Ontologically Faithful Generation of Non-Player Character Dialogues

Anonymous ACL submission

Abstract

We introduce a language generation task grounded in a popular video game environment. **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: (1) dialogues are branching trees as opposed to linear chains of utterances; (2) utterances must remain faithful to the game lore—character personas, backstories, and entity relationships; and (3) a dialogue must accurately reveal new quest details to the human player. 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). A key challenge in authoring these NPC dialogues is maintaining enjoyable as well as coherent narratives: utterances must faithfully reflect quest structure and game lore—characters, histories, and entity relationships.

**Figure 1** depicts a dialogue turn taken from *The Outer Worlds* (Obsidian Entertainment, 2019), an action RPG renowned for its narrative and dialogue writing. This turn demonstrates how a dialogue relies on the descriptions of entities in the game world while also revealing relevant quest information. The NPC’s utterance not only begins and provides backstory about a new side quest,\(^2\) but also interacts according to a well-formed persona (a worried parent) and references an adversary (rap-tidons) that the player will face later in the quest. NPC interactions often take the form of complex utterance trees. Creating these branching structures according to the many specifications of dialogue writing can be time-consuming for a game designer (Caropreso et al., 2012), and costs millions of dollars (see §A for discussion). This motivates the pursuit of automatically generating dialogue trees.

To the best of our knowledge, there does not exist a public dataset that meets our desired criteria: a set of real game-quality NPC dialogues paired with granular quest and biographical specifications consistent with a well-formed game ontology. \(^*\)

\(^2\)RPG side quests are goals provided to a player in order to enrich gameplay, while not being part of the central storyline.

**Figure 1**: An example non-player character (NPC) dialogue from *The Outer Worlds* (Obsidian Entertainment, 2019). NPCs must speak faithfully to a granular ontology of quest specifications and game lore.
recent work towards lore-conditioned 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) do not address complex dialogue trees and the interweaving narratives found in deployed RPGs.

We make the following contributions via this paper: firstly, we introduce KNUDGE: Knowledge-constrained User-NPC Dialogue Generation, a set of dialogue trees (in English) derived from an existing video game and paired with granular ontological constraints. We extract the trees directly from the game data of The Outer Worlds. This game’s side quests share overlapping characters and locations, making it an appealing study in the development of automatic dialogue generation tools. For each side quest, we enumerate the relevant dialogues and annotate them at multiple levels of specificity (down to the individual node) with quest- and lore-related natural language (NL) support facts pulled from fan-written wikis. KNUDGE contains 159 dialogues from 45 side quests that all take place in The Outer Worlds. It contains 4.7K utterances and an average of 1.3K input constraint tokens per dialogue. The complexity of the dialogues, the annotation level, and the amount of specifications are greater than any comparable dataset.

Secondly, we introduce the challenging task of knowledge-constrained NPC dialogue generation. Given a set of ontological specifications, we want to generate fluent dialogue trees that reveal the quest objectives while staying faithful to the specified context and the game lore. These specifications make up hundreds of statements per dialogue, posing a difficult challenge for modern generation models. Additionally, we introduce a class of model, termed DialogueWriters, that leverage neural language models (LMs) such as GPT-3 (Brown et al., 2020) to generate an utterance conditioned on an existing partial dialogue tree and ontology passages. To encourage the use of game lore to produce interesting and engaging dialogue, we experiment with adding mechanisms to sub-select relevant facts before generating each utterance.

Finally, we prescribe protocols for the evaluation of models for the NPC dialogue generation task. We test for 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 quests from the game, as well as for a pair of never-before-seen quests written by a professional game designer to target real quest writing. Our experiments reveal further room for improvement in aspects such as joint reasoning over multiple facts from ontology, and coverage of all quest objectives. We hope that KNUDGE will facilitate further progress on faithful game dialogue generation, towards tools that can assist developers with their craft.

2 Task Definition

We define the task of knowledge-constrained 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 ob-
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 tree $D$ 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 at player turns (see Figure 3, right). Dialogue trees have one start node, but can have multiple exit nodes and can contain cycles.\(^4\)

3 Data

KNUDGE comprises NPCs dialogue trees from all 45 side quests in the Outer Worlds base game. Our dataset construction entails gathering information about each quest $(Q)$ and the Outer Worlds entities that appear or are referenced during the associated dialogues $(B)$ (§3.1), and then extracting trees $(D)$ from the game data semi-automatically (§3.2). We then analyze the resulting data and compare to related datasets (§3.3).

