Creating an AI Observer: Generative Semantic Workspaces

Anonymous ACL submission

Abstract

An experienced human Observer reading a document –such as a crime report– creates a succinct plot-like *"Working Memory"* comprising different actors, their prototypical roles and states at any point, their evolution over time based on their interactions, and even a map of missing Semantic parts anticipating them in the future. *An equivalent AI Observer currently does not exist*. We introduce the **[G]**enerative **[S]**emantic **[W]**orkspace (GSW) – comprising an *"Operator"* and a *"Reconciler"* – that leverages advancements in LLMs to create a generative-style Semantic framework, as opposed to a traditionally predefined set of lexicon labels. Given a text segment C_n that describes an ongoing situation, the *Operator* instantiates actor-centric Semantic maps (termed "Workspace instance" $\mathcal{W}_n$). The *Reconciler* resolves differences between $\mathcal{W}_n$ and a "Working memory" $\mathcal{M}_n^*$ to generate the updated $\mathcal{M}_{n+1}^*$. GSW outperforms well-known baselines on several tasks ($\sim 94\%$ vs. FST, GLEN, BertSRL - multi-sentence Semantics extraction, $\sim 15\%$ vs. NLI-BERT, $\sim 35\%$ vs. QA). By mirroring the real Observer, GSW provides the first step towards Spatial Computing assistants capable of understanding individual intentions and predicting future behavior.

1 Overview

We introduce a novel computational framework termed the **[G]**enerative **[S]**emantic **[W]**orkspace for estimating the latent Semantic structure in natural language from its embodiment in large-scale data using Large Language Models (LLM) (see Fig. 1). GSW differentiates itself from existing methods (Kingsbury and Palmer, 2002; Baker et al., 1998; Schuler, 2005) by introducing an "Observer"-centric formalism (see Fig. 1, Sec. 2). In addition to bringing in a novel computational aspect to the study of Semantics, such a framework would enable applications such as digital assistants in Spatial Computing (SC) and Augmented Reality (AR) interfaces, that require an Observer-centric view of the world.

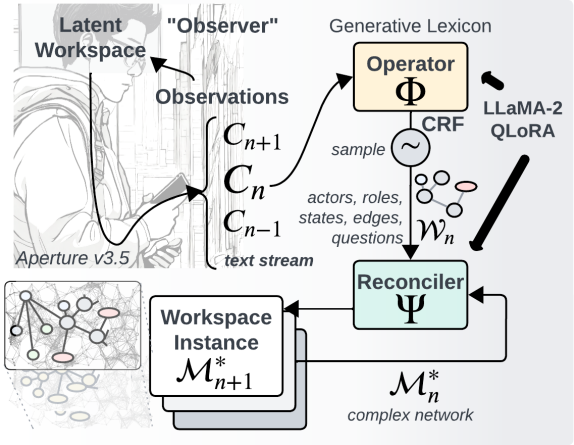

Figure 1: **The GSW framework:** The Observer constructs an internal map of the Semantics common across several instances of a situation, and in the process, identifies the recurring and prototypical conceptual gestalts. Any unfolding instance of a situation is processed through such semantic "lenses" and embedded in natural language with the use of grammar: a sample from the Observer's semantic map, with encodings of instance-specific actors, their interactions, and their evolution over space and time. In the GSW model, the goal is to construct an AI equivalent that interprets situations as they are encoded in text, to create a succinct *"Workspace"* instance – or working memory – that contains an extendable layout of the Semantics (see Tab. 6 for comparison to baselines). The Workspace is modeled as a Conditional Random Field (CRF), with Actors, Roles, States, Questions, and Predicates sampled from a conditional distribution. The CRF is estimated using a multi-task LLaMA + LoRA setup (*Operator*). The *Reconciler* computes the similarity between a pair of *"Workspace instances"* sampled from the CRF (see Tab. 8 for a comparison to baselines).

Computationally, GSW comprises two parts: (a) The *Operator*, a parameterized generative model of actor-centric Semantics as a Conditional Random Field (CRF) – a sample instance is presented in Tab. 1, and (b) The *Reconciler*, a fine-tuned LLM model that compares and resolves a pair of

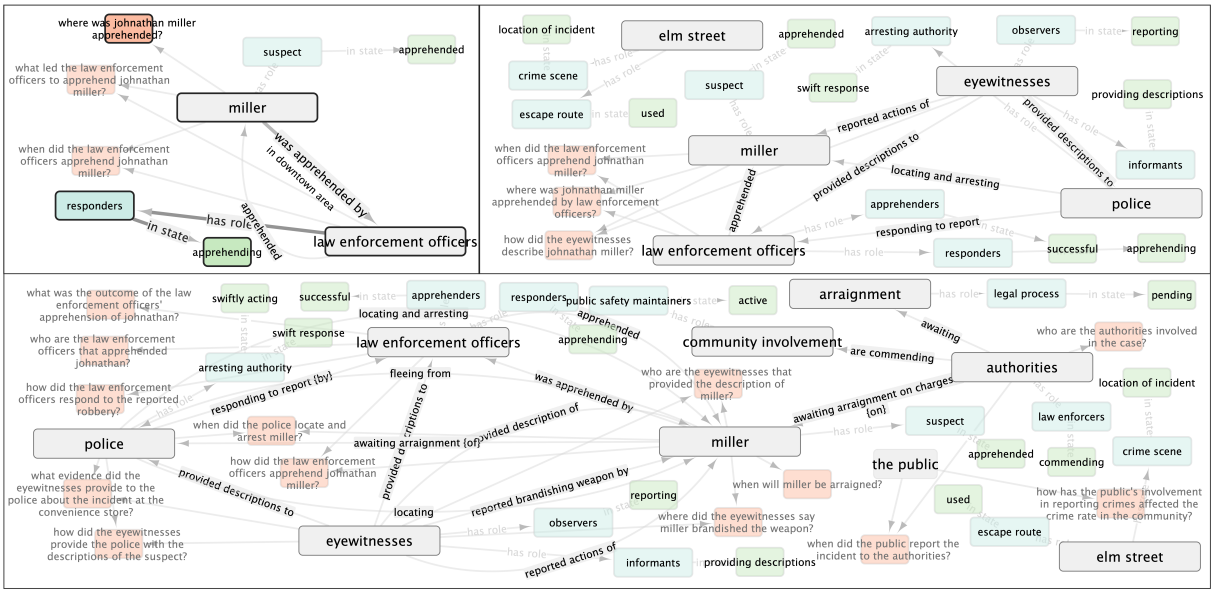

Figure 2: **Operator + Reconciler - Dynamics of the Generative Semantic Workspace [GSW]:** Snapshots of the evolving Workspace instance are depicted (from top-left, clockwise) along different time points, as the GSW framework processes a story $[C_1, \ldots, C_N]$ (see Fig. 9 for full story). We denote the lexicon types - Actor: grey, Role: blue, State: green, Predicate: edge, and Question: red - the first instance of each is in bold in the first pane. The workspace instances $\mathcal{W}_n$ (for each C_n) are aggregated into a consensus Workspace Instance $\mathcal{M}^*_{n+1}$ (see Fig. 1) using the *Reconciler* to compare the consensus Semantics at the previous step $\mathcal{M}^*_n$ with the latest Operator output $\mathcal{W}_n$. More workspace instances are presented in App. Sec. G.

Operator-generated instances, so that logical and temporal continuity of actor semantics is preserved (see Tab. 9).

The rest of the paper is arranged as follows: In Sec. 2, we introduce an Observer-centric view of Semantics, App. Sec. A contains related computational efforts, Sec. 3.1 enumerates the two challenges in existing methods, Sec. 4 introduces the GSW formalisms, Secs. 5, Sec. 6 presents results and discussion. FAQ is presented in App. Sec. B, and GSW-generated Semantic maps are presented in App. Sec. G.

2 A Cognitive Perspective for Observer-centric Semantics

We begin by walking through the following post from an Observer's perspective to get a better sense of the Semantics critical to them:

[Sample S1] Yesterday, in a swift response to a reported ***robbery***, **law enforcement officers** apprehended **Johnathan Miller**, a 32-year-old resident of **Greenview Avenue**, in the downtown area.

The Observer upon reading this first encounters the actors "law enforcement officer", "johnathan miller", "greenview avenue", "robbery", and, next comes up with *"consensus"* identifiers that convey the underlying Semantics: Stereotypical **role** identifiers such as the *Suspect*, *Crime Scene*, and *Apprehenders*, along with their respective **state** identifiers (e.g., "Where in the process is the suspect now?" $\rightarrow$ "apprehended"), delineate the responsibilities of the various **actors** in the situation. Role-states pre-condition the potential inter-actor relationships or *predicates* ("responded to", "apprehended by"). Unanswered **questions** ("What led to...," "How did law enforcement...") act as placeholders for the missing Semantics parts that are currently unanswered in the post. A summary is presented in Tab. 1.

The goal of an AI Observer is to be able to generate such semantic maps. It is thus instructive to create a framework of how human Observers are able to create and understand such maps. An Observer's focus on Semantics is oriented toward the actors involved in a situation and the inter-actor dynamics, a fundamental principle encapsulated in *"Situated Cognition"* (Clancey, 1997). The Observer lives in an uncertain world with partial information forcing them to create a structured probabilistic model of potential outcomes via the roles, and states (constraining the inter-actor interactions) that the Observer attributes to the actors. *Thus, a particular role, state, or interaction signifies a distribution* over ensuing outcomes, and cannot be

described in entirety by few samples. This poses a communication challenge when two Observers –who have experienced the same situation – have to communicate Semantic information over natural language. The Observers are forced to come up with a restricted set of consensus labels and syntax (or ***Identifiers***) that elicit similar distributions over the Semantics of the situation in both parties (Kripke, 1972; Jackendoff, 2017). Thus, the usage of the word "suspect" intends to convey an underlying distribution of outcomes in terms of other labels or words – such as accusation, freedom, penalties, and victimhood – which themselves correspond to distributions. A Conditional Random Field (CRF) (Sutton et al., 2012) is one of many probabilistic modeling tools that can succinctly represent such scenarios. If an Identifier cannot convey such a distribution over related concepts in a reliable way, then it is not an effective conveyor of Semantic information. These motivations are used to construct the GSW framework in Sec. 4 but first we review why some of the existing frameworks for constructing Semantic maps cannot be easily adapted for this task.

3 Can existing frameworks construct Observer-centric Semantics?

Over the last few decades, seminal works including PropBank, FrameNet, GLEN, ACE, and ConceptNet (Kingsbury and Palmer, 2002; Baker et al., 1998; Zhan et al., 2023; Doddington et al., 2004; Speer et al., 2017) have posited computational approximations to model the Semantic space and have addressed its modeling in two parts: (A) Constructing an explicit ***Lexicon*** – a set of pre-indexed *Identifiers* that act as proxies for the underlying concept distributions, and (B) An accompanying ***Text-to-Semantics*** model to map natural language (like **Sample S1**) onto the lexicon. For example, PropBank (Kingsbury and Palmer, 2002) *enumerates* role identifiers that attempt to identify stereotypical (thematic) participants in the usage of a *verb*; FrameNet (Baker et al., 1998) groups semantic role identifiers into larger groups known as Frames and adds Lexical Units (LU) – mostly verbs and nouns – that map a Frame to natural language. Frame Elements (FE) correspond to the roles. The Text-to-Semantics model is a post-hoc model that is rule-based or data-driven model (Chanin, 2023; Shi and Lin, 2019; Lu et al., 2021), which maps individual text segments onto this lexicon. The study of lexicons and Text-to-Semantics models has a long and rich history that is captured in great detail in the Related Works (see Appendix Sec. A).

Semantic Parts	Identifiers
ACTOR **specific named entities, and other noun phrases**	law enforcement officer, Johnathan Miller, Greenview Avenue, robbery
ROLE **stereotypical purpose embodied by each actor**	apprehenders, suspect, suspect residence, crime event
PREDICATE **interactions between actors in role-states**	law enforcement officers – responded to – reported robbery [yesterday] Johnathan Miller – apprehended by – law enforcement officers [in downtown area]
STATE **attributes of actor-roles determined by context**	successful, apprehended, inhabited, reported
QUESTION **unresolved valences that need answering**	What led to the apprehension of Johnathan Miller by the law enforcement officers? How did the law enforcement officers apprehend Johnathan Miller?

Table 1: **An Observer's distribution of semantics are aliased to identifiers in NL:** Identifiers serve to imply the Observer's internal Semantics distribution of the roles and states played by the different actors as well as the predicates with which these actors interact. Missing facets that the Observer expects to have answered in the future, are represented as unresolved questions. Readers interpret these identifiers by reconstructing the original Semantic distributions. GSW is able to model such distributions and constructs extendable Semantic maps as a result (see Fig. 2, App. Sec. G).

3.1 Challenges

Consider **Sample S1** mapped onto the FrameNet (Baker et al., 1998) lexicon using the Frame Semantic Transformer (FST) (Chanin, 2023), a Text-to-Semantics model (Tab. 2)[1]. In this setup, predicates (verbs and many nouns) behave as anchors that map to a specific subset of enumerated frame identifiers – such as *Response*, *Statement*, *Robbery*, *Arrest*, *Residence* – and text spans are associated to frame elements (FE) that are part of each frame. For example, the "Authorities" FE is the identifier representing a semantic role in the *Arrest* frame and corresponds to the "law enforcement officers" span. A thorough analysis reveals the following properties:

[L1] Fragmented Semantics: While the FrameNet lexicon identified (rows in Tab. 1) from

[1] Similar comparisons to Event and SRL frameworks have been performed at scale in Sec. 6.1

Frame	FE	Span
Response	Manner	swift
	Trigger	to a reported robbery
Statement	Message	Robbery
Robbery		
Arrest	Authorities	law enforcement officers
	Suspect	Johnathan Miller, a 32-year-old resident of greenview avenue
Residence	Resident	Johnathan Miller
	Location	of greenview avenue

Table 2: **Frame Semantic Transformer applied on Sample S1:** The retrieved frames, frame elements and text spans are presented above. These are compared to the Semantics identified by GSW in Tab. 6.

Sample S1 are meaningful at the frame-level, challenges arise when trying to ascertain the Semantics associated to a specific actor and their inter-actor relationships encoded in **Sample S1**. For instance, the *Response* frame neglects to address "Who responded?", the "Responding Entity" FE. Next, although "what the *Responders* responded to?" is mapped to the phrase "to a reported robbery" (FE "Trigger"), the frame "Robbery" is without any frame elements, though "Johnathan Miller" is a candidate for the "Perpetrator" FE, thus missing a vital connection between the *Response* and *Robbery* frames. Additionally, the *Statement* frame is misplaced – the lexical unit "reported" is being taken out-of-context – adding noise to the Semantic map since it's treated equally alongside more contextually pertinent frames.

[L2] Discontinuous Semantics: Fragmented Semantics at the sentence level exacerbates the challenges when trying to construct a dynamic semantic map, where actors, their roles, states, and their interactions evolve according to a consensus plot-like situation model. Consider for example, the result of applying FST on **Sample S2** which follows **Sample S1** (see Fig. