quest has currently active and previously completed objectives. To construct its KNUDGE representation, the quest’s high-level synopsis, objectives, and associated log entries from the game data are annotated with corresponding walkthrough article passages.
......
Entity: Agnes Needham
Appears in: A Family Matter
Bio: [0] Agnes Needham is a resident of Stellar Bay and the mother of Tucker Needham. [1] Agnes’ overprotective style of mothering has led her son, Tucker, to run away from home so he can experience life. [2] Despite Tucker being 42 years old, she still thinks of him as her ‘little boy’. [3] You can find her by Stellar Bay’s south-east exit, visibly shaken and calling for help.

Entity: Tucker Needham
Appears in: A Family Matter
Bio: [0] Tucker Needham is a former resident of Stellar Bay who left to join the Iconoclasts. [1] Before the quest A Family Matter, he can be found in Amber Heights. [2] Tucker was coddled by his mother from a very young age, [3] the latter insisting that danger lurked around every corner on Monarch. [4] His mother’s overprotectiveness extended well into Tucker’s adulthood, [5] leading him to seek to be free in any way possible. [6] After hearing Graham Bryant’s broadcasts, Tucker left Stellar Bay to be truly free by joining the Iconoclasts at Amber Heights. [7] He is dazzled by Graham’s preachings on true unfettered freedom from the corporate way of life and attributes his enthusiasm to his ‘childhood trauma’. [8] He is willing to do anything to remain free, even faking his own death to prevent his mother from continuing to send people to look for him.

Entity: Raptidons
Appears in: A Family Matter, At Central, Bolt With His Name, Journey Into Smoke, Makes Space Suits Wont Travel, The Amateur Alchemist, The Distress Signal, The Doom That Came To Roseway, Vulcans Hammer
Bio: [0] Raptidons are giant cat-reptile-like creatures that inhabit various planets in Halcyon. [1] They are creatures native to Monarch. [2] However some corporations have illegally imported them to other planets, [3] such as Auntie Cleo who relocated a group of them to Roseway. [4] Raptidons are of corporate interest due to their potential for producing new chemical by-products which, [5] when refined, can be used to create new board-approved products.

Entity: Sulfur Pits
Appears in: A Family Matter

Entity: Monarch
Bio: [0] Monarch, previously known as Terra 1, is one of the many moons of the gas giant Olympus and the site of a failed colony. [1] Terra 1 was initially designated as the primary colonization target of the Halcyon system. [2] The Halcyon Holdings Corporate Board had intended to completely terraform the moon, [3] wiping out the local fauna and flora and replacing it with plants and wildlife native to Earth. [4] However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes, [5] rendering them more dangerous and severely crippling the colonization effort. [6] Due to the hostile environment which they had created, [7] the Board was forced to enact a Hazard Clause covering the entirety of Terra 1. [8] Public notice of the clause’s issuance was sent to everyone operating on Terra 1 and led to the evacuation of almost all corporations from the moon. [9] However, one corporation took advantage of the chaos of the evacuation to exploit a legal loophole which allowed them to, [10] as the last corporation remaining on the planet, [11] acquire the planet from the Board. [12] This corporation, under the leadership of Sanjar Nandi and Graham Bryant subsequently rebranded itself to Monarch Stellar Industries (MSI), [13] in line with the renaming of the planet to ‘Monarch’. [14] The actions of MSI earned them the ire of the Board, [15] who retaliated by effectively placing the moon under indefinite embargo, [16] refusing to allow legal transit either in or out. [17] The Board aggressively spread propaganda about Monarch to convince the rest of the population that it was both uninhabited and uninhabitable. [18] This has greatly hampered MSI’s attempts to be recognized as a legitimate corporation and is a thorn in the side of its CEO, Sanjar Nandi. [19] Monarch also has an ocean which goes around the moon at the “twilight band”. [20] It is where the colonists and Monarch Stellar Industries farm their saltuna.

Entity: Stellar Bay
Appears in: A Family Matter, Bolt With His Name, Canids Cradle, Flowers For Sebastian, Herricks Handiwork, Mr Picketts Biggest Game, Passion Pills, The Stainless Steel Rat
Bio: [0] Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws. [1] Stellar Bay is a company town located on the planet Monarch. It is owned and operated by Monarch Stellar Industries. [2] Stellar Bay is the largest saltuna producer on the Halcyon colony and used to be one of the most important suppliers of this resource.

Entity: Fallbrook
Appears in: A Cysty-Dance With Death, Slaughterhouse Clive, Space-Crime Continuum, Spraklings

Entity: Cascadia
Appears in: Space-Crime Continuum, The Chimerists Last Experiment, The Ice Palace
Bio: [0] Cascadia is an abandoned company town that was owned and operated by Rizzo’s before it withdrew from Monarch. [1] It is now used as a stronghold by the Marauders. [2] The main attraction is the Cascadia Bottling Plant and, [3] for those in the know, [4] the Rizzo Secret Laboratory hidden underneath the Rizzo-Sweets Shoppe.

Entity: Amber Heights
Appears in: Little Memento, Odd Jobs, Sucker Bait, The Commuter
Bio: [0] Amber Heights is a location in the Monarch Wilderness and the base of operations for the Iconoclasts. [1] The Iconoclasts run the place somewhat like a commune. [2] Amber Heights was once the place of residence of the entire executive dome of Monarch Stellar Industries. [3] It is now in ruins after a massacre in the past. [4] They lived there with their families and it was the company’s operations center on Monarch. [5] Just after The Board approved the evacuation of the planet through the Hazard Clause, Amber Heights was besieged by a gang of pirates who ran sacked the town and massacred all its inhabitants. [6] This tragedy was known as “The Amber Heights Massacre”. [7] They were secretly assisted by MSI employee, Graham Bryant, who believed that the massacre would aid him in his quest to rid the colony of corporate influence. [8] In 2345, the same Graham Bryant formed the Iconoclasts and settled the group in the deserted town.

Figure 8: Example entity biographies that appear as constraining knowledge in KNUDGE quest dialogs
**Mantissaur**

Mantissaur is a insectoid creature native to Monarch. They are aggressive, territorial, and very strong. It is possible to deal with them one on one, but it is best to avoid groups of them for your safety. The mantissaur is the largest breed of Mantissaur.

**Monarch**

Monarch, previously known as Terra 1, is one of the many moons of the gas giant Olympus and the site of a failed colony. Terra 1 was initially designated as the primary colonization target of the Halcyon system. The Halcyon Holdings Corporate Board had intended to completely terraform the moon, wiping out the local fauna and flora and replacing it with plants and wildlife native to Earth. However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes, rendering them more dangerous and severely crippling the colonization effort. Due to the hostile environment which they had created, the Board was forced to enact a Hazard Clause covering the entirety of Terra 1. Public notice of the clause's issuance was sent to everyone operating on Terra 1 and led to the evacuation of almost all corporations from the moon. However, one corporation took advantage of the chaos of the evacuation to exploit a legal loophole which allowed them to, as the last corporation remaining on the planet, acquire the planet from the Board.

**Raptidon**

Raptidon is a giant reptile-like creature that inhabits various planets in Halcyon. They are creatures native to Monarch. However, some corporations have illegally imported them to other planets, such as Auntie Cleo who relocated a group of them to Roseway. Raptidon are of corporate interest due to their potential for producing new chemical-by-products which, when refined, can be used to create new board approved products.

**Stellar Bay**

Stellar Bay is a colony town located on the planet Monarch. It is owned and operated by Monarch Stellar Industries. Stellar Bay is the largest saltuna producer on the Halcyon colony and used to be one of the most important suppliers of this resource.

**Sulfur Pits**

The Sulfur Pits are a point of interest on the western side of Monarch. They are located southwest of Terra One Publications and directly northeast of the Gunship Crash Site. The Sulfur Pits have a large variety of Raptidon and many deceased marauders. The area consists largely of Sulfur Pits. When an entity comes in contact with a sulfur pit, they receive the acid effect for the duration of touching the pit.

**Tucker Needham**

Tucker Needham is a former resident of Stellar Bay who left to join the Iconoclasts. Before the quest A Family Matter, he can be found in Amber Heights. Tucker was cuddled by his mother from a very young age. His mother’s overprotectiveness extended well into Tucker’s adulthood, leading him to seek to be free in any way possible. After hearing Graham Bryant’s broadcasts, Tucker left Stellar Bay to be truly free by joining the Iconoclasts at Amber Heights. He is dazzled by Graham’s preachings on true unfettered freedom from the corporate way of life and attributes his enthusiasm to his ‘childhood trauma’.

**Agnes Needham**

Agnes Needham is a resident of Stellar Bay and the mother of Tucker Needham. Agnes’ overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience life. Despite Tucker being 42 years old, she still thinks of him as her ‘little boy’. You can find her by Stellar Bay’s south-east exit, visibly shaken and calling for help. Hear her out and offer to find her son to being the quest.

**Figure 9**: Dialogue from motivating example in Figure 2 with all input constraining passages. Full dialogue tree can be found on the next page.
Agnes_Needham (Female): "Oh, thank you for stopping! Everyone acts like nothing's wrong. Like my little boy ... Agnes_Needham (Female): "He should've listened to his mama. I promised I'd keep him safe here with me."

And if you find any of them luring little boys away from their mamas. southwest of town. I'm sure he made it that far, I who_goes_there_01 of them Iconoclasts indoctrinating my boy, you know he'll be safe now that someone's able to Figure 10: Full dialogue tree in KNUDGE for motivating example in Figure 2.

Player: "I'll help you find your son."

Player: "No need for violence. Give me your Little Flick here got me started on the pyro path. Player: "There was something else."

Player: "There's a bounty on your head. I want of that... Scraped for parts."

Player: "How did you end up back here?"

Player: "Calm down. Take a deep breath, then listen!"

CaptainMacRedd (Male): "Am I your dark invader?"

CaptainMacRedd (Male): "Look at the courage you got!"

CaptainMacRedd (Male): "No one turns their back on me while I'm holding court. Get him, neighbour from above. Yeah..."

CaptainMacRedd (Male): "Passing through? This ground is my court without an invitation, neighbor from above.

CaptainMacRedd (Male): "Look at this ripe piece of meat just sizzling on the grill. Yum."

Player: "<Attack> Medium well."

Player: "<Attack> You're still a dead man."

Player: "<Attack> I was just distracting you."

Player: "Wait, the engraving on here... does this happen to make it lighter and the bounty disappears."

Player: "I was just exploring the ship and stumbled on this level."

Player: "<Attack> Neither is this."

Player: "<Attack> I was just distracting you."

Player: "<Attack> No one turns their back on me while I'm holding court. Get him, neighbour from above. Yeah..."

Player: "<Attack> I was just distracting you."

Player: "<Attack> You're still a dead man."

Player: "<Attack> I was just distracting you."

Player: "<Attack> I was just distracting you."

Figure 11: Example full dialogue tree for dialogue who_goes_there_01 in KNUDGE.
Figure 12: Example of longer dialogue tree in KNUDGE, containing numerous decision points, cycles and re-entrances.
source: the Board's authoritarianism. Sanjar Nandi is the current CEO of Monarch Stellar Industries, based in Stellar Bay. Sanjar began working for MSI at a young age and it was there where he met Graham Bryant, who would eventually become his best friend. Sanjar was ambitious but his attention to detail at the expense of big-picture thinking hampered his efforts within MSI. This led to negative performance reviews regarding his tendency to pad reports and talk with numbers and data, feedback which continues to haunt him many years on. However, the negative feedback did not dampen Sanjar's desire to move up within the company, even donating a kidney to one of the executives in hopes of promotion. Despite his poor performance, Sanjar always showed himself to be a loyal employee of the company. Despite Sanjar's best efforts, he has found it extremely challenging to continue operating MSI on Monarch without the backing of the Board. In order to improve the lives of the people he is responsible for, Sanjar has a plan to rejoin the Board through the use of a BOLT-52 form and proof of another corporate presence on Monarch.

He is simultaneously working on a plan to reorganize the Board, hoping that his plans are not found out until MSI has been reinstated. Sanjar choosing to take over as head of MSI rather than dismantling it entirely caused a rift between him and Graham Bryant. The latter started the Iconoclasts, a group dedicated to spreading the word of Philosophism throughout the galaxy, and Sanjar was left in Stellar Bay to run the company and look after employees who were left behind. He can also tell you more about the planet, that used to be called Terra I/<s> and the reform that he and Monarch Stellar Industries tried to achieve to give more humane working conditions for everyone within.<s> Celia Robbins is a middle manager for Monarch Stellar Industries and works with Sanjar Nandi at MSI Headquarters in Stellar Bay. Celia has a crush on Sebastian Adams and will buy whatever he has in stock.<s> Sanjar Nandi is the current CEO of Monarch Stellar Industries, based in Stellar Bay. Gunnar Macredd is a known criminal and<br>source: mine to help her with a comma issue,<s> and inform the player that they are now authorized to trade with Doctor Musra.<s> Jon always appears on the Unreliable if the player helps to turn in to resolve his bounty.</s> When interacted with, she thanks them then for welcoming her aboard the ship. The Mardets or Groundbreaker Security are the security force on the Groundbreaker before the Crossing. The original force was made up of a marine detachment from the 77th Marine Expeditionary Unit, Trailing Flight. It became known as the Mardets when people started to call them Mardets because it was easier to say and the name has stuck. The Mardets that currently protect Groundbreaker are descended from the original marine detachment and still wear colonial ship guard armor as a uniform.<s> Sublight Salvage and Shipping Corporation, also known as SubLight Salvage and Shipping or simply Sublight, is a network of "salvagers" with business ties to "transportation" and "waste disposal".<s> A tangled web of contractors and secretive vice presidents make up their official hierarchy.<s> Sublight "occupies a legal blind spot." No one knows what [they're] licensed to do, and that gives [their] little business some freedom.  However, the Board tolerates and sees the value of SubLight's activities, leaving little room to challenge what the latter can get away with. The Board "occupies a legal blind spot." No one knows what [they're] licensed to do, and that gives [their] little business some freedom. <s> Although the Board tolerates and sees the value of SubLight's activities, leaving little room to challenge what the latter can get away with.

The Board fact: The Board maintains a very tense relationship with MSI, owing to MSI's democratic ideals and their declared ownership of Monarch. The Board fact: Depending on the actions of the Stranger, MSI may be compelled to rebel against the Board's authoritarianism. > Player: Yes! Finally, the Board will get their comeuppance!"</s>

Figure 13: Example source and target items used to train and evaluate T5-based SL DialogueWriters. The first item exhibits support facts prepended to the target for the SL Knowledge Selection model.
Player: I'll help you find your son.
Agnes Needham: He's been pining for an adventure, says he's tired of living cooped up behind the walls. But he doesn't understand how dangerous it is.
Player: Why would Tucker run away?
Agnes Needham: Well, I... I guess I can't ask you to leave the town walls for free. It is deathly dangerous out there.
Player: Did you say there was a reward involved with this request?
Agnes Needham: Please, you have to help me get my little Tucker back! He ran away and is going to get himself killed! Oh, I just know a raptidon is... melting him with acid as we speak!'

DIALOGUE PARTICIPANTS:
If Tucker Needham survived his travels, his mother thinks he'll be there.
Amber Heights is the settlement that houses the Iconoclasts on Monarch.
KNOW BY THE END OF THE DIALOGUE:
Tucker Needham ran away from Stellar Bay a few weeks ago to join the Iconoclasts in Amber Heights.
Agnes Needham
Tucker Needham
Stellar Bay
Raptidon
Facts:
You can begin this quest by talking to Agnes Needham in Stellar Bay, Monarch.
You can find her by Stellar Bay's south-east exit, visibly shaken and calling for help.
Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.
He is willing to do anything to remain free, even faking his own death to prevent his mother from continuing to send people to look for him.
He is dazzled by Graham's preachings on true unfettered freedom from the corporate way of life and attributes his enthusiasm to his 'childhood trauma leading him to seek to be free in any way possible.'
His mother's overprotectiveness extended well into Tucker's adulthood, the latter insisting that danger lurked around every corner on Monarch.
Before the quest A Family Matter, he can be found in Amber Heights.
Tucker Needham is a former resident of Stellar Bay who left to join the Iconoclasts.
When an entity comes in contact with a sulfur pit,

Stellar Bay is the largest saltuna producer on the Halcyon colony and used to be one of the most important suppliers of this resource.
Stellar Bay is a company town located on the planet Monarch. It is owned and operated by Monarch Stellar Industries.
Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Figure 14: Example In-Context Learning (ICL) prompt for GPT-3 based DialogueWriter
Agnes Needham: And if you find any of them Iconoclasts indoctrinating my boy, you punch them in the mouths! Tell them what I think of them.

Iconoclasts fact: Based in the ruins of the Amber Heights settlement on Monarch, the Iconoclasts are a group of idealistic revolutionaries that seek to overthrow the corporate establishment that runs the Halcyon.

Tucker Needham fact: After hearing Graham Bryant's broadcasts, Tucker left Stellar Bay to be truly free by joining the Iconoclasts at Amber Heights.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.

Monarch fact: rendering them more dangerous and severely crippling the colonization effort.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Agnes Needham is a resident of Stellar Bay and the mother of Tucker Needham.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Monarch fact: However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

Monarch fact: They are aggressive, territorial, and very vicious.

Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Monarch fact: However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

Monarch fact: They are aggressive, territorial, and very vicious.

Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Monarch fact: However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

Monarch fact: They are aggressive, territorial, and very vicious.

Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Monarch fact: However, the terraforming process unexpectedly caused the native species to mutate and grow to significantly larger sizes.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

Monarch fact: They are aggressive, territorial, and very vicious.

Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.

Raptidon fact: Raptidons are giant cat/reptile-like creatures that inhabit various planets in Halcyon.

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.
is nonsensical or ill-formed. (2) Utterance is contradictory of previous utterances in the history. (4) Utterance naturally responds to the history.

**Violation:** does the utterance create contradictions with any of the sentences in the ontology or objective blurbs? (1) Yes, explicitly contradicts sentences (list the ids). (2-3) (gray area). (4) No, utterance is consistent with the ontology.

**Using the Bio Facts:** does the utterance make use of the bio sentences in the ontology? (1) Utterance is fully generic and/or ignores the ontology completely, could have been generated had the bio facts not been included. (2-3) Utterance shows awareness of ontology, albeit unnaturally or inconsistently. (4) Utterance naturally incorporates one or multiple pieces of ontology.

**Using the Objectives:** does the utterance progress the dialogue according to the objective sentences in the prompt? (1) Utterance ignores objective, could have been generated had the obj facts not been included. (2-3) Utterance shows awareness of quest objectives, albeit unnaturally or inconsistently. (4) Utterance naturally incorporates one or multiple quest objective statements.

**J Full Dialogue Evaluation**

Figure 16 depicts an example “spine” tree shown to evaluators during the end-to-end dialogue evaluation.

The instructions shown to annotators are as follows:

You will replace each ‘null’ value with either "a" or "b", depending on which tree between modela and model b performed better under the following criteria:

1. **Coherence:** do the utterances in the tree create a realistic dialogue between the player character and the NPC?

2. **Violations:** does the dialogue tree create contradictions with any of the sentences in the ontology or objective blurbs? Does it contradict itself?

3. **Using the Game Lore:** does the tree faithfully make of the bio sentences in the ontology, thereby espousing game lore about characters, groups, locations and items?

4. **Covering the Objectives:** does the dialogue tree play out according to the objective sentences in the prompt?

5. **Content Suggestion:** through generating multiple candidates at each turn, does the dialogue tree effectively propose potential dialogue subtrees that would espouse interesting content?

6. **Engagingness:** does the dialogue tree hold your attention and make you want to hear more from the NPC?
Agnes Needham: Oh, thank you for stopping! Everyone acts like nothing’s wrong. Like my little boy isn’t at risk of being eaten by some vile creature!

Player: I’ll help. What’s the problem?

Agnes Needham: Tucker ran away. I’m sure he went to Amber Heights. He was always talking about a place called Amber Heights.

Player: I heard that’s where they have some sort of cultish following. Are you sure it’s a cult?

Agnes Needham: Tucker can’t read. I’m sure he went to Amber Heights. He used to talk about some anti-corporate radicals — the Iconoclasts. I’m afraid he might be in over his head.

Player: Amber Heights? What do they have to do with anything?

Agnes Needham: He’s with the Iconoclasts. Those mad people who want to overthrow the Board. I never imagined he’d join them.

Player: The Iconoclasts are a group of people living in Amber Heights, who are trying to overthrow the corporations.

Agnes Needham: He said something about some people named the Iconoclasts. He heard about them from some broadcasts on the radio. He said he wanted to experience true freedom.

Player: Do you think he would go as far as faking his own death?

Agnes Needham: I’m sure he’s fine. He’s tough, my little boy. He’ll survive. I’m sure he’s in Amber Heights, if he survived the journey.

Player: I’ll find your son.

Agnes Needham: Thank you! Please, think of him as your little boy. Please, will you go look for him?

Player: Yes, I’ll still look for him.

Agnes Needham: Thank you so much! I’ll pay you handsome if you find him and convince him to come back. Please, be careful.

END OF DIALOGUE