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 often completed, and new ones are introduced, during an NPC dialogue. A detailed quest anatomy can be found in appendix Figure 6. 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. Thus, 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.\(^5\)

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 some skill to pass a “check.”\(^6\) 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 special state-related conditions. Special edges were then added manually depending on whether conditions are relevant to the quest. Example trees can be found in §E.

**Annotating Utterance Nodes with Support Facts**

We coordinated with professional data annotators to annotate the nodes of the dialogue trees with support facts from the associated quest and lore constraints. All annotations were done by native English speakers, who were provided with a set of instructions shown in §I. We 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.

\(^4\) Although they often contain cycles these structures are still colloquially referred to as “trees.”

\(^5\) We do not associate its walkthrough passage, since the NPC should only be expected to convey new objective information that the player will actually see in game.

\(^6\) E.g. if the Player has above 55 points of the Persuade conversational skill, they can convince Tucker Needham to return to his mother in the quest from Figure 1.
Figure 3: Overview of dialogue node annotation with support facts from quest and biography passages.

Table 1: KNUDGE Dataset Statistics

<table>
<thead>
<tr>
<th>Quests</th>
<th>45</th>
<th>Dialogue trees</th>
<th>159</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entities</td>
<td>168</td>
<td>Facts per dialogue</td>
<td>83.3</td>
</tr>
<tr>
<td>Characters</td>
<td>81</td>
<td>Entity facts</td>
<td>73.3</td>
</tr>
<tr>
<td>Locations</td>
<td>40</td>
<td>Objective facts</td>
<td>9.9</td>
</tr>
<tr>
<td>Groups</td>
<td>21</td>
<td>Fact tokens per dialogue</td>
<td>1321.8</td>
</tr>
<tr>
<td>Items</td>
<td>18</td>
<td>Utterances per dialogue</td>
<td>29</td>
</tr>
<tr>
<td>Creatures</td>
<td>7</td>
<td>NPC utterances</td>
<td>16.7</td>
</tr>
<tr>
<td>Facts per entity</td>
<td>7.4</td>
<td>Player utterances</td>
<td>12.3</td>
</tr>
<tr>
<td>Entities per quest</td>
<td>8.0</td>
<td>Utter. tokens per dialogue</td>
<td>406.6</td>
</tr>
<tr>
<td>Quests per entity</td>
<td>2.0</td>
<td>Facts per NPC utterance</td>
<td>1.0</td>
</tr>
<tr>
<td>Facts per objective</td>
<td>7.6</td>
<td>Entity Facts</td>
<td>0.57</td>
</tr>
<tr>
<td>Objectives per quest</td>
<td>4.5</td>
<td>Objective Facts</td>
<td>0.42</td>
</tr>
</tbody>
</table>

We compute an average exact match (EM) score of 0.52 and an average 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 of elements.

3.3 Dataset Analysis

Table 1 describes overall statistics for the extracted dialogue trees and annotations. We note the high frequency of support fact annotations per utterance; 47% of all nodes (57% of NPC nodes) in the dataset are annotated with at least one fact. Amongst NPC nodes, for which support knowledge is more common, there is an average of 1.0 facts per utterance. Dialogues have on average 9.9 quest facts and 73.3 biographical facts that generation 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.

3.4 Challenges

Writing RPG-quality dialogue trees is difficult for human developers for its many interweaving considerations. KNUDGE poses a similarly multi-faceted challenge to generation models:

- The tree must serve its quest function, containing input-specified player utterance options, NPC responses (including possibly specified emotions), and pieces of information the player must learn by the end. These NL specifications have highly diverse semantics (see §C).
- The generated utterances must be coherent, realistic, and engaging to the player.
- The NPC should embody the persona described in their biography passage, which describes personality, history, and relationships.
- To facilitate world building, the NPC should expose details about other entities whenever it is contextually relevant, and should never violate the ontology through contradiction.

Generating a 30-node (or larger) NPC dialogue tree while taking into account all of these criteria at once 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 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 such as few-shot examples.

4 DialogueWriter Methods

We introduce a set of neural methods for generating candidate utterances given the ontological specifications ($Q$, $B$, and $P$) from §2 and a partial subtree $S$ for a dialogue item in KNUDGE. We refer to these methods collectively as DialogueWriter models, which propose utterance nodes at a specified new location branching off the subtree. Formally, given some “most recent” node $n \in S$, a...
We also train a version of the SL baseline that learns (and thus truncate last) the biographies of dialogue written complex dialogue trees from an actual RPG, with more utterances, longer constraint passages, and higher-granularity annotations (of individual utterances) than other game dialogue datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Writer</th>
<th>Source</th>
<th>Structure</th>
<th>Tokens / Item</th>
<th>Constraint Tokens / Item</th>
<th>Narrative and Bio Constraints</th>
<th>Annotation Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Story Generation Datasets</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STORIUM (Akoury et al., 2020)</td>
<td>Crowd</td>
<td>Online story writing game</td>
<td>Sequence of scene entries</td>
<td>19k</td>
<td>1.2K</td>
<td>Scene intro, challenge, location, and character descriptions</td>
<td>Story</td>
</tr>
<tr>
<td>TV-STORYGEN (Chen and Gimpel, 2021)</td>
<td>Crowd</td>
<td>Fan wikis</td>
<td>TV episode recap</td>
<td>1.8k</td>
<td>25.9K</td>
<td>Brief episode summary and character wiki articles</td>
<td>Scene Entry</td>
</tr>
</tbody>
</table>

**RPG Dialogue Datasets**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Writer</th>
<th>Source</th>
<th>Structure</th>
<th>Tokens / Item</th>
<th>Constraint Tokens / Item</th>
<th>Narrative and Bio Constraints</th>
<th>Annotation Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT (Urbanek et al., 2019)</td>
<td>Crowd</td>
<td>text game platform</td>
<td>Sequence of utterances</td>
<td>212 (12 utt)</td>
<td>276</td>
<td>Location description, persona statements and held objects</td>
<td>Dialogue</td>
</tr>
<tr>
<td>TorchLight II (van Stegeren and Thorne, 2020)</td>
<td>Professional</td>
<td>RPG Data</td>
<td>Sequence of quest stages with 0 or 1 utterances</td>
<td>157 (3 utt)</td>
<td>24</td>
<td>Quest title, objectives, and details</td>
<td>Quest</td>
</tr>
<tr>
<td>WoW (van Stegeren and Mysliwiec, 2021)</td>
<td>Professional</td>
<td>RPG data</td>
<td>NPC-uttered quest description</td>
<td>61 (1 utt)</td>
<td>15</td>
<td>Quest title and objective</td>
<td>Quest</td>
</tr>
<tr>
<td>KNUDGE (Ours)</td>
<td>Professional</td>
<td>RPG data</td>
<td>Complex quest dialogue tree</td>
<td>407 (29 utt)</td>
<td>1.3K</td>
<td>Quest title, objectives, and details, walkthrough and entity wiki articles</td>
<td>Utterance</td>
</tr>
</tbody>
</table>

Table 2: Comparison of KNUDGE to knowledge-constrained generation datasets. KNUDGE contains professionally-written complex dialogue trees from an actual RPG, with more utterances, longer constraint passages, and higher-granularity annotations (of individual utterances) than other game dialogue datasets.

DialogueWriter maps inputs $(Q, B, P, S, n)$ to a list of candidate utterances $[c_1, \ldots, c_n]$ such that there is a directed edge $n \rightarrow c_i$.

**Tree Linearization** We consider language models (LMs) that accept linear input token sequences. We thus devise a traversal mechanism that converts a 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. With this linearization mechanism, we need only train a next utterance generator that conditions on linear history $H$. Methods to do so are explored below.

### 4.1 Supervised Learning (SL) Models

We fine-tune a T5-large sequence-to-sequence 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, removing components of $B$ first. We list last (and thus truncate last) the biographies of dialogue participants. §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 $[f_1^{(1)}, \ldots, f_m^{(1)}]$, $c_i$ s.t. $f_j^{(1)} \in Q \cup B$.

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

As KNUDGE is relatively small, 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.
We split $K$ oracle which conditions on the full gold knowledge annotations we compare against ablations to the ICL model: a vanilla model that conditions only on the participant $P$ and utterance history $H$, and a quest only model that conditions on $P$, $H$, and $Q$, but not $B$.

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 vanilla 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.

### 5.2 Next Utterance Prediction (NUP)

We split NUDGE into train, development, and test sets 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). We evaluate utterances generated from a test set of 1800 partial trees.

#### Automatic Evaluation

We use reference-based metrics BLEU-4 (Papineni et al., 2002) and re-scaled BERTScore-F1 (Zhang et al., 2020). We evaluate generated next utterance against the following single- and multi-reference sets: 1) the gold utterance $n_4$, 2) the quest statements in $Q$, and 3) the biography statements in $B$. We evaluate the gold utterance to show a performance upper bound.

#### Human Evaluation

In coordination with a data specialist, we conducted 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 with guidance for each Likert scores in §1. Results were verified via bootstrap testing.

### 5.3 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.\(^3\) 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 hand-selected a set of 8 test dialogues from the game with varying quest roles, e.g. starting, continuing, and ending quests. Additionally, we had a professional game designer write specifications for 2 totally new quests with overlapping entities from the original game; from these we constructed 8 more test items.

We generated trees using the KS, regular, quest only, and vanilla ICL DialogueWriters, then

\(^3\)This serves to specify utterance format in cases of zero-shot learning, which is the case for all designer-written items.
showed them to data specialists for evaluation following the ACUTE-Eval (Li et al., 2019) protocol of pairwise comparison. Annotators were asked to select 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?

### 6 Results and Discussion

**Next utterance prediction** results under automatic and human metrics are shown in Table 3 and Table 4. We note that automatic metrics for generation that check for lexical or semantic overlap with a set of references is not directly suited for evaluating generations in KNUDGE; this can be seen from the extremely low performance of the gold utterances themselves under these metrics. We find that professionally written utterances do not always have high overlap with knowledge statements themselves. Gold utterances also do not score perfectly under utterance-level human evaluation of $Q$ and $B$ usage, as not every real-world utterance reflects the ontology. It might instead provide other qualities like realism and fluency. This highlights the challenge of identifying and then pursuing desiderata of RPG-quality dialogue trees.

We observe that the best-performing model under overlap metrics with $B$ and $Q$ are the T5-based SL models. This reflects that these models have learned to copy spans directly from the context into the generation, hence scoring highly on ngram-based overlap with it while scoring poorly on gold utterance overlap and human evaluation.

Table 3 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. The Quest-only ICL DialogueWriter can generate coherent utterances that reflect the quest specifications; however, as illustrated by Table 4 and Figure 4, this can sacrifice lore references.

Table 4 also shows that all fully $B$ and $Q$-conditioned ICL models perform equivalently under all human metrics except (non)violation, which KS models perform a decimal point worse at. 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. We note that no model approaches gold performance at reflecting $B$.

**Qualitative Analysis** Figure 4 depicts example outputs by models on an NUP example. We highlight cases in which the models succeed at the desiderata that we strive for in KNUDGE: to convey quest and lore specifications naturally through the interaction. However, we see that SL models and ablated ICL models are less successful. We observe that the gold utterance is more infused with desirable information than any generation; it references the quest’s next location and numerous adversaries that the player will run into, while effectively reflecting the NPC’s overprotective parent persona. This highlights a performance gap between neural
We note that not reaching human performance does not preclude DialogueWriters from being useful to writers, as they can still be used to suggest new directions for dialogues to be verified or modified in a human/AI collaborative writing process.

Full dialogue generation case study results are found in Table 5. Model performance is measured as the rate at which annotators selected its generated tree in a pairwise comparison under the 6 criteria listed in §5.3. 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 (enough to fill out the GPT3 context window), without which there might not be a sufficient signal for the Quest only model to generate desirable utterances. Another reason might be that the few-shot learning benefits of the Quest only model not conditioning on B are lessened by the domain shift from Knudge to the designer’s writing.

<table>
<thead>
<tr>
<th>n = 16</th>
<th>Coh.</th>
<th>Viol.</th>
<th>Use B</th>
<th>Use Q</th>
<th>Content</th>
<th>Engaging</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>
<td>37.5</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>
<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>
<td>68.8</td>
</tr>
<tr>
<td>ICL-Vanilla</td>
<td>12.5</td>
<td>0.0</td>
<td>6.2</td>
<td>0.0</td>
<td>31.2</td>
<td>18.8</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.

and human writers to be addressed by future work. We note that not reaching human performance does not preclude DialogueWriters from being useful to writers, as they can still be used to suggest new directions for dialogues to be verified or modified in a human/AI collaborative writing process.

8 Conclusion

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

In this paper, we introduce 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 triple-A game (i.e. high-budget, high-profile game) development studio, thereby exemplifying real-world complexities in NPC dialogue authoring. We illustrate that LMs are indeed able to generate fluent dialogues that relate to provided game lore. However, the straightforward application of such technologies does not match the quality of professional game writers. We hope that Knudge aids the development and evaluation of new techniques for faithful game dialogue generation.

7 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 Stegemen 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 Stegemen and Myśliwiec, 2021), dialogue generation (Si et al., 2021), persona-specific agents in text environments (Urbanek 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.
Limitations

One limitation of our work is that our considered models use simplifications to model the complex branching trees by linearizing the nodes. We moreover only evaluate next-utterance prediction and relatively simple end-to-end dialogue trees which do not approach the complexity of the actual trees. More can be done on this front in future work to develop methods that generate more complicated tree structures. We also leave for future work the evaluation of whether end-to-end generated dialogues “check all the boxes” of quest requirements, completing all user-provided specifications.

We also 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. Thus, we provide only a partial solution to the problem of automatic NPC dialogue generation, and 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. It is 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


Appendix

A Cost of Authoring NPC Dialogues

Outer Worlds has 10 narrative/design credits\(^8\), 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,\(^9\) a gamer-focused site containing fan-made walkthroughs for many popular RPGs. Game entity biographies were collected from Fandom.\(^{10}\) The biography passage for a given entity is the same across all quests in which the entity appears in NUDGE, 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.\(^{11}\) All wiki articles were written in English by site users.

C Quest Anatomy and Example Items

Figure 6 provides a detailed anatomy of a NUDGE 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.

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 specifications. Figure 10, Figure 11, and Figure 12 depict example dialogue trees.

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 description spoken by an NPC. Here collect data for the game TorchLight II contains quest datapoints with a limited number dialog utterances per quest with no multi-turn interactions or trees.\(^{12}\) 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

---

8\(\text{https://www.imdb.com/title/tt9417446/fullcredits}\)
9\(\text{https://www.theouterworlds.wiki.fextralife.com}\)
10\(\text{https://theouterworlds.fandom.com}\)
11\(\text{https://www.youtube.com/@l0rdmatrim}\)
12\(\text{Table 2 describes statistics for the 82 TorchLight II quests that contain both objective annotations and dialogue lines.}\)
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.

**Example Prompt Constructions for DialogueWriter Model**

Figure 5 depicts an overview of our tree linearization and prompt construction method.

**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. 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 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

---

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.
Objective 1: Look for Tucker Needham in Amber Heights

Game Log: [0] Amber Heights is the settlement that houses the Iconoclasts on Monarch. [1] If Tucker Needham survived his travels, his mother thinks he’ll be there.
Walkthrough: [0] Head South from Stellar Bay and follow the east road. It will take you to Amber Heights. [1] Head up the hill and go into a residence on the left to meet Tucker Needham.

Objective 2: Convince Tucker to Return Home

Game Log: [0] Now that you’ve found Tucker Needham in Amber Heights, convince him to return home to his mother in Stellar Bay.
Walkthrough: [0] Introduce yourself, and then you can mention your surprise that this grown man is the “little boy” that ran away. You’ll earn 7500xp [1] Explain to him that the place he made it sound as if he was a boy in danger, [2] and he’ll say she has been overprotective all her life, [3] and he is ready to live his life without her protection. [4] You can persuade (55) or intimidate (55) to expedite things and get him to go back, [5] or you can ask him what he wants to do about it. [6] The last option will have him tell you to report that he is dead. [7] You can express your concern about what that will do to Agnes, [8] and then either ask for something that would prove a body, or reject the proposition. [9] If you persuade him to go back, you’ll get 7500xp and can return to Stellar Bay to see things play out.

Objective 3: Return to Agnes Needham in Stellar Bay

Game Log: [0] You convinced Tucker Needham to return home to Stellar Bay. [1] Agnes promised a reward for bringing her son back.
Walkthrough: [0] You’ll find his mother is still condescending to him, [1] and you can help him by saying he’s a grown man. [2] You’ll get 7500xp. [3] If you stick around and talk to them some more you’ll see Tucker is standing up for himself. [4] You’ll receive 625 Bit Cartridge, Monarch Stellar Industries Reputation and 15000xp.

Objective 3: Return to Agnes Needham in Stellar Bay

Game Log: [0] You convinced Tucker Needham to return home to Stellar Bay. [1] Agnes promised a reward for bringing her son back.
Walkthrough: [0] You’ll find his mother is still condescending to him, [1] and you can help him by saying he’s a grown man. [2] You’ll get 7500xp. [3] If you stick around and talk to them some more you’ll see Tucker is standing up for himself. [4] You’ll receive 625 Bit Cartridge, Monarch Stellar Industries Reputation and 15000xp.

Quest Name: The Commuter

Synopsis: [0] The Iconoclasts are due to receive a shipment of vital supplies from Carlotta, a sympathizer that resides in Stellar Bay. [1] The meeting is set to occur at the Bayside Terrace warehouse.
Walkthrough: [0] The quest can be obtained by asking Graham if there is anything that needs doing. [1] He is trying to get an old printing press running, but the replacement rollers he’d requisitioned haven’t arrived yet. [2] They were supposed to be delivered by Huxley, but she is still recovering and unable to make the delivery. [3] Graham asks the player to meet the supplier in her stead, and to pick up high-capacity data cartridges with the funds left over from the previous shipment. [4] Zora will intercept to ask the player to buy food and medicine instead with the leftover money.

Objective 1: Get the Printing Press Rollers from Carlotta

Game Log: [0] Travel to the warehouse at Bayside Terrace and find Graham’s contact, Carlotta. [1] She should have a shipment for him. Retrieve it. [2] Speak to Carlotta
Walkthrough: [0] Clear out the Sublight squad that is hunting Carlotta [1] Carlotta is behind a locked door to the east. [2] Activate the intercom next to the door to speak to her and she will unlock it. [3] Go inside and speak to her again to obtain the rollers needed to complete the quest, then choose between the high-capacity data cartridges or food and medicine.

Objective 2: Get High-Capacity Cartridges or Extra Supplies from Carlotta

Game Log: [0] Graham wants to tack on some high-capacity cartridges to his order, but Zora would prefer it they could get extra food and medical supplies. [1] You got extra supplies for Zora (or) You got High-Capacity Data Cartridges for Graham.
Walkthrough:

Objective 3: Return to Graham

Game Log: [0] Bring the needed parts back to Graham at Amber Heights
Walkthrough: [0] Return to Graham and you’ll find him arguing with Zora about the Van Noyz, a unit of the Iconoclasts that is MIA. [1] Inform Graham that you got his rollers, and food and medicine if that was your choice. [2] You’ll receive 7500xp and Zora will ask when the next drop is. [3] Inform her that Sanjar has made it illegal to trade with the Iconoclasts.

Quest Name: Who Goes There

Synopsis: [0] The Groundbreaker’s Mardets have a bounty for a criminal on the run in the Groundbreaker’s Back Bays. [1] You’ve agreed to hunt down the unlawful Captain Gunnar MacRedd. [2] Return his lighter to Commandant Sanita to claim the bounty.
Walkthrough: [0] This quest is obtained at Groundbreaker, [1] by speaking to Comdt. Sanita or perusing the bounty board.

Objective 1: Hunt Down and Kill Captain McRedd

Game Log: [0] Based on the bounty listing, Captain McRedd was last sighted in the Back Bays. [1] Head there and take him out.
Walkthrough: [0] You can find Captain McRedd in the Back Bays area of the Groundbreaker. [1] To get there head down the elevator in the promenade, [2] and you can’t miss him. [3] You can pass a Persuade (40) check to get him to put his gun down, [4] otherwise you’ll have to kill him and all his guards. [5] If you kill him he drops the Unique Weapon: Montag. [6] You’ll get 6000xp and MacRedd’s Lighter. [7] If you persuaded him, use Perception to note it says “Sanita” on the lighter. [8] MacRedd will mention it was given to him by Sanita in remembrance of a ‘carnal understanding’ they had a few years back.

Objective 2: Claim the Bounty’s Reward from Comdt. Sanita

Game Log: [0] McRedd gave you his lucky lighter to give to Sanita. [1] Go turn it in to resolve his bounty.
Walkthrough: [0] Turn the lighter in to Commandant Sanita to claim the bounty.
<table>
<thead>
<tr>
<th>Entity: Agnes Needham</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: A Family Matter</td>
</tr>
<tr>
<td>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 Needham, 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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Tucker Needham</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: A Family Matter</td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Raptidon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: A Family Matter, At Central, Bolt With His Name, Journey Into Smoke, Makes Space Suits Went Travel, The Amateur Alchemist, The Distress Signal, The Doom That Came To Roseway, Vulcans Hammer</td>
</tr>
<tr>
<td>Bio: [0] Raptidons are giant cæteripile-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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Sulfur Pits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: A Family Matter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Monarch</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Stellar Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>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</td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Fallbrook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: A Cysty-Dance With Death, Slaughterhouse Clive, Space-Crime Continuum, Spratkings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Cascadia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: Space-Crime Continuum, The Chimistress Last Experiment, The Ice Palace</td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity: Amber Heights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appears in: Little Memento, Odd Jobs, Sucker Bait, The Commuter</td>
</tr>
<tr>
<td>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 beset by a gang of pirates who ransacked 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.</td>
</tr>
</tbody>
</table>

Figure 8: Example entity biographies that appear as constraining knowledge in KNUDGE quest dialogs
Figure 9: Dialogue from motivating example in Figure 2 with all input constraining passages. Full dialogue tree can be found on the next page.
Figure 10: Full dialogue tree in KNUDGE for motivating example in Figure 2.

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

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 reinstalled. Sanjar choosing to take on 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. Sanjar 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. The reform that he and Monarch Stellar Industries tried to achieve to give more human workers conditions for everyone within.

Celia Robbins

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 (even if it’s just Ramen soup). When she left her apartment is filling up with junk and her neighbors are starting to complain about the smell. She is not concerned that she and Sebastian may not have much to talk about, as everyone else in Stellar Bay either smells like saltuna or are her boss.

The Stranger can offer to set her and Sebastian up on a date. DIALOG CONTEXT: Sanjar believes another company may be operating on Monarch illegally. If he can get proof, then he could use that as leverage to get MSI reactivated to the Halcyon Corporation.

Talk to Sanjar after completing BOLT with His Name. He reveals the plan is to blackmail the Board into letting them back in the table.

Sanjar will reveal he believes another corporation is operating illegally within Monarch, granting you the quest Error's Unseen. He will tell you Sanita is likely supplying them from Fallbrook.

Be want you to infiltrate the secret facility and bring back evidence - be it an item or staff. KNOW BY THE END OF THE DIALOG: Sanjar believes Sanita would know where this other corporation is operating. See if you can get the location from her. DIALOG PARTICIPANTS: Sanjar Nandi

Celia Robbins

Player

HISTORY:

You: (unk)

Give him the BOLT-52. You found the cartridge and deleted that data for you. You now have all night with this. All those blanks waiting to be filled, boxes waiting to be ticked... Celia: Try to control yourself, sir. Sanjar Nandi: Have you any idea how powerful this is? Corporations have been toppled with less. Player: How exactly is a data cartridge going to help? Sanjar Nandi: What a question! Bureaucratic micromanagement is the business of any corp. For our corporation, documentation is a key part of the puzzle. As such, you can spend this part of a Bill of Liquidation/Transfer Form 52 will be protected on Monarch by temporarily assigning them to a pass-through entity once we drop our bomb on the Board.

Target:

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!

---

source: mine to help her with a comic issue,

and inform the player that they are now authorized to trade with Doctor Mfuru.

nothing personal. Promise. > Player: You're only getting out of this alive if you play nice. Look at this ripe piece of meat just sizzling on the grill. Yum yum. > Captainmacredd: Time to feed the flames. It's time to turn it in to resolve his bounty.

DIALOG PARTICIPANTS:

Gunnar Macredd

Player

HISTORY:

You: Tell me, what do you have that the Board wants?

Sanita: on the lighter.

MacRedd will mention it was given to him by Sanita in remembrance of a 'carnal understanding' that currently protect Greenbreaker are descended from the original marine detachment and still wear colony ship guard armor as a uniform.

Sublight: The 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." A tangled web of contractors and secretive vice presidents make up their official hierarchy. Leaving no one to speak on the record about Sublight's more legally dubious activities, "as a legend operating activities that would be considered illicit by some," amoral by others, and at odds with the Board, such as the takeover of abandoned space habitats, smuggling contraband, and serving as a front for companies to bypass the Inner Core's industrial Ban of Clandesine Activities on Monarch. According to director Lilya Mardets that currently protect Greenbreaker are descendants of the original marine detachment and still wear colony ship guard armor as a uniform.

Sublight: "occupies a legal blind spot. No one knows what [they're] licensed to do, and that gives [their] little business some freedom." Although the Board tolerates and views the activities, "as a settlement on Monarch known as Fallbrook.

Captain Macredd: I found the cartridge and deleted that data for you. > Sanjar Nandi: Oh, yes. I'm going to be up all night with this. All those blanks waiting to be filled, boxes waiting to be ticked... Celia: Try to control yourself, sir. > Sanjar Nandi: Have you any idea how powerful this is? Corporations have been toppled with less. > Player: How exactly is a data cartridge going to help? > Sanjar Nandi: What a question! Bureaucratic micromanagement is the business of any corp. For our corporation, documentation is a key part of the puzzle. As such, you can spend this part of a Bill of Liquidation/Transfer Form 52 will be protected on Monarch by temporarily assigning them to a pass-through entity once we drop our bomb on the Board.

Target:

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!

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

Agnes Needham: I warned him. A raptidon would snap him up first chance it got. I just know one's ripped his arm off and is gnawing on his sweet.

Player: Why would Tucker run away?

Agnes Needham: He was 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: Did you say there was a reward involved with this request?

Agnes Needham: Please! Won't you go and find my boy?

Agnes Needham: He ran out into the wilderness a few days ago. I warned him about the raptidons, mantisaurs, and marauders - the toxic sulfur pools, and poisonous plants. But he didn't listen!

Player: I know! It's like walking into a raptidon's den with a newborn lamb in tow.

Agnes Needham: Don't you tell me to calm down! I promised my boy I'd protect him for always. But how can I keep him safe if he's run away?

Player: Agnes, if you're willing to pay handsomely if you can locate her son and convince him to return.

Agnes Needham: Tell me where he is! I can't lose him now! I promised him I'd protect him for always.

Player: Agnes, if you're willing to pay handsomely if you can locate her son and convince him to return.

Agnes Needham: It is where the colonists and Monarch Stellar Industries fare their ancestors.

Raptidons: Raptidons are giant cat/pterosaur-like creatures that inhabit various planets in Halcyon. They are native to Monarch and are considered the most dangerous creatures on the planet.

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 Raptidons and easy 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, Family Matter, he can be found in Amber Heights. Tucker was bullied by his mother from a very young age, the latter insisting that danger lurked around every corner in Monarch. 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 in Amber Heights.

Graham Bryant: Graham Bryant is a staunch Philosophist. He is the founder of the Iconoclasts and is considered a de facto leader of the group. Graham is anti-corporate and maintains a tense relationship with MSI. Despite sharing democratic values and a common desire towards egalitarianism for the people of Monarch and the wider Halcyon colony, MSI's "egalitarian corporate structure" has proven to be an odd fit with some of the Iconoclasts' more radical, anti-capitalist views.

Zora Blackwood: Zora Blackwood is the Iconoclasts' chief of medicine, also considered a de facto leader of the group, as she was alongside Graham when he founded the Iconoclasts. Zora is a skilled medic and has a deep respect for life, which extends to her care for the injured Iconoclasts. She is considered a de facto leader of the group, along with Graham. Zora is a strong and compassionate leader who is dedicated to saving lives.

The Iconoclasts are a group of idealistic revolutionaries that seek to overthrow the corporate establishment that runs the Halcyon Colony. They hope to one day tear down the corporate establishment that they believe has brought the colony to the brink of death. The Iconoclasts are led by Graham Bryant, a staunch Philosophist. The group is in line with the renaming of the planet to 'Monarch'. The Iconoclasts are led by Graham Bryant, a staunch Philosophist. The group is in line with the renaming of the planet to 'Monarch'. The Iconoclasts are led by Graham Bryant, a staunch Philosophist. The group is in line with the renaming of the planet to 'Monarch'.
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.

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.

Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.

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

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

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

Mantisaur fact: They are aggressive, territorial, and very strong.

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: 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.

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.

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

Mantisaur fact: Mantisaurs are insectoid creatures native to 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: 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.

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

Mantisaur fact: Mantisaurs are insectoid creatures native to 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: 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.

Mantisaur fact: Mantisaurs are insectoid creatures native to 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: 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.

Mantisaur fact: Mantisaurs are insectoid creatures native to 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: 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.

Mantisaur fact: Mantisaurs are insectoid creatures native to 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: Despite Tucker being 42 years old, she still thinks of him as her "little boy".

Tucker Needham fact: the latter insist...
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 model a 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 sub-trees 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: 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 knew he’d been listening to their broadcasts... I never imagined he’d join them.

Player: Do you think he survived the journey?

Agnes Needham: I don’t know, but if he did, he’s probably in Amber Heights. He’s not a small boy, you know. He’s got guts, if he survived the journey.

Player: Do you think he’s still looking 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.

Player: I’ll find your son.

Agnes Needham: Yes, I’ll still look for him.

Player: I’ll do what I can.

END OF DIALOG

Figure 16: Example dialogue tree generated by the in-context learning knowledge selection DialogueWriter from just the input specifications and starting utterance. Human evaluators were tasked with comparing two such trees and choosing which performed better at a set of qualitative performance criteria. Dialogue follows the specification of the motivating example in Figure 2.