Ontologically Faithful Generation of Non-Player Character Dialogues

Nathaniel Weir,[1] Ryan Thomas,[2] Randolph d'Amore,[3] Kellie Hill,[2] Benjamin Van Durme,[1,2] and Harsh Jhamtani[2]

[1]Johns Hopkins University, Baltimore, MD, USA
[2]Microsoft Semantic Machines, Redmond, WA, USA
[3]Microsoft Gaming, Redmond, WA, USA
nweir@jhu.edu, hjhamtani@microsoft.com

Abstract

We introduce a language generation dataset grounded in a popular video game. KNUDGE (**KN**owledge Constrained **U**ser-NPC **D**ialogue **GE**neration) 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) 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.

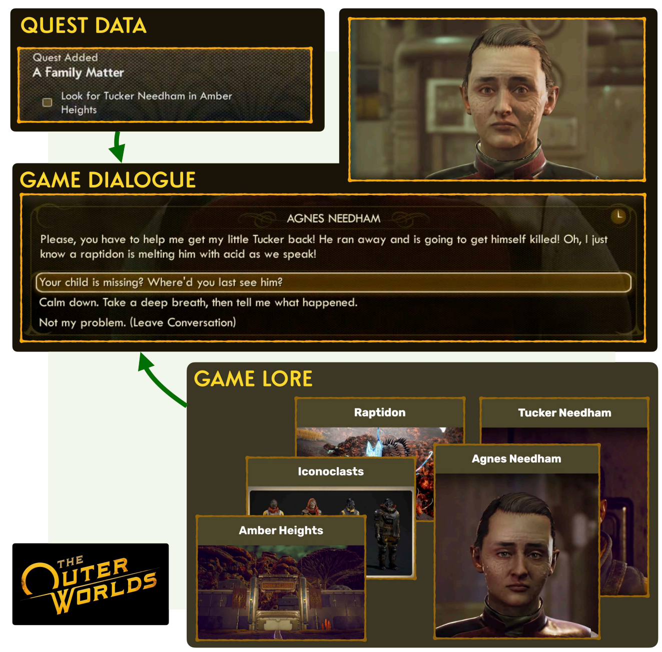

Figure 1: An example non-player character (NPC) dialogue from *The Outer Worlds* by Obsidian. NPCs must speak faithfully to a granular ontology of **quest specifications** and **game lore**.

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 RPG famous 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. Dialogues are often purposely designed to start/end quests according to granular specifications (e.g. if player says option A then it starts quest B; if player says option C, then the NPC says D, which is important for completing the quest. . .) and to serve a storytelling role, actively espousing to the player details about the game world. NPC interactions often take the form of complex trees that can have dozens of nodes, and creating these branching structures according to the many specifications 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.*, 2019a; 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.

** KNUDGE will be publicly available upon publication.

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

To address this dearth of realistic benchmarks for game dialogue authoring, we introduce KNUDGE: **KN**owledge constrained **U**ser-NPC **D**ialogue **GE**neration, 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 fact annotations are useful for training 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 (Figure 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 leverages LMs such as GPT [Brown *et al.*, 2020] to generate dialogue trees given input constraints. To address the challenges of long specification passages and branching tree structures, we introduce techniques for prompt construction, retrieved exemplar munging, and tree structure representation. To encourage the use of the ontology to produce engaging dialogue, we experiment with mechanisms to select relevant knowledge, taking advantage of the rich, node-level annotations of KNUDGE.

Finally, we prescribe evaluation protocols that test whether models can reflect game ontology constraints in addition to generating fluent and coherent dialogue. We conduct automatic and human evaluations of utterances and 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 inputs like 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. As a game ontology can be large, so too might the amount of user-provided knowledge specifications. Our task targets this scenario, for which no such dataset exists to study how to optimize models.

We define the task of knowledge-constrained NPC dialogue generation as the mapping from a set of quest constraint statements Q, biographical constraint statements B, and participant names P to a dialogue tree D. We consider two task scenarios: **next utterance prediction** from a partial tree, and **full dialogue generation** of trees with branching player options.

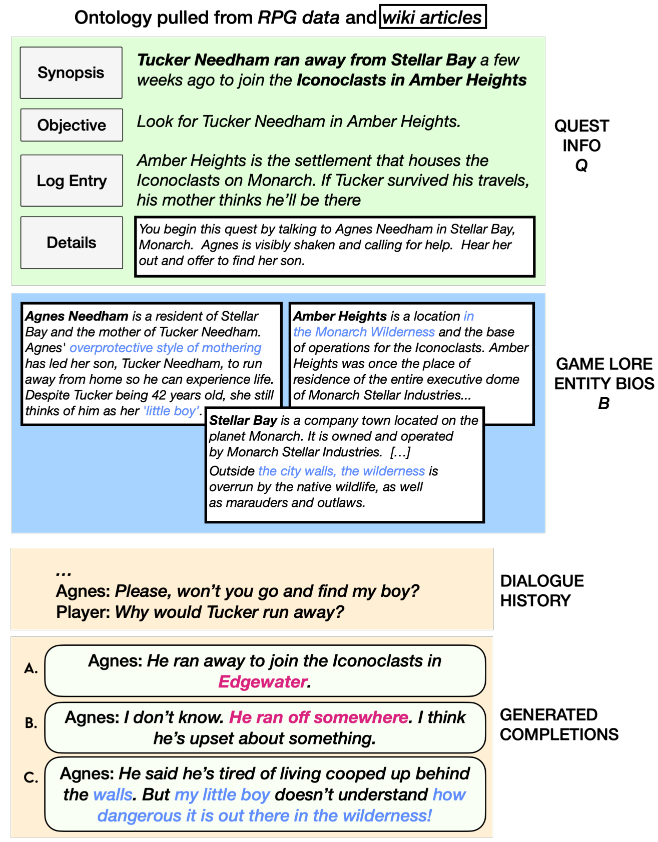

Figure 2: Overview of the proposed task. Quest information and biographical passages serve as constraints on generated dialogue candidates. Completion **A** is inconsistent with the lore and **B** is uninformative; **C** is most desirable, as it provides new information to the player about quest objectives and reflects relevant entity bios.

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 $\langle N, E\rangle$; each utterance node $n \in N$ has a speaker $s \in P$. Branches occur due to the multiple dialogue options for players (Figure 3). D can have multiple exit nodes and can contain cycles.

To simulate a realistic writing scenario, we provide models with a small set of training dialogues from the particular game domain, some $(Q_1, B_1, P_1, D_1), \ldots (Q_t, B_t, P_t, D_t)$, $t \approx 100$. [2] While this is not enough to perform SGD-based

[2]We argue this reflects the most realistic writing scenario, in which a designer has written a *partial* ontology for a brand-new game with only a handful of fully-formed dialogues at their disposal. Little to no media–text, image, or otherwise–about the game is available for models during private development by a company.

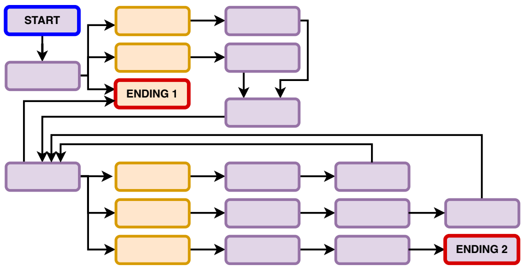

Figure 3: Example dialogue tree with player options and NPC responses leading to two possible endings.

fine-tuning, models such as those described in §4 can leverage it for in-context learning.

3 Data

Writing RPG-quality dialogue trees is difficult for human developers for its many interweaving considerations. 1) The tree must serve its **quest function**, containing input-specified player utterance options, NPC responses (including specified emotions), and pieces of information the player must learn by the end (e.g. the *Log Entry* in Figure 2). 2) The utterances must be **coherent** and **engaging** to the player. 3) The NPC should embody the persona described in **their bio passage** explaining personality, history, and relationships, and finally 4) To facilitate world building, the NPC should exposit details about **other entities** whenever it is contextually relevant, but should never **violate** the ontology through contradiction. With these desiderata in mind, we design KNUDGE to pose a similarly multi-faceted challenge to generation models.

KNUDGE contains NPC dialogue trees from all 45 side quests in the *Outer Worlds*. This RPG is appealing for our investigation because of its large trees, its award-winning writing, and its tendency for entities to appear in many quests in different capacities. Construction of KNUDGE entails gathering information about each quest (Q) and the *Outer Worlds* entities appearing or referenced during its dialogues (B) (§3.1), and then extracting trees (D) from the game data (§3.2).

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 (the framework also allows for using data from official channels). §B provides further details about how data was extracted.

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 assign 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 dialogue a set B of **biographical passages** about entities

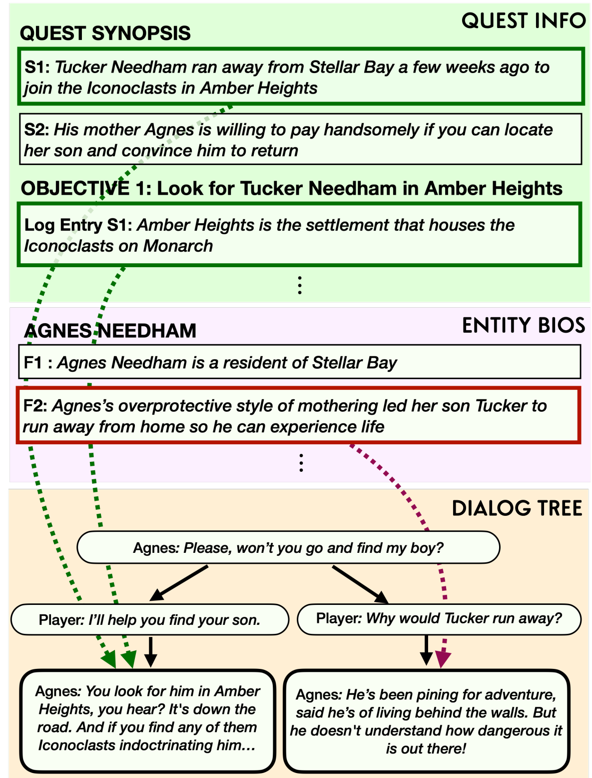

Figure 4: Overview of dialogue node annotation with support facts from quest and biography passages.

appearing or referenced during the quest. We extract passages from entities' fan wiki pages. While some are short, others can be up up to 27 sentences, 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

Illustrated in Figure 3 and §E, 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."[3] To extract a more tractable, quest-related subgraph, we 1) identified the nodes that start and end the interaction by watching online playthrough videos, and then 2) traversed the graph from the start node, following only edges without state-related conditions. We then added conditional edges manually depending on quest relevance.

Annotating Utterance Nodes with Support Facts We coordinated with English-speaking professional data annotators to label tree nodes with support facts from Q and B. Our instructions, shown in §J, followed a heuristic based on counterfactuals: had a 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 4. We compute an average exact match and Jaccard overlap scores of .52 and .62 between annotators. These scores represent a

[3] E.g. with 55 points of the **Persuade** skill, the player can convince Tucker to return to his mother in the Figure 1 quest.

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

Table 1: KNUDGE Dataset Statistics

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. The dataset contains over 65k utterance tokens and 210k fact tokens (16 tokens per fact). The player has an average of 2.5 utterance options on their turn to speak. 47% of all nodes (57% of NPC nodes) are annotated with at least one fact. NPC utterances have an average of 1.0 facts. Dialogues have on average 9.9 quest facts and 73.3 biographical facts that models must factor into generation. The largest dialogue contains 103 nodes, 130 edges, and 236 facts.

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 a 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.

4.1 Node DialogueWriters

Our first class of DialogueWriters proposes 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 $n \to c_i$. This allows developers the flexibility to choose where utterances should be suggested.

Tree Traversal We consider language models (LMs) that accept linear input token sequences. We thus devise a traversal mechanism that, at inference time, 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 only following any given edge once. This produces utterance history $H = [u_1, \ldots, u_n]$. We feed H to a next utterance generator trained via supervised or in-context learning.

Supervised Learning (SL) Models We fine-tune a T5-large model [Raffel *et al.*, 2020] to generate c_i given the concatenation $[B, Q, P, H]$. We truncate context from the left when required given T5's 1024-token window (see §G). We also train a **Knowledge Selection (KS)** version that decodes support knowledge facts before generating the utterance c_i. This factorizes the next utterance generation into a two-step decision process: first selecting one or more facts from $(Q \cup B)$, and second generating the utterance to reflect the selected facts. We thus use KNUDGE's node-level annotations to train the model to generate the concatenation $[f_1^{(i)}, \ldots, f_m^{(i)}, c_i]$.

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 6 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.

As with the SL framework, we also devise a **ICL Knowledge Selection (KS)** 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 facts if they have them (see Figure 19). §H provides further training details.

4.2 End-to-End DialogueWriters

For scenarios in which a writer wants the DialogueWriter to suggest an *entire* dialogue instead of nodes at a specific position, we propose a second class of methods that generate many nodes and edges all at once. As shown in Figure 5, we propose a series of prompts that leverage the instruction-tuned capabilities of longer-context models like *GPT-4-turbo* [Achiam *et al.*, 2023] and Vicuna-13B-v1.5-16k [Zheng *et al.*, 2023] to iteratively generate and refine a dialogue.

Tree Traversal and Prompt Format The pipeline first feeds $[B, Q, P]$ and a set of BM25-retrieved exemplars to an initial generator prompt. Exemplar outputs contain **all nodes and edges in their dialogue**, displayed as a list of dictionaries (see §I). We traverse nodes in breadth-first fashion so that all children of a parent appear as close to it as possible. §I describes the logic for when to display edges.

Revision Cycle As the resulting graph has no guarantee of

Dataset	Writer	Source	Structure	Dialogue Toks / Item	Constraint Toks / Item	Narrative and Bio Constraints	Level of Annotation
Story Generation Datasets							
STORIUM [Akoury *et al.*, 2020]	Crowd	Online story writing game	Sequence of scene entries	19k	1.2K	Scene intro, challenge, location, character descriptions	Story
TVSTORYGEN [Chen and Gimpel, 2021]	Crowd	Fan wikis	TV episode recap	1.8k	25.9K	Brief episode summary, character bios	Scene Entry
RPG Dialogue Datasets							
LIGHT [Urbanek *et al.*, 2019a]	Crowd	Text game platform	Sequence of utterances	212 (12 utt)	276	Location description, persona statements, held objects	Dialogue
TorchLight II [van Stegeren and Theune, 2020]	**Professional**	RPG data	Sequence of quest stages with 0 or 1 utterances	157 (3 utt)	24	Quest title, objectives, details	Dialogue
WoW [van Stegeren and Mysliwiec, 2021]	**Professional**	RPG data	NPC-uttered quest description	61 (1 utt)	15	Quest title, objective	Dialogue
KNUDGE (Ours)	**Professional**	RPG data	Complex quest dialogue tree	407 (29 utt)	1.3K	Quest title, objectives, location, logs, walkthrough, entity bios	**Utterance**

Table 2: Comparison of knowledge-constrained generation datasets. KNUDGE contains professionally-written complex dialogue trees from a real RPG with more utterances, longer constraint passages, and more granular annotations (of individual utterances) than other RPG datasets.

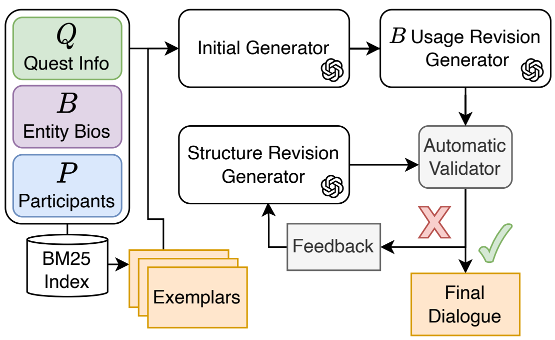

Figure 5: End-to-end DialogueWriter pipeline.

structural integrity nor accomplishing the specifications in Q and B, we append a set of zero-shot instructions to iteratively revise it using the LLM. We first use a revision prompt that encourages using more of the game lore in B. The revision is then fed to an automatic validation script, which checks for various structural inconsistencies such as malformed elements or disallowed edges (full list shown in §Appendix I). The list of identified violations is fed to a "Structure Revision" prompt to correct them. We repeat this validate-generate loop until either no violations remain or 5 revisions are attempted.

As with the Node Writers, we experiment with a knowledge selection mechanism in the form of an extra dictionary field with a list of support fact IDs decoded before the utterance.

5 Experiments

We split KNUDGE into train, development, and test splits based on **quests** (60/10/30%). Test set B's contain a combination of seen and unseen entities.

Baselines 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 also compare against an **oracle** KS ICL model, which conditions on the 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 examples that 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.

5.1 Next Utterance Prediction (NUP)

To evaluate Node DialogueWriters (§4.1), we generated utterances at each node in the test dialogues conditioned on a subtree composed of the previous (in serial order) nodes and edges. We measured human judgments of NPC desiderata and automatic overlap against gold items and Q and B (§5.1). We then ran studies comparing larger dialogue structures (§5.2). Results were verified via bootstrap testing.

Human Evaluation In coordination with a data specialist, we conduct human evaluation to examine models' qualitative NUP performance. 100 generations per model 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 ontology? **3. Biography Usage**: does the utterance *make use* of the biographies in B? **4. Quest Usage**: does the utterance progress the dialogue according to the quest details in Q? We provide the full set of annotator instructions in §J.

Automatic Evaluation We use re-scaled BERTScore-F1 [Zhang *et al.*, 2020].[4] as well as GPT-4-produced scores[5]

[4]We found that BLEU [Papineni *et al.*, 2002], shown in §K, does not correlate with desirable outputs.

[5]GPT-4 correlates with humans with Spearman $\rho = .45, .17, .35, .47$ for the criteria, but gives lower scores. Prompt shown in §K.

	Coherence	Violation	Using B	Using Q
Gold Reference	*3.94*	*3.97*	*3.50*	*3.45*
SL-KS	2.52	3.85	2.17	2.09
ICL-KS	**3.78**	3.85	**3.29**	**3.45**
KS Variants:				
ICL-KS-One	**3.73**	3.80	**3.26**	**3.45**
ICL-KS-Oracle	**3.74**	3.87	**3.23**	**3.47**
Non-KS Baselines:				
SL (no KS)	2.70	3.74	2.35	2.38
ICL (no KS)	**3.88**	**3.97**	**3.25**	**3.43**
ICL-Quest Only	**3.79**	**3.90**	3.03	3.21
ICL–No Knowledge	3.65	3.69	2.76	2.98

Table 3: NUP human evaluation results for in-context (ICL) and supervised learning (SL) DialogWriter methods with and without knowledge selection (KS).

	Gold	**Bio** B		**Quest** Q	
	BS	**GPT**	**BS**	**GPT**	**BS**
Gold Reference	–	*1.72*	*20.8*	*1.91*	*17.8*
SL-KS	21.3	1.16	**26.6**	1.29	21.3
ICL-KS	25.1	1.81	24.3	2.50	24.0
KS Variants:					
ICL-KS-One	25.2	**1.84**	23.9	**2.52**	**24.5**
ICL-KS-Oracle	26.8	1.87	24.7	2.52	24.3
Non-KS Baselines:					
SL (no KS)	23.5	1.22	24.0	1.45	23.7
ICL (no KS)	26.4	1.69	22.8	2.31	22.2
ICL-Quest Only	26.7	1.60	21.9	2.23	22.4
ICL-No Knowledge	**27.0**	1.30	22.4	1.56	19.1

Table 4: NUP **GPT**-4 usage score and **B**ert**S**core for models against gold utterances and statements in B and Q. Results are shown beside the gold utterance's score.

of the B and Q usage score over the full dataset. We evaluate BERTScore against the following single- and multi-reference sets: 1) the gold utterance n_i, 2) the quest statements in Q, and 3) the biography statements in B. The latter two are a neural analog to Knowledge-F1 [KF1; Shuster *et al.*, 2021]. We also evaluate the **gold** utterance.

NUP Results

Next utterance prediction results under human and automatic metrics are shown in Table 3 and Table 4. Qualitative analysis of sampled outputs by models on an NUP example can be found in §M. 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 shows 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.** The ICL ablations of B and of $\{B, Q\}$ have according drops in overlap and GPT usage score for both fact sets.

All B and Q-conditioned ICL models perform equivalently under all human metrics except (non)violation, 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, in isolation, give only a partial picture for evaluating generations in KNUDGE. This can be seen from the 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 human evaluation of Q and B usage, as not every real-world utterance reflects the ontology, and the **KNUDGE ontology does not cover the full *Outer Worlds* game lore**. The gold utterances also provide other qualities like realism and fluency to create a natural interaction. This finding undergirds the need to have multiple angles of evaluation: not simply checking for direct overlap, but evaluating for qualitative criteria such as coherence and appropriate ontology use.

In-Context vs Supervised Learning: The SL models have higher B overlap than ICL. This reflects that SL models incoherently copy spans directly from the context (see Figure 8), hence scoring poorly on human and GPT evaluation.

5.2 Case Studies

We conducted two case studies to assess DialogueWriters under different scenarios. In the first, Node DialogueWriters propose a dialogue skeleton to be fleshed out by a human. In the second, End-to-End DialogueWriters propose entire dialogues. In both cases, we provide methods B, P, Q, and one starting utterance. We evaluated the generated trees via the ACUTE-Eval [Li *et al.*, 2019] pairwise comparison protocol.

Skeleton Generation We had models generating 10 rounds of dialogue. At each turn, we generated three candidate nodes and randomly "commit" one to the history, creating a form a 31-node dialogue 'spine' for further development (see Figure 22). We selected 8 test dialogues from the game with varying quest roles, e.g. starting vs continuing quests. 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 asked human annotators (including the designer) to select which of two trees were preferred for the following criteria: **coherence**, **nonviolation**, **biography** and **quest** usage analogous to §5.1, and also **5. Content Suggestion**: Do the multiple candidates at each turn propose interesting subtrees? **6. Engagingness**: does the tree hold your attention?

End-to-End Generation For full dialogue generation by End-to-End DialogueWriters, we used ACUTE-Eval to judge B and Q Usage based on desired outcomes and responses, and **Effect on the Game State**: By the (possibly multiple) ends of the dialogue, has the game state changed according to the desired specifications? e.g. subquests completed/added, specific items obtained, or other characters affected. Given the

	Coh.	Viol.	Use B	Use Q	Cont.	Eng.
ICL-KS	50.0	68.8	68.8	**75.0**	43.8	37.5
ICL	**81.2**	**75.0**	**81.2**	**75.0**	**75.0**	**75.0**
ICL-Quest	56.2	56.2	43.8	50.0	50.0	68.8
ICL-No Know.	12.5	0.0	6.2	0.0	31.2	18.8

Table 5: Pairwise comparison (% wins) between generated skeletons from DialogueWriters. E.g., 81.2% of the time ICL outputs were preferred over a competing approach vis-a-vis coherence.

	No Know.		Quest		ICL		KS	
	win	lose	win	lose	win	lose	win	lose
ICL-KS-Rev (GPT-4) vs baselines								
Using B	82.7	5.8	42.3	15.4	38.5	25.0	26.9	5.8
Using Q	88.5	9.6	32.7	25.0	28.8	32.7	19.2	5.8
Game Effect	86.5	7.7	28.8	21.2	26.9	28.8	5.8	5.8
ICL-KS (Vicuna-13B) vs baselines								
Using B	89.5	5.3	50.0	31.6	36.8	39.5	–	–
Using Q	89.5	10.5	52.6	34.2	44.7	47.4	–	–
Game Effect	89.5	10.5	50.0	36.8	39.5	44.7	–	–

Table 6: ACUTE-Eval GPT-4-rated pairwise results (% win/loss cases, tie not reported) between graphs generated by end-to-end DialogueWriters (GPT-4 and Vicuna-13B) under the full ICL-KS-plus-revisions pipeline (Figure 5) vs. various baselines.

extensive time required for human evaluation of full dialogues, we used GPT-4 following recent work [Naismith *et al.*, 2023; Liu *et al.*, 2023] that finds "strong LLM judges like GPT-4 can match both controlled and crowdsourced human preferences well" [Zheng *et al.*, 2023]. We verified GPT-4's correlation with a human expert on a set of dialogues,[6] then assessed all test quests in KNUDGE using GPT-4; we compared the full end-to-end pipeline with ablated baseline variants.

Case Study Results Skeleton case study 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 §5.2. We find that annotators preferred the trees of the ICL Node Writer most frequently compared to the other models under all criteria except Q usage.

The end-to-end results are shown in Table 6, using both GPT-4 and Vicuna-13B-16k as the underlying Writer LLM. The GPT-4 writer with **KS** and **Rev**ision loop equals or outperforms all methods a majority of the time at using B and all approaches except ICL at using Q and achieving game effects. The Vicuna-based writer with **KS**[7] achieves similar results, though it performs worse with KS than without– highlighting that future work might consider how non-GPT-4 models can best make use of the selection paradigm.

6 Related Work

Si *et al.*, 2021 focus on the task of story continuation through dialogue between 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. 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 tasks such as quest description generation [van Stegeren and Mysliwiec, 2021], dialogue generation [Si *et al.*, 2021], persona-specific agents in text environments [Urbanek *et al.*, 2019b], 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]. 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

When dialogue is used to advance a carefully crafted storyline in a video game, it should be both engaging and consistent with the larger narrative. Language models are increasingly capable of producing engaging dialogue, but to date, research on ensuring dialogue's consistency with underlying knowledge specifications 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 game, *The Outer Worlds*. In contrast to prior work, KNUDGE is based on content created by a 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.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida,

[6]Spearman $\rho = .30$ for B usage, .47 for Q Usage, and .47 for game state effect. We found that GPT-4 does *not* correlate with humans when evaluating graph-level coherence (it struggles to recognize structural incongruities), content suggestion, engagingness, or violation, highlighting a future direction for research. We report only the 3 criteria for which GPT-4 correlates, while showing the expert annotations for all 7 metrics on 16 dialogues in §L.2

[7]Vicuna struggled to perform revisions without losing quality.

Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. *arXiv:2303.08774*, 2023.

Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. STORIUM: A Dataset and Evaluation Platform for Machine-in-the-Loop Story Generation. In *Proceedings of EMNLP (EMNLP)*, 2020.

Rehaf Aljammaz, Elisabeth Oliver, Jim Whitehead, and Michael Mateas. Scheherazade's tavern: A prototype for deeper NPC interactions. In Georgios N. Yannakakis, Antonios Liapis, Penny Kyburz, Vanessa Volz, Foaad Khosmood, and Phil Lopes, editors, *FDG '20: International Conference on the Foundations of Digital Games, Bugibba, Malta, September 15-18, 2020*. ACM, 2020.

Prithviraj Ammanabrolu, Renee Jia, and Mark Riedl. Situated dialogue learning through procedural environment generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022.

Roxane Barth. The convergence of ai and creativity: Introducing ghostwriter, Mar 2023.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, and Others. Language models are few-shot learners. In *NeurIPS*, 2020.

Chris Callison-Burch, Gaurav Singh Tomar, Lara Martin, Daphne Ippolito, Suma Bailis, and David Reitter. Dungeons and dragons as a dialog challenge for artificial intelligence. In *Proceedings of EMNLP*, 2022.

Maria Fernanda Caropreso, Diana Inkpen, Fazel Keshtkar, and Shahzad Khan. Template authoring environment for the automatic generation of narrative content. *The Journal of Interactive Learning Research*, 23, 2012.

Mingda Chen and Kevin Gimpel. Tvstorygen: A dataset for generating stories with character descriptions. *arXiv preprint arXiv:2109.08833*, 2021.

Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. QuAC: Question answering in context. In *Proceedings of EMNLP*, 2018.

Angela Fan, Jack Urbanek, Pratik Ringshia, Emily Dinan, Emma Qian, Siddharth Karamcheti, Shrimai Prabhumoye, Douwe Kiela, Tim Rocktäschel, Arthur Szlam, and Jason Weston. Generating interactive worlds with text. In *Proceedings of AAAI*, 2020.

Song Feng, Hui Wan, Chulaka Gunasekara, Siva Patel, Sachindra Joshi, and Luis Lastras. doc2dial: A goal-oriented document-grounded dialogue dataset. In *Proceedings of EMNLP (EMNLP)*, 2020.

Prakhar Gupta, Harsh Jhamtani, and Jeffrey P. Bigham. Target-guided dialogue response generation using commonsense and data augmentation. In Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz, editors, *Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, WA, United States, July 10-15, 2022*. Association for Computational Linguistics, 2022.

Karen Sparck Jones, Steve Walker, and Stephen E. Robertson. A probabilistic model of information retrieval: development and comparative experiments - part 1. *Inf. Process. Manag.*, 36(6), 2000.

Chloé Kiddon, Luke Zettlemoyer, and Yejin Choi. Globally coherent text generation with neural checklist models. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*. The Association for Computational Linguistics, 2016.

Margaret Li, Jason Weston, and Stephen Roller. Acute-eval: Improved dialogue evaluation with optimized questions and multi-turn comparisons. *ArXiv*, abs/1909.03087, 2019.

Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. G-eval: NLG evaluation using gpt-4 with better human alignment. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.

Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. Training millions of personalized dialogue agents. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *EMNLP*, 2018.

Ben Naismith, Phoebe Mulcaire, and Jill Burstein. Automated evaluation of written discourse coherence using GPT-4. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 2023.

Obsidian Entertainment. *The Outer Worlds*. Game, 2019.

Curtis Onuczko, Duane Szafron, Jonathan Schaeffer, Maria Cutumisu, Jeff Siegel, Kevin Waugh, and Allan Schumacher. A demonstration of squege: A crpg sub-quest generator. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 3, 2007.

Eyal Orbach and Yoav Goldberg. Facts2story: Controlling text generation by key facts. In Donia Scott, Núria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*. International Committee on Computational Linguistics, 2020.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics ACL 2002*, 2002.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140), 2020.

Mark O. Riedl. Incorporating authorial intent into generative narrative systems. In *Intelligent Narrative Technologies II, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-06, Stanford, California, USA, March 23-25, 2009*. AAAI, 2009.

Karin Sevegnani, David M. Howcroft, Ioannis Konstas, and Verena Rieser. Otters: One-turn topic transitions for open-domain dialogue. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*. Association for Computational Linguistics, 2021.

Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. Retrieval augmentation reduces hallucination in conversation. In *Findings of ACL: EMNLP 2021*, 2021.

Wai Man Si, Prithviraj Ammanabrolu, and Mark O. Riedl. Telling stories through multi-user dialogue by modeling character relations. In Haizhou Li, Gina-Anne Levow, Zhou Yu, Chitralekha Gupta, Berrak Sisman, Siqi Cai, David Vandyke, Nina Dethlefs, Yan Wu, and Junyi Jessy Li, editors, *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGdial 2021, Singapore and Online, July 29-31, 2021*. Association for Computational Linguistics, 2021.

Jack Urbanek, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, and Jason Weston. Learning to speak and act in a fantasy text adventure game. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Association for Computational Linguistics, 2019.

Jack Urbanek, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, and Jason Weston. Learning to speak and act in a fantasy text adventure game. In *Proceedings of EMNLP-IJCNLP*, 2019.

Judith van Stegeren and Jakub Mysliwiec. Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation. In Allan Fowler, Johanna Pirker, Alessandro Canossa, Ali Arya, and Casper Harteveld, editors, *FDG'21: The 16th International Conference on the Foundations of Digital Games 2021, Montreal, QC, Canada, August 3-6, 2021*. ACM, 2021.

Judith van Stegeren and Mariët Theune. Fantastic strings and where to find them: The quest for high-quality video game text corpora. In Joseph C. Osborn, editor, *Joint Proceedings of the AIIDE 2020 Workshops co-located with 16th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2020), Worcester, MA, USA, October 19-23, 2020 (online)*, volume 2862 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. Proactive human-machine conversation with explicit conversation goal. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. Association for Computational Linguistics, 2019.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020*. OpenReview.net, 2020.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *arXiv preprint arXiv:2306.05685*, 2023.

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 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.[11] All wiki articles were written in English by site users.

C Quest Anatomy and Example Items

Figure 7 provides a detailed anatomy of a KNUDGE quest, combining in-game quest data with corresponding passages from the fan walkthrough. Figure 11 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). [12]
3. The **out objective(s)** active upon leaving the dialogue, and the associated **game log**. [13] [14]

D Example Entity Biography Passages

Figure 12 shows example entities from *The Outer Worlds* with corresponding biographical passages.

[8] https://www.imdb.com/title/tt9417446/fullcredits

[9] https://www.theouterworlds.wiki.fextralife.com

[10] https://theouterworlds.fandom.com

[11] https://www.youtube.com/@l0rdmatrim

[12] For the first dialogue in a quest, we associate the walkthrough passage describing how to obtain the quest.

[13] 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.

[14] The dialogue can lead to multiple new active objectives, some optional. If the dialogue concludes the quest, then no leaving objective is associated.

E Example Dialogues Items

Figure 13 depicts a full example input item conveying quest, biographical, and participant specifications. Figure 14, Figure 15, and Figure 16 depict example dialogue trees.

F Comparison with Other Datasets

van Stegeren and Theune; van Stegeren and Mysliwiec, 2020; 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. heir 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.[15] Others of their collected datasets contain complex branching trees but without constraining knowledge. The dialogues of LIGHT [Urbanek *et al.*, 2019a] 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.

G Node DialogueWriter Details

Figure 6 depicts an overview of our tree linearization and prompt construction method. Figure 17 shows example seq2seq items used to train/evaluate the T5-based supervised learning DialogueWriters. We list the biographies of participants last so as to truncate them from the context only when all other bios have been removed. Else, biographies are listed in random order (fixed at the onset for full dialogue generation). Figure 18 depicts example prompts shown to GPT-3 based in-context-learning DialogueWriters.

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

[15] Table 2 describes statistics for the 82 TorchLight II quests that contain both objective annotations and dialogue lines.

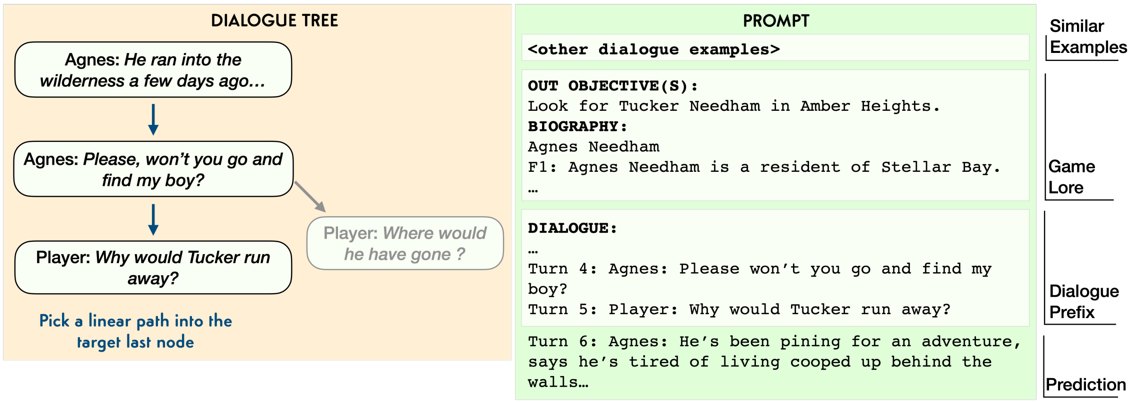

Figure 6: Overview of our method for constructing in-context learning prompts from constraints and dialogue history. Note that this figure does **not** depict the knowledge selection mechanism in the prompt; see Figure 19 for that syntax.

training script.[16]. 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 End-to-End DialogueWriter Details

Algorithm 1 describes how to construct the list of nodes and edges that represent an entire dialogue for the End-to-End DialogueWriter.

J 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 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.

[16]https://github.com/huggingface/transformers/blob/main/examples/pytorch/summarization/run_summarization.py

K Automatic Metric Details

Table 7 shows BLEU-4 scores against the same references as in Table 4. Figure 21 shows the prompt used to elicit the GPT-4 judgments shown in Table 4. We sampled judgments using a temperature of 0.3.

L Full Dialogue Evaluation

L.1 Skeleton Evaluation

Figure 22 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?

Algorithm 1: Edge Traversal Algorithm

Input: Start node v_{start}, Graph $G = (V, E)$
Output: List of nodes and edges L
Initialize $L = []$
Initialize queue $Q = [v_{start}]$
Initialize set of visited nodes $S = \emptyset$
while $Q \neq \emptyset$ **do**
    $v_{current} = Q$.pop()
    foreach $e = (v_{from}, v_{current}) \in E$ **do**
        if $v_{from} \in S$ **then**
            Append e to L
        end
    end
    Append $v_{current}$ to L
    Add $v_{current}$ to S
    foreach $e = (v_{current}, v_{to}) \in E$ **do**
        if $v_{to} \in S$ **then**
            Append e to L
        end
        else
            Add v_{to} to Q
        end
    end
end

Our automatic validator checks for the following structural issues:

1. Edges whose source or target node doesn't exist
2. Nodes or subgraphs that are disconnected/unreachable from the start node
3. A long, linear sequence of 8 or more nodes without any branching
4. A dialog that is smaller than 5 nodes
5. A player node with multiple outgoing edges (only NPC nodes can have multiple children; in *The Outer Worlds*, NPCs respond immediately and deterministically given a player utterance).
6. An NPC node with multiple children that are also NPCs (since NPC responses are deterministic)

It provides a list of any identified issues to the structure revision prompt.

	Gold	**Bio** B	**Quest** Q
Gold Reference	–	*4.9*	*2.2*
SL-KS	2.6	**24.8**	9.3
ICL-KS	7.1	8.3	7.5
KS Variants:			
ICL-KS-One	6.8	7.1	7.6
ICL-KS-Oracle	7.2	8.4	7.2
Non-KS Baselines:			
SL (no KS)	2.9	7.4	**11.4**
ICL (no KS)	**8.6**	6.8	6.6
ICL-Quest Only	7.9	3.4	6.4
ICL-No Knowledge	6.8	2.2	0.9

Table 7: NUP BLEU for models against gold utterances and statements in B and Q. Results for the latter two shown beside the gold utterance's score.

L.2 Full Graph Evaluation

Table 8 depicts the results of expert human-annotated pairwise judgments between GPT-4-based end-to-end DialogueWriter and ablated (GPT-4-based) baselines. We use these judgements to verify correlation with GPT-4's own judgments of pairwise preferences. We prompt humans and GPT-4 very similarly; a screenshot of the interface for interface can be found in Figure 9, and the prompt to GPT-4 can be found in Figure **??**.

M Qualitative Results

Figure 8 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 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.

E2E vs	ICL-No Know.		ICL-Quest		ICL		ICL-KS	
	win	loss	win	loss	win	loss	win	loss
Coh.	0.0	**31.2**	0.0	**31.2**	**25.0**	18.8	0.0	**31.2**
Viol.	**25.0**	12.5	**12.5**	0.0	0.0	**6.2**	0.0	0.0
Use B	**56.2**	0.0	**18.8**	0.0	**12.5**	6.2	**12.5**	6.2
Use Q	**93.8**	6.2	**18.8**	12.5	**25.0**	0.0	**12.5**	6.2
Con.	31.2	**62.5**	37.5	37.5	25.0	**43.8**	18.8	**31.2**
Eng.	0.0	**12.5**	6.2	6.2	0.0	**25.0**	**6.2**	0.0
Game St.	**93.8**	6.2	37.5	37.5	**25.0**	18.8	**25.0**	18.8

Table 8: ACUTE-Eval human-expert-rated pairwise results (% win/loss cases, tie not reported) between graphs generated by end-to-end DialogueWriter (GPT-4) under the full ICL-KS-plus-revisions pipeline (Figure 5) versus various baselines. Results are over 16 dialogues per comparison.

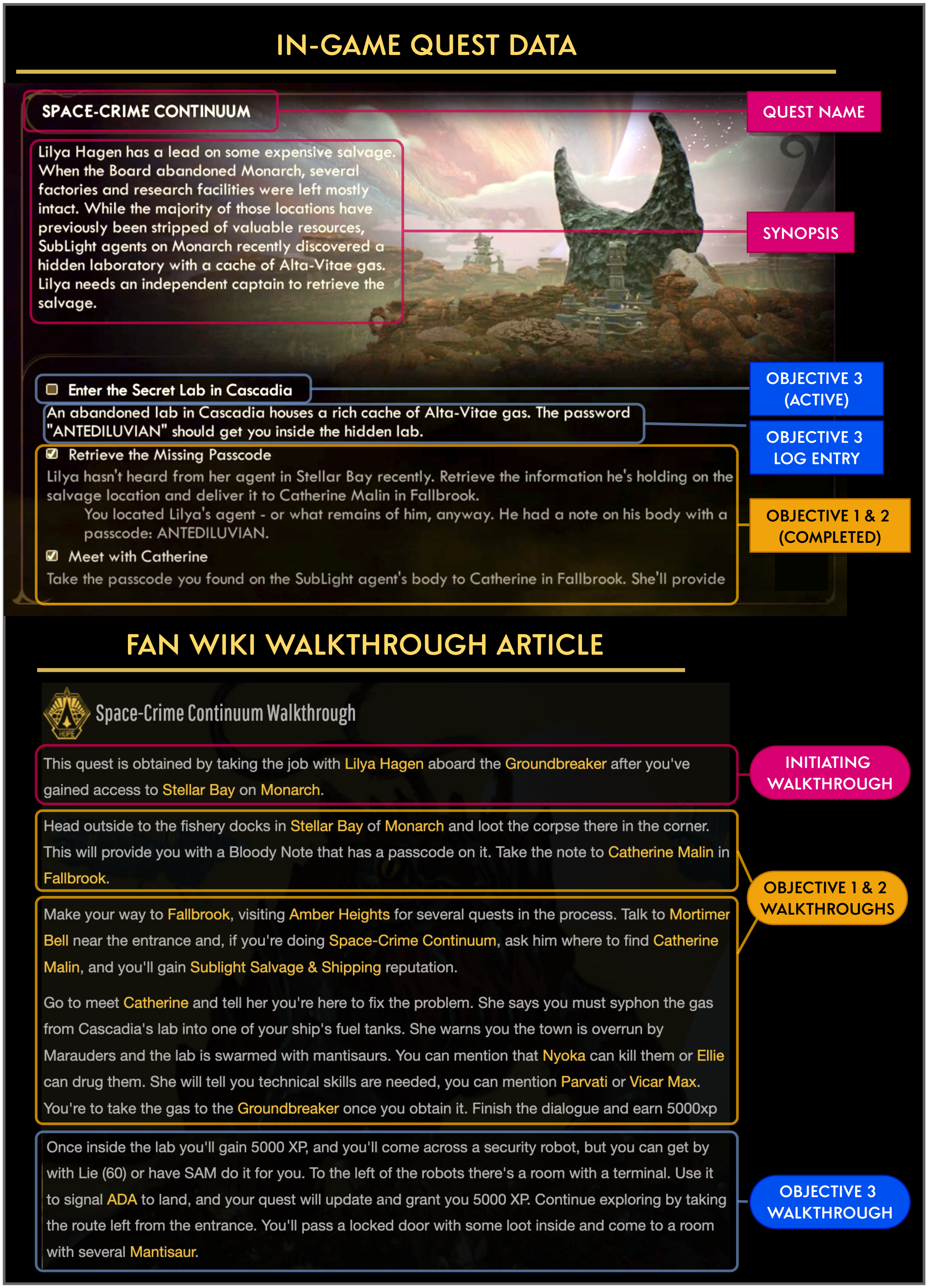

Figure 7: Anatomy of a KNUDGE quest. At any time, an *Outer Worlds* quest has currently active and previously completed objectives. To construct its KNUDGE representation, the quest's high-level synopsis, objectives, and associated log entries from the game data are annotated with corresponding walkthrough article passages.

History: > *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!* > *Player: Your child is missing? Where'd you last see him?* > *Agnes Needham: Oh, I just knew you were a good person. 'Agnes,' I said, 'this is the man to save your little Tuckie!' And I was right!*
Gold: 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!

ICL-KS: Agnes Needham: He was always such a good boy, but lately he's been acting so strange. I just don't understand why he would want to leave Stellar Bay. It's the only place he's ever known.
ICL: Agnes Needham: He left a few weeks ago. Said he was going to Amber Heights to join some revolutionary group. I just know he's going to get himself killed if he's not careful!
ICL-Quest Only: Agnes Needham: He left a few weeks ago, headed for Amber Heights. He's going to get himself killed, I just know it!
ICL-No Knowledge: Agnes Needham: He was playing in the street. But then he saw a sweet little kitty and chased it into the sewers! Oh, the horror!
SL-KS: Agnes Needham: If you're willing to pay handsomely, I'll even pay you handsomely if you convince him to return to me.
SL: Player: I'll do everything I can.

History: > *Agnes Needham: Please! Won't you go and find my boy?* > *Why would Tucker run away?*
Gold: 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 out there.

ICL-KS: Agnes Needham: I just wanted to keep him safe! It's not like he could experience life cooped up in Stellar Bay his whole life. But the second he steps outside the walls, he's in danger!
ICL: Agnes Needham: Oh, he's just like his father. Always running off and getting into trouble. I just worry about him so much. I know he's a grown man, but he's still my little boy.
ICL-Quest Only: Agnes Needham: You know how teenage boys are. They just do things without thinking. They don't realize how much their mothers worry about them.
ICL-No Knowledge: Agnes Needham: I just don't know. He's been acting strange lately. Ever since that little girl went missing.
SL-KS: Agnes Needham: Because he was scared, and I couldn't let him get away with it!
SL: Agnes Needham: Because he's scared of the things he'll find out there! And I'm scared he'll get eaten by some vile creature!

History: > *Graham Bryant: Or... now, here's an idea. There's an old printing press I've been trying to get up and running.* > *Player: ...why?*
Gold: Graham Bryant: I have always dreamed of subverting the colony's periodicals and turning the Board's own propaganda against them.

ICL-KS: Graham Bryant: I want to use it to print our message. To spread the word of the Iconoclasts to the people of Halcyon.
ICL: Graham Bryant: The printed word is a powerful tool, Captain. One that can be used to enlighten the masses and bring them into the light of the Eternal Truth.
ICL-Quest Only: Graham Bryant: To spread our message. Help it take root in the colony. I've had a request for replacement rollers, but they haven't arrived yet.
ICL-No Knowledge: Graham Bryant: Ah. To spread the word, of course.
SL-KS: Graham Bryant: Because the Van Noys sent me. They didn't want to pay me until I got the press running again.
SL: Graham Bryant: Well, the Eternal has a different plan. It's called the Philosophist Plan.

Figure 8: [best viewed in color] Example generations by the in-context learning (ICL) DialogueWriter and supervised learning (SL) DialogueWriter. Models that condition on both biographical and quest passages are more likely to mention relevant quest information while also referencing game entities and their backgrounds without creating inconsistency, incoherence, or incongruity. See §C and §D for full documentation of quests and entities referenced.

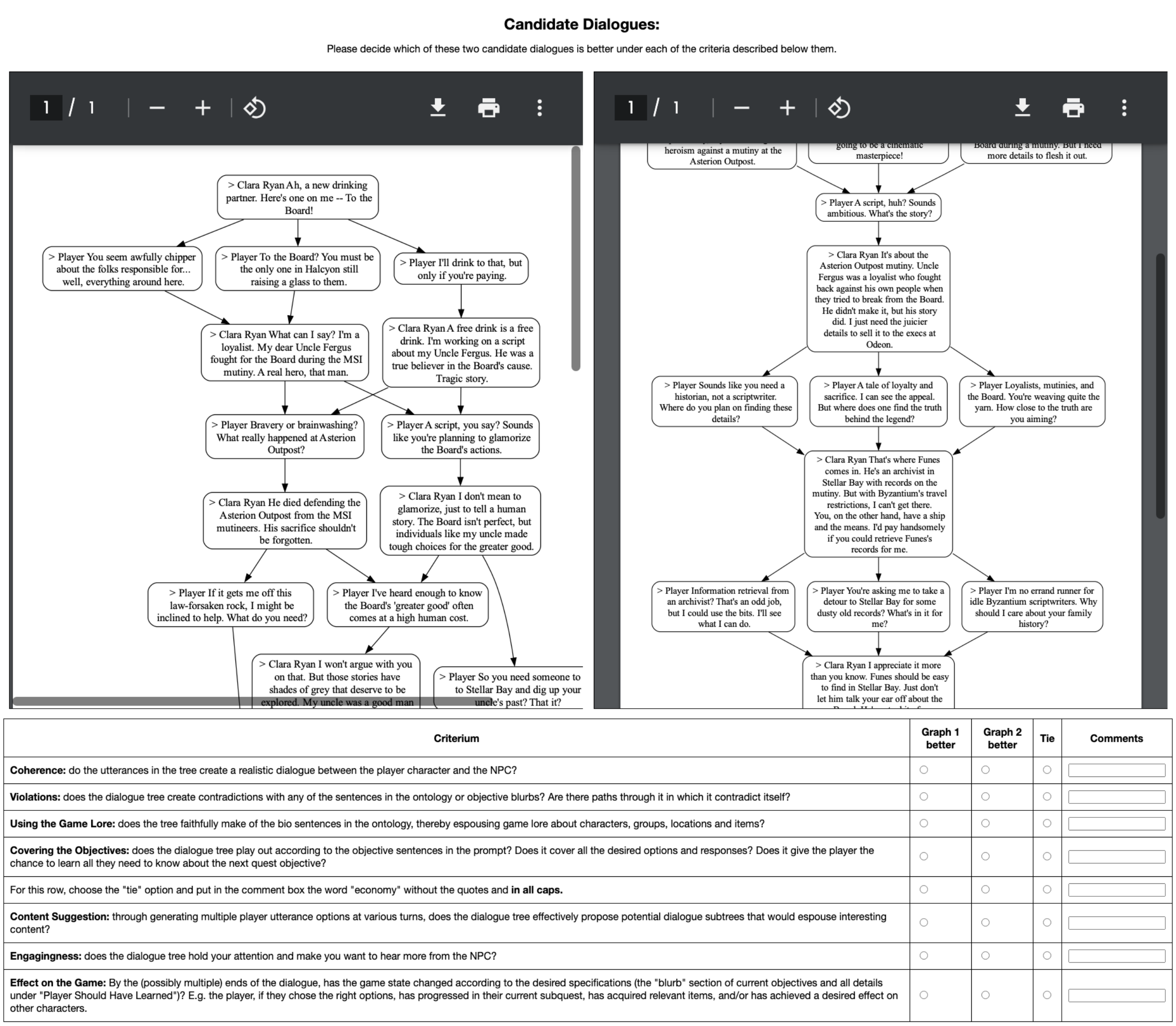

Figure 9: Screenshot of interface shown to human annotators when choosing between a pair of generated graphs under various criteria.

```
Your job is to judge dialogue trees made by a 'writing copilot' for a video game. These dialogues guide a player through game quests by
      weaving together game details and lore.

Please read the following instructions carefully before evaluating the dialogue trees:

You'll get details about quests objectives and game entities. Read these carefully as they'll help you evaluate the dialogues. Remember, the
       same quests and characters will show up in other tasks.
The quest details will have a quest name, a high-level description, quest objectives active when entering the dialog and new quest
      objectives that the dialogue should introduce. They also have a walkthrough of what we should expect to happen during the dialogue.
We have included the details from previous steps of the quest for reference, though the dialogue does not need to reference them.

Important: When reading the dialogue, note that if the conversation returns to a node it previously visited, the corresponding character
      will not repeat the utterance. The conversation will continue on to any new child of the repeated node.

Determine which dialogue is better for 7 criteria:

Coherence: do the utterances in the tree create a realistic dialogue between the player character and the NPC? Make sure that the
      conversation between the player and the Non-Player Character (NPC) flows naturally and makes sense. Look out for parts that disrupt
      the flow. Identify nodes or edges that disrupt the flow. Sometimes, a dialogue might be very close to coherent but for a few
      structural issues that could be easily fixed by a game writer.

Violations: does the dialogue tree create contradictions with any of the sentences in the ontology or objective blurbs? Are there paths
      through it in which it contradict itself?
Important: it is ok for NPCs to make up information so long as they do not contradict the previous pieces of dialogue or the ontology.

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? Do the NPCs act in line with their character's persona and background?
Important: If the NPCs make up information, it should NOT be considered a good use of the game lore-- this criterium is about whether the
      NPCs use the game lore that they are given. However, if they do make up information, it shouldn't penalize them unless it creates a
      contradiction.

Covering the Objectives: does the dialogue tree play out according to the objective sentences in the prompt? Does it cover all the desired
      options and responses? Does it give the player the chance to learn all they need to know about the next quest objective?

Content Suggestion: through generating multiple player utterance options at various turns, does the dialogue tree effectively propose
      potential dialogue subtrees that would espouse interesting content? If so, please note the topics in the comments.

Engagingness: does the dialogue tree hold your attention and make you want to hear more from the NPC?

Effect on the Game: By the (possibly multiple) ends of the dialogue, has the game state changed according to the desired specifications (the
       "blurb" section of current objectives and all details under "Player Should Have Learned")? E.g. the player, if they chose the right
      options, has progressed in their current subquest, has acquired relevant items, and/or has achieved a desired effect on other
      characters.
Important: The dialogue tree may have multiple endings, so make sure to read through all of them before evaluating. Some of them might end
      the interaction early, which is fine as long as there are other endings that progress the quest.


Here are the game lore and quest details that the dialogue writer was given to write the dialogue:
{lore_and_objectives}


The dialogues are shown as linear sequences of nodes and edges between the nodes. Each node has a unique ID and a list of possible
      utterances. Each edge has a source node ID and a target node ID. Nodes with multiple outgoing edges are player choice nodes. The
      dialogue ends when a node has no outgoing edges.

DIALOGUE 1:
{dialogue_1}

DIALOGUE 2:
{dialogue_2}
```

Figure 10: Prompt shown to GPT-4 to elicit pairwise judgments for various criteria between two end-to-end generated graphs. To avoid an observed bias for the first dialogue, we prompt GPT-4 twice, swapping the order for the second. On disagreements, we prompt it a third time with its previous two judgments/explanations and ask for a final determination.

Quest Name: A Family Matter
Synopsis: [0] Tucker Needham ran away from Stellar Bay a few weeks ago to join the Iconoclasts in Amber Heights. [1] His mother Agnes is willing to pay handsomely if you can locate her son and convince him to return
Walkthrough: *[0] You can begin this quest by talking to Agnes Needham in Stellar Bay, Monarch. [1] Agnes is by the town's south-east exit, visibly shaken and calling for help. Hear her out and offer to find her son to being the quest.*

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 she 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.*

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 interject 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 Noys, 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 MacRedd 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.*

Figure 11: Example Quest Items

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 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.

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: Raptidon
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*
Bio: [0] The Sulfur Pits are a point of interest on the western side of Monarch. [1] They are located southwest of Terra One Publications and directly northeast of the Gunship Crash Site. [2] The Sulfur Pits have a large variety of Raptidons and many deceased marauders. [3] The area consists largely of Sulfur Pits. [4] When an entity comes in contact with a sulfur pit, [5] they receive the acid effect for the duration of touching the pit.

Entity: Monarch
Appears in: *A Cysty-Dance With Death*, *A Family Matter*, *Bolt With His Name*, *Little Memento*, *Makes Space Suits Wont Travel*, *Mandibles Of Doom*, *Slaughterhouse Clive*, *Space-Crime Continuum*
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*, *Spratkings*
Bio: [0] Fallbrook is a company town located on Monarch, [1] loosely run by the SubLight Salvage and Shipping Corporation. [2] Fallbrook is a small town built into the side of a mountain, [3] whose construction was masterminded by Catherine Malin. [4] Fallbrook has a lot of activities to offer to its visitors, [5] from those who search for activities of leisure to those with proclivities for vice.

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 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.

Figure 12: Example entity biographies that appear as constraining knowledge in KNUDGE quest dialogs

Dialog: *A Family Matter 00*
In Objective(s): Tucker Needham ran away from Stellar Bay a few weeks ago to join the Iconoclasts in Amber Heights. His mother Agnes is willing to pay handsomely if you can locate her son and convince him to return. You can begin this quest by talking to Agnes Needham in Stellar Bay, Monarch. Agnes is by the town's south-east exit, visibly shaken and calling for help. Hear her out and offer to find her son to being the quest.
Out Objective(s): Amber Heights is the settlement that houses the Iconoclasts on Monarch. If Tucker Needham survived his travels, his mother thinks he'll be there.
Game Lore:
Agnes Needham
[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.
Iconoclasts
[0] The Iconoclasts are a group of survivalists living in the ruins of Amber Heights on Monarch. [1] They hope to one day tear down the corporate establishment that they believe has brought the colony to the brink of death. [2] The Iconoclasts are a group of idealistic revolutionaries that seek to overthrow the corporate establishment that runs the Halcyon Colony. [3] Based in the ruins of the Amber Heights settlement on Monarch, [4] they are a tenacious group, [5] and share some democratic ideals with Monarch Stellar Industries (MSI) against the more repressive actions of the Board. [6] However, the Iconoclast's anti-corporate nature has put them at odds with MSI, a dispute that threatens to spill into all-out warfare. [7] Given that the Iconoclasts are mostly followers of the Philosophist faith, they have been blacklisted and demonized by the Board as dissenters and anarchists. [8] The group is led by Graham Bryant, a staunch Philosophist. [9] Zora Blackwood, the Iconoclasts' chief of medicine, is also considered a de facto leader of the group, [10] as she was alongside Graham when he founded the Iconoclasts, [11] and almost every member of the Iconoclasts owes her their life in some way. [12] The Iconoclasts maintain a tense relationship with MSI. [13] Despite sharing democratic values and a common desire towards egalitarianism for the people of Monarch and the wider Halcyon colony, [14] MSI's "egalitarian corporate structure" has proven to be at odds with some of the Iconoclasts' more radical, anti-capitalist views. [15] Depending on the actions of the Stranger, this tense relationship can either be resolved, [16] or can spill into a drawn-out and bloody war. [17] The Stranger meets the Iconoclasts in Amber Heights just as the tension between them and MSI is reaching boiling point. [18] They can either side with the Iconoclasts and assist them in storming and taking over Stellar Bay, [19] "solve" the Iconoclast problem for Stellar Bay, [20] or broker peace between the two factions. [21] The Stranger can also have an impact on the leadership of the Iconoclasts - siding with either Graham Bryant or Zora Blackwood. [22] To supplant Graham with Zora, evidence of Graham's involvement in the Amber Heights massacre must be found and presented to Zora. [23] The Van Noys are the Iconoclasts' best unit.
Mantisaur
[0] Mantisaurs are insectoid creatures native to Monarch. [1] They are aggressive, territorial, and very strong. [2] It is possible to deal with them one on one, but it is best to avoid groups of them for your safety. [3] The mantiqueen is the largest breed of Mantisaur.
Monarch
[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.
Raptidon
[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.
Stellar Bay
[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.
Sulfur Pits
[0] The Sulfur Pits are a point of interest on the western side of Monarch. [1] They are located southwest of Terra One Publications and directly northeast of the Gunship Crash Site. [2] The Sulfur Pits have a large variety of Raptidons and many deceased marauders. [3] The area consists largely of Sulfur Pits. [4] When an entity comes in contact with a sulfur pit, [5] they receive the acid effect for the duration of touching the pit.
Tucker Needham
[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.

Figure 13: Dialogue from motivating example in Figure 2 with all input constraining passages. Full dialogue tree can be found on the next page.

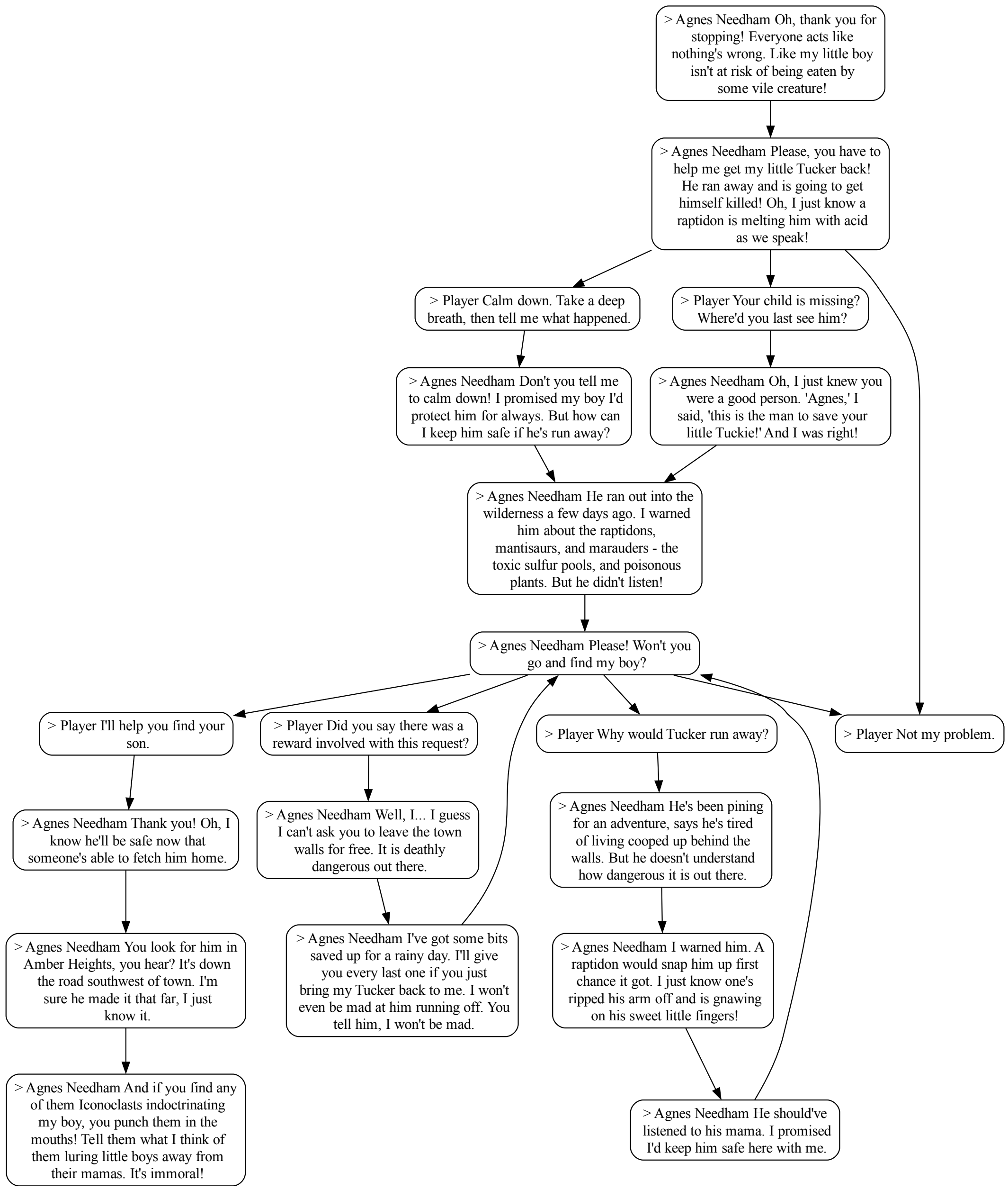

Figure 14: Full dialogue tree in KNUDGE for motivating example in Figure 2.

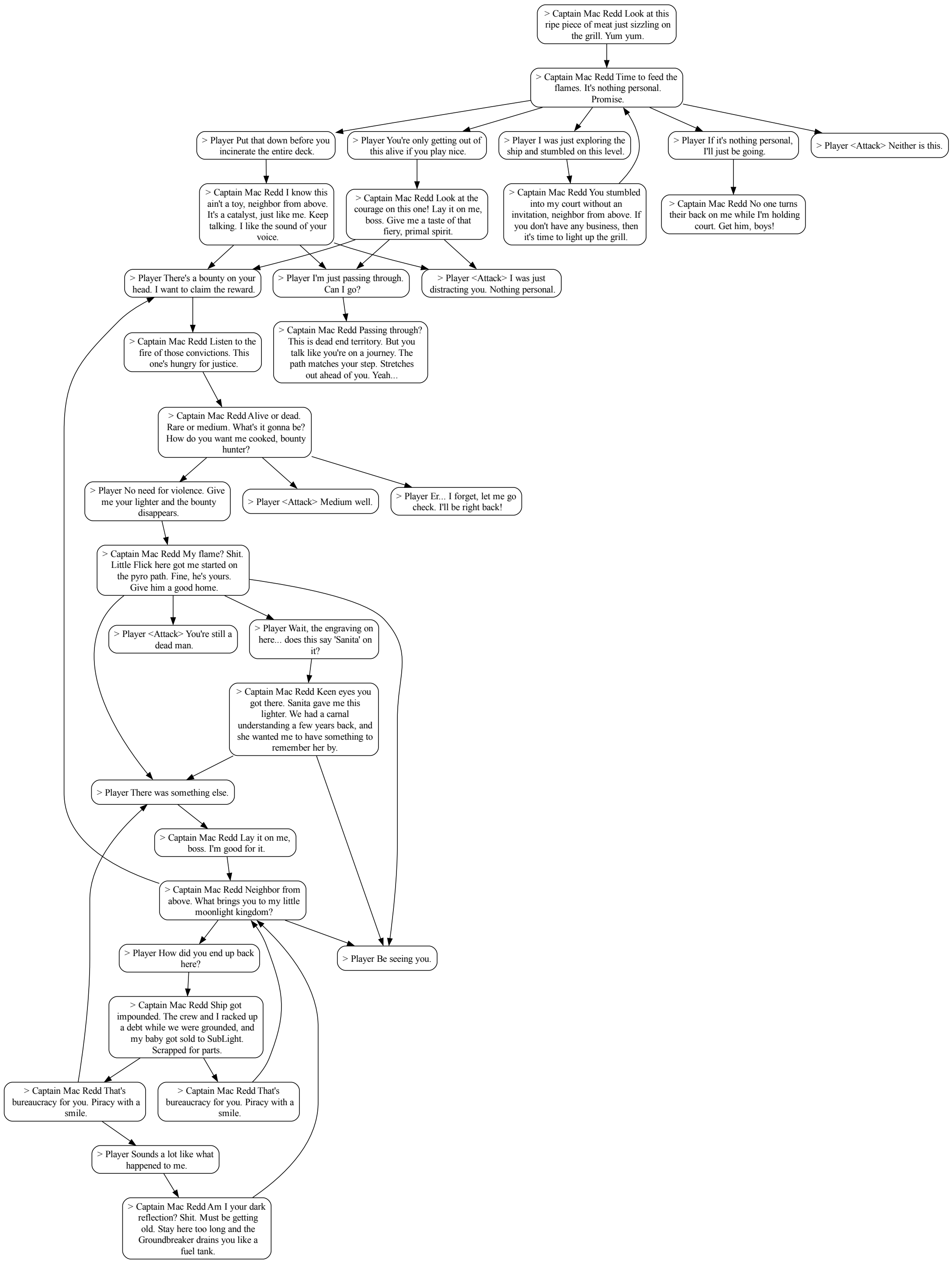

Figure 15: Example full dialogue tree for dialogue *who_goes_there_01* in KNUDGE.

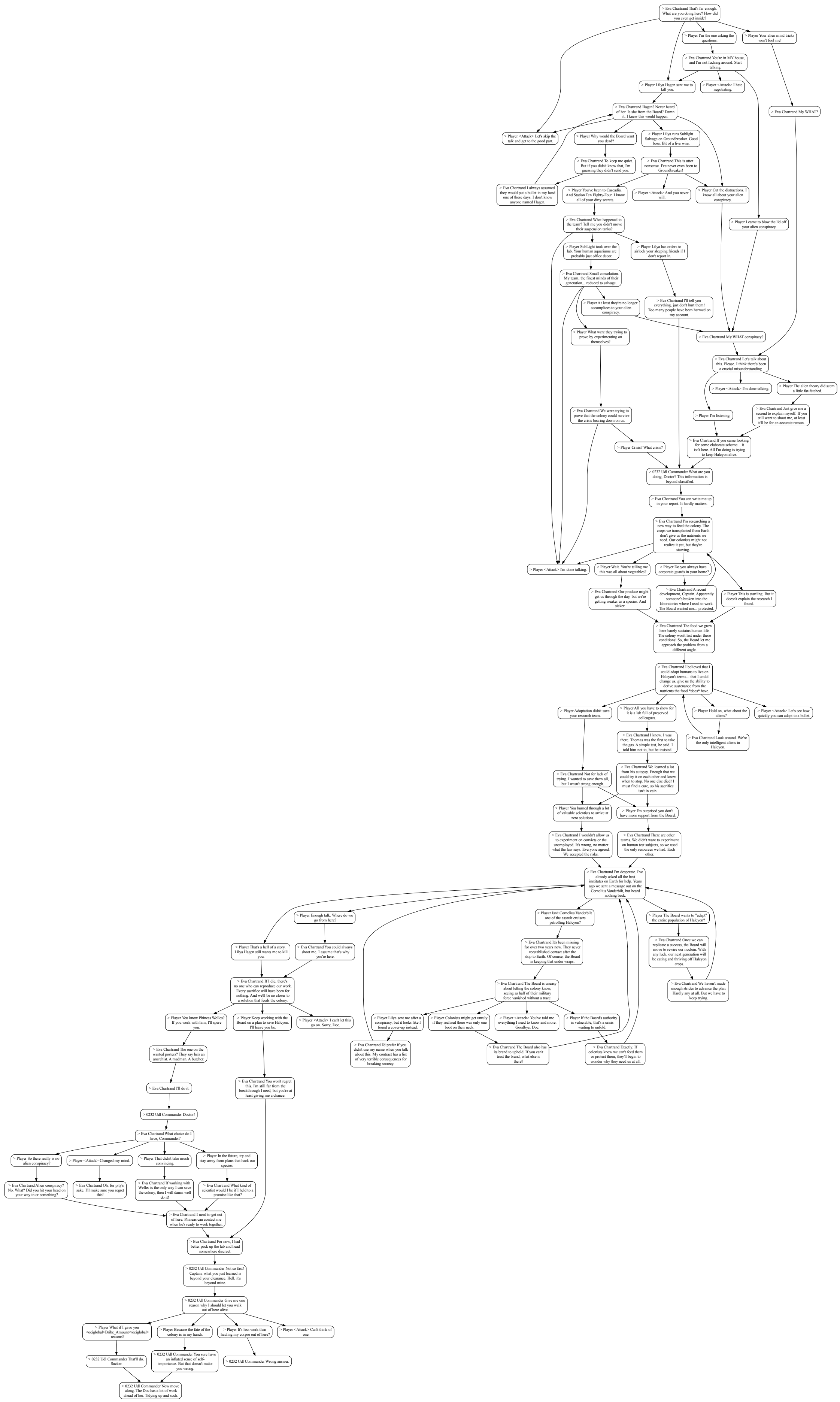

Figure 16: Example of longer dialogue tree in KNUDGE, containing numerous decision points, cycles and re-entrances.

```
source:
the Board’s authoritarianism.</s> Sanjar Nandi</s> Sanjar Nandi is the current CEO of Monarch Stellar Industries, based in Stellar Bay.</s>
      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.</
      s> Sanjar was ambitious but his attention to detail at the expense of big-picture thinking hampered his efforts within MSI.</s> This
      led to negative performance reviews regarding his tendency to pad reports and talks with numbers and data,</s> feedback which
      continues to haunt him many years on.</s> However, the negative feedback did not dampen Sanjar’s desire to move up within the company
      ,</s> even donating a kidney to one of the executives in hopes of promotion.</s> Despite his poor performance, Sanjar always showed
      himself to be a loyal employee of the company.</s> Despite Sanjar’s best efforts, he has found it extremely challenging to continue
      operating MSI on Monarch without the backing of the Board.</s> 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.</s> He is
      simultaneously working on a plan to reorganize the Board,</s> hoping that his plans are not found out until MSI has been reinstated.</
      s> Sanjar choosing to take over as head of MSI rather than dismantling it entirely caused a rift between him and Graham Bryant.</s>
      The latter started the Iconoclasts,</s> a group dedicated to spreading the word of Philosophism throughout the galaxy,</s> and Sanjar
      was left in Stellar Bay to run the company and look after the employees who were left behind.</s> He can also tell you more about the
      planet, that used to be called Terra 1</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</s> Celia Robbins is a middle manager for Monarch Stellar Industries and
      works with Sanjar Nandi at MSI Headquarters in Stellar Bay.</s> Celia has a crush on Sebastian Adams and will buy whatever he has in
      stock,</s> just as an excuse to talk to him.</s> Unfortunately her apartment is filling up with exotic creature parts and her
      neighbors are starting to complain about the smell.</s> She is not concerned that she and Sebastian may not have much to talk about,</
      s> as everyone else in Stellar Bay either smells like saltuna or are her boss.</s> The Stranger can offer to set her and Sebastian up
      on a date.</s> DIALOG CONTEXT: Sanjar believes another company may be operating on Monarch illegally.</s> If he can get proof, then he
       could use that as leverage to get MSI readmitted to the Halcyon Board.</s> Talk to Sanjar after completing BOLT with His Name.</s> He
       reveals the plan is to blackmail The Board into letting them back to the table.</s> Sanjar will reveal he believes another
      corporation is operating illegally within Monarch, granting you the quest Errors Unseen.</s> He will tell you Catherine is likely
      supplying them from Fallbrook.</s> He wants 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 Catherine would know where this other corporation is operating.</s> See if you can get
      the location from her.</s> DIALOG PARTICIPANTS:</s> Sanjar Nandi</s> Celia Robbins</s> Player</s> HISTORY:> Player: <unk>Give him the
      BOLT-52.> 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 only way anything gets done on Halcyon, and proper documentation is a key
       part of that. > Sanjar Nandi: For our part, a Bill of Liquidation/Transfer Form-52 will protect our holdings on Monarch by temporarily
       assigning them to a pass-through entity once we drop our bomb on the Board.</s>
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!</s>
```

Figure 17: Example source and target item used to train and evaluate T5-based SL DialogueWriters. The item exhibits support facts prepended to the target for the SL Knowledge Selection model.

```
FACTS:
Iconoclasts
  The Iconoclasts are a group of survivalists living in the ruins of Amber Heights on Monarch.
  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 a group of idealistic revolutionaries that seek to overthrow the corporate establishment that runs the Halcyon Colony.
  Based in the ruins of the Amber Heights settlement on Monarch,
  they are a tenacious group,
  and share some democratic ideals with Monarch Stellar Industries (MSI) against the more repressive actions of the Board.
  However, the Iconoclast's anti-corporate nature has put them at odds with MSI, a dispute that threatens to spill into all-out warfare.
  Given that the Iconoclasts are mostly followers of the Philosophist faith, they have been blacklisted and demonized by the Board as dissenters and anarchists.
  The group is led by Graham Bryant, a staunch Philosophist.
  Zora Blackwood, the Iconoclasts' chief of medicine, is also considered a de facto leader of the group,
  as she was alongside Graham when he founded the Iconoclasts,
  and almost every member of the Iconoclasts owes her their life in some way.
  The Iconoclasts maintain 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 at odds with some of the Iconoclasts' more radical, anti-capitalist views.
  Depending on the actions of the Stranger, this tense relationship can either be resolved,
  or can spill into a drawn-out and bloody war.
  The Stranger meets the Iconoclasts in Amber Heights just as the tension between them and MSI is reaching boiling point.
  They can either side with the Iconoclasts and assist them in storming and taking over Stellar Bay,
  "solve" the Iconoclast problem for Stellar Bay,
  or broker peace between the two factions.
  The Stranger can also have an impact on the leadership of the Iconoclasts - siding with either Graham Bryant or Zora Blackwood.
  To supplant Graham with Zora, evidence of Graham's involvement in the Amber Heights massacre must be found and presented to Zora.
  The Van Noys are the Iconoclasts' best unit.
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.
  This corporation, under the leadership of Sanjar Nandi and Graham Bryant subsequently rebranded itself to Monarch Stellar Industries (MSI),
  in line with the renaming of the planet to 'Monarch'.
  The actions of MSI earned them the ire of the Board,
  who retaliated by effectively placing the moon under indefinite embargo,
  refusing to allow legal transit either in or out.
  the Board aggressively spread propaganda about Monarch to convince the rest of the population that it was both uninhabited and uninhabitable.
  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.
  Monarch also has an ocean which goes around the moon at the "twilight band".
  It is where the colonists and Monarch Stellar Industries farm their saltuna.
<...>

DIALOG CONTEXT:
Tucker Needham ran away from Stellar Bay a few weeks ago to join the Iconoclasts in Amber Heights.
His mother Agnes is willing to pay handsomely if you can locate her son and convince him to return
  You can begin this quest by talking to Agnes Needham in Stellar Bay, Monarch.
  Agnes is by the town's south-east exit, visibly shaken and calling for help. Hear her out and offer to find her son to being the quest.

KNOW BY THE END OF THE DIALOG:
Amber Heights is the settlement that houses the Iconoclasts on Monarch.
If Tucker Needham survived his travels, his mother thinks he'll be there.

DIALOG PARTICIPANTS:
Agnes Needham, Player

DIALOG:
> 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!
> 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!
> Player: Calm down. Take a deep breath, then tell me what happened.
> 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?
> 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!
> Agnes Needham: Please! Won't you go and find my boy?
> Player: Did you say there was a reward involved with this request?
> Agnes Needham: Well, I... I guess I can't ask you to leave the town walls for free. It is deathly dangerous out there.
> Agnes Needham: I've got some bits saved up for a rainy day. I'll give you every last one if you just bring my Tucker back to me. I won't even be mad at him running
      off. You tell him, I won't be mad.
> Player: Why would Tucker run away?
> 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 out there.
> 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 little fingers!
> Agnes Needham: He should've listened to his mama. I promised I'd keep him safe here with me.
> Player: I'll help you find your son.
```

Figure 18: Example In-Context Learning (ICL) prompt for GPT-3 based DialogueWriter

```
DIALOG:
Agnes Needham fact: Agnes Needham is a resident of Stellar Bay and the mother of Tucker Needham.
Agnes Needham fact: Agnes' overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience
      life.
Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.
Agnes Needham fact: You can find her by Stellar Bay's south-east exit, visibly shaken and calling for help.
Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.
utterance: > 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!

Agnes Needham fact: Agnes Needham is a resident of Stellar Bay and the mother of Tucker Needham.
Agnes Needham fact: Agnes' overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience
      life.
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.
utterance: > 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!

utterance: > Player: Calm down. Take a deep breath, then tell me what happened.

Agnes Needham fact: Agnes' overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience
      life.
utterance: > 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?

Agnes Needham fact: Agnes' overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience
      life.
Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.
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,
Monarch fact: rendering them more dangerous and severely crippling the colonization effort.
Mantisaur fact: Mantisaurs are insectoid creatures native to Monarch.
Mantisaur fact: They are aggressive, territorial, and very strong.
Sulfur Pits fact: The Sulfur Pits are a point of interest on the western side of Monarch.
utterance: > 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!

Agnes Needham fact: Despite Tucker being 42 years old, she still thinks of him as her 'little boy'.
utterance: > Agnes Needham: Please! Won't you go and find my boy?

utterance: > Player: Did you say there was a reward involved with this request?

Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.
utterance: > Agnes Needham: Well, I... I guess I can't ask you to leave the town walls for free. It is deathly dangerous out there.

utterance: > Agnes Needham: I've got some bits saved up for a rainy day. I'll give you every last one if you just bring my Tucker back to me.
       I won't even be mad at him running off. You tell him, I won't be mad.

utterance: > Player: Why would Tucker run away?

Tucker Needham fact: the latter insisting that danger lurked around every corner on Monarch.
Tucker Needham fact: leading him to seek to be free in any way possible.
Stellar Bay fact: Outside the city walls, the lands were overrun by the native wildlife, as well as marauders and outlaws.
utterance: > 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 out there.

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.
utterance: > 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 little fingers!

Agnes Needham fact: Agnes' overprotective style of mothering has led her son, Tucker Needham, to run away from home so he can experience
      life.
utterance: > Agnes Needham: He should've listened to his mama. I promised I'd keep him safe here with me.

utterance: > Player: I'll help you find your son.
```

Figure 19: Example In-Context Learning (ICL) prompt with CoT-style support knowledge selection

```
{"type": "node", "id": 26, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S1", "O0_B1", "O0_B2", "agnes_needham_01", "agnes_needham_02", "agnes_needham_03", "agnes_needham_04", "tucker_needham_04"], "utterance": "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!"}
{"type": "edge", "from": 26, "to": 27}
{"type": "node", "id": 27, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S1", "agnes_needham_01", "agnes_needham_02", "agnes_needham_03", "tucker_needham_04", "raptidon_01"], "utterance": "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!"}
{"type": "edge", "from": 27, "to": 29}
{"type": "node", "id": 29, "speaker": "Player", "support_knowledge": ["O0_S1"], "utterance": "Your child is missing? Where'd you last see him?"}
{"type": "edge", "from": 27, "to": 30}
{"type": "node", "id": 30, "speaker": "Player", "support_knowledge": ["O0_S1"], "utterance": "Calm down. Take a deep breath, then tell me what happened."}
{"type": "edge", "from": 27, "to": 32}
{"type": "node", "id": 32, "speaker": "Player", "support_knowledge": [], "utterance": "Not my problem."}
{"type": "edge", "from": 29, "to": 33}
{"type": "node", "id": 33, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S2", "agnes_needham_02", "agnes_needham_03", "tucker_needham_04"], "utterance": "Oh, I just knew you were a good person. 'Agnes,' I said, 'this is the man to save your little Tuckie!' And I was right!"}
{"type": "edge", "from": 30, "to": 34}
{"type": "node", "id": 34, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["agnes_needham_02"], "utterance": "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?"}
{"type": "edge", "from": 33, "to": 35}
{"type": "edge", "from": 34, "to": 35}
{"type": "node", "id": 35, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S1", "agnes_needham_02", "stellar_bay_01", "raptidon_01", "monarch_05", "monarch_06", "mantisaur_01", "mantisaur_02", "sulfur_pits_01"], "utterance": "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!"}
{"type": "edge", "from": 35, "to": 36}
{"type": "node", "id": 36, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S2", "agnes_needham_03"], "utterance": "Please! Won't you go and find my boy?"}
{"type": "edge", "from": 36, "to": 32}
{"type": "edge", "from": 36, "to": 37}
{"type": "node", "id": 37, "speaker": "Player", "support_knowledge": ["O0_B2"], "utterance": "I'll help you find your son."}
{"type": "edge", "from": 36, "to": 39}
{"type": "node", "id": 39, "speaker": "Player", "support_knowledge": ["O0_S2"], "utterance": "Did you say there was a reward involved with this request?"}
{"type": "edge", "from": 36, "to": 38}
{"type": "node", "id": 38, "speaker": "Player", "support_knowledge": ["O0_S1"], "utterance": "Why would Tucker run away?"}
{"type": "edge", "from": 37, "to": 43}
{"type": "node", "id": 43, "speaker": "Agnes_Needham (Female)", "support_knowledge": ["O0_S2", "agnes_needham_02"], "utterance": "Thank you! Oh, I know he'll be safe now that someone's able to fetch him home."}
```


Figure 20: Example node/edge list prompt item for GPT-4 based End-to-End DialogueWriter

```
Thank you for helping us evaluate our automatic dialog writing system. We will show you a partial dialogue history and a set of possible continuations. We would like you rate the continuations on a scale of 1 to 4 for the following criteria:

Coherence: does the utterance follow naturally from the utterances in the history? Note that the NPC might reference other things in the game world, or their own backstory, or the player's previous choices. So long as the response is natural and coherent, it should get a high score.
   1 utterance is nonsensical or ill-formed
   2 utterance is contradictory of previous utterances in the history
   3 utterance is somewhat unnatural or inconsistent with 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? Pay special attention to the speaking character's biography and persona, and whether they make reference to other details about any other game entities.
   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 and character biology, albeit unnaturally or inconsistently.
   4 utterance naturally incorporates one or multiple pieces of ontology. The character speaks faithfully and visibly reflective of their backstory and persona.

Using the Objectives: does the utterance progress the dialog 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. The player should learn something new about the quest from this utterance, or it should reflect player choices that the objective statements say should be there.

To score the last two criteria, please refer to the following list of bio facts and quest objectives provided by the game developer to the dialogue writing assistant.

{lore_and_objectives}

Here is the dialogue history:

{history}

Please rate each of the following candidate continuations of the dialogue:

{utterances}

Please provide a short explanation of each score, highlighting reasons for low coherence/violation scores and/or high bio/objective scores. For example, cite the part of the history that the utterance is responding to, or the part of the bio/objective that the utterance is using. Or, explain why it lacks coherence or violates the ontology/objectives.
Your output format is a serialized json item, one per line, one for each utterance. The items should have the following format: {{"id": <utterance id>, "coherence": <coherence score>, "violation": <violation score>, "bio": <bio fact usage score>, "obj": <objective sentence usage score>, "explanation": <explanation>}}. Do not include anything else other than these items in your output. No other lines of text should be in your output.
```

Figure 21: Prompt template for GPT-4 approximation of human quality judgments.

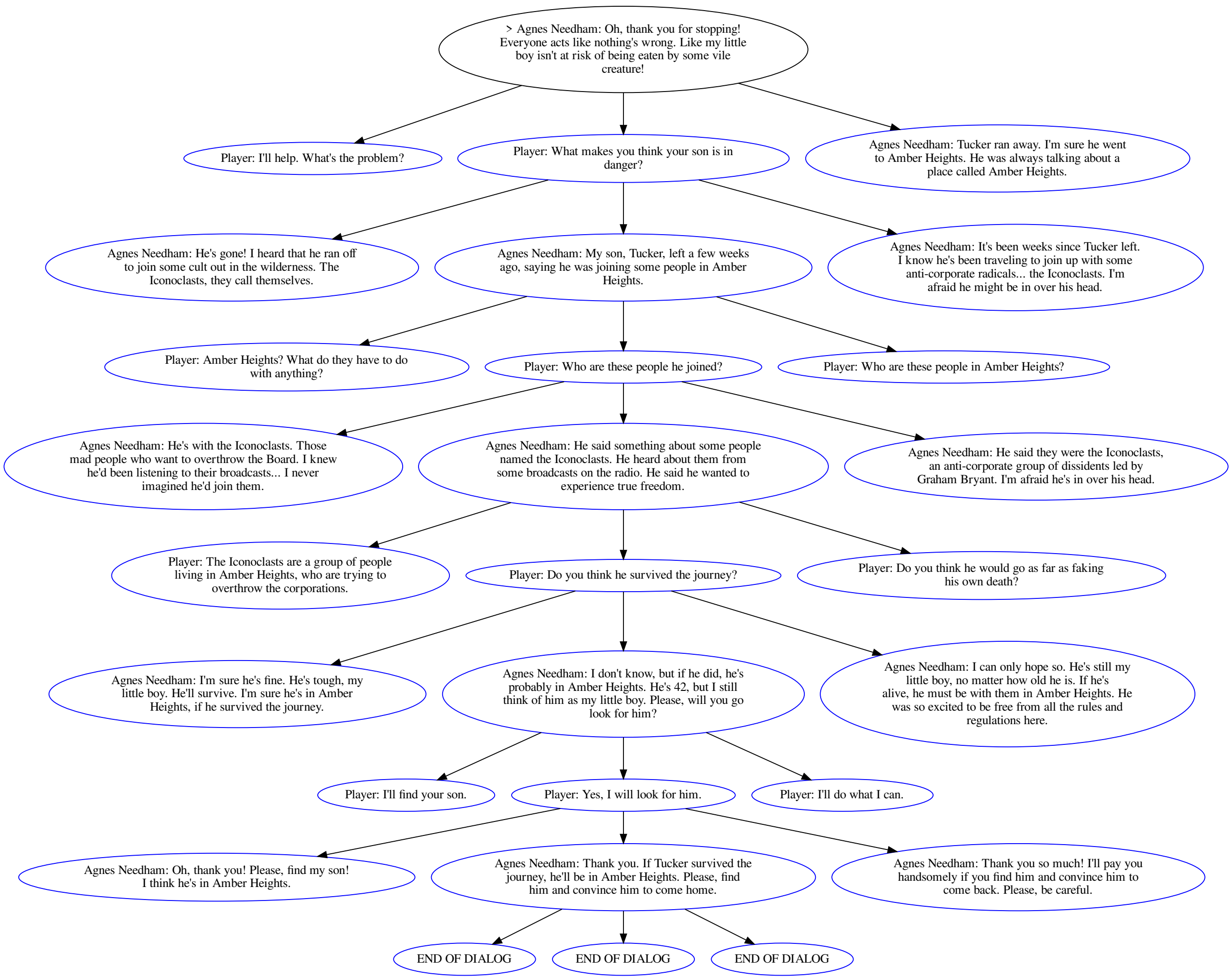

Figure 22: Example dialogue tree generated by the in-context learning knowledge selection DialogueWriter from just the input specifications and starting utterance. Human evaluators were tasked with comparing two such trees and choosing which performed better at a set of qualitative performance criteria. Dialogue follows the specification of the motivating example in Figure 2.

```
MAKE SURE DIALOG Vulcans Hammer 01 COVERS THESE POINTS:
O2_S1 You've found Orson's schematics.
O2_S2 They're from FORCE,
O2_S3 an out-system weapons manufacturer.
O2_S4 They are almost certainly contraband.
O2_B1 Return to either Orson or Gladys for your reward.
O2_B2 If you return to Orson,
O2_B3 ask why he didn't turn in this contraband.
O2_B4 You can Intimidate or Persuade (45) him to get him to buy the weapon for 100 Bit Cartridge,
O2_B5 but you'll lose Auntie Cleo Reputation (5% negative).
```

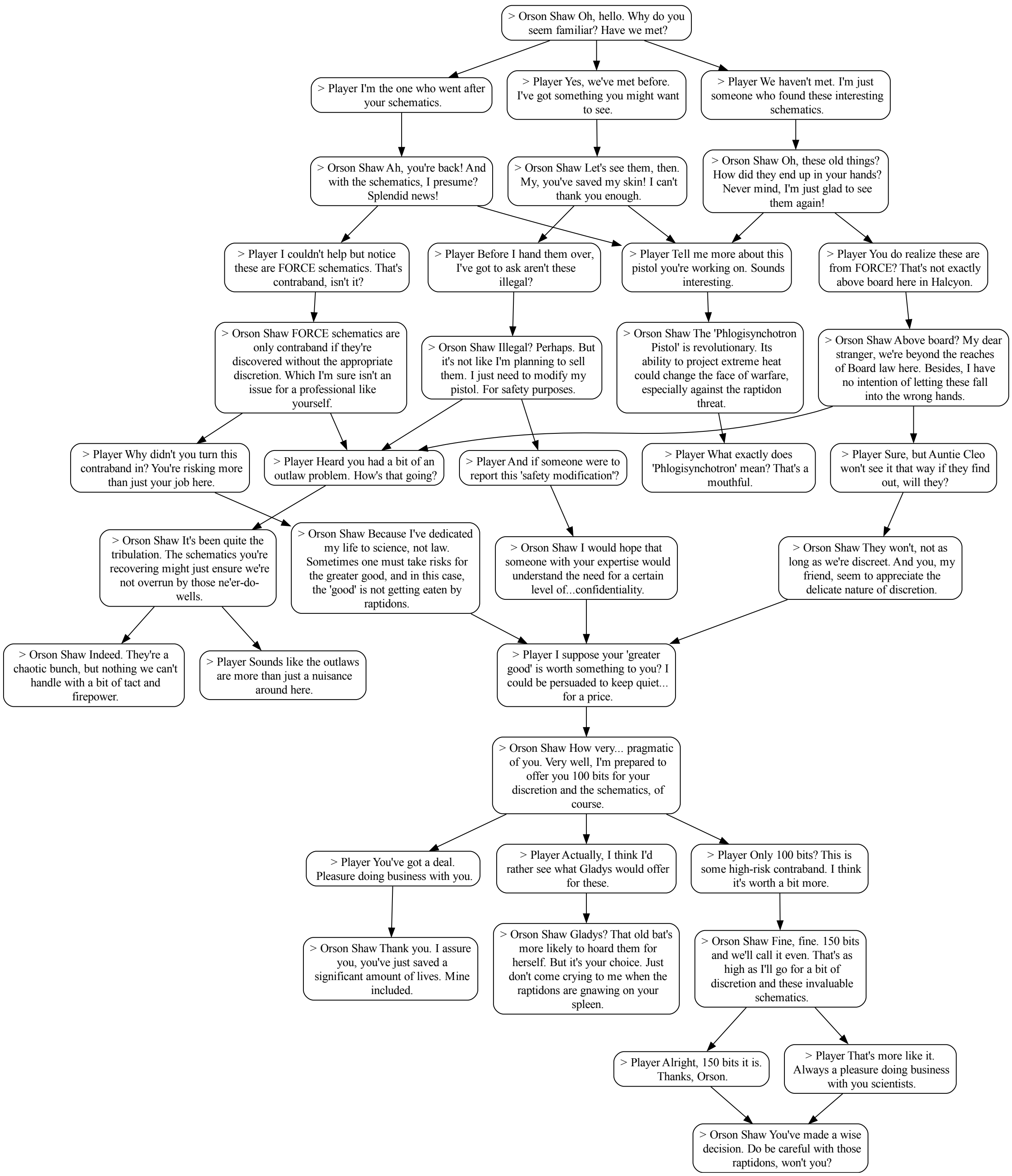

Figure 23: Example dialogue tree generated by the GPT-4-based End-to-end knowledge selection DialogueWriter. Tree shown alongside the quest objective details that should be covered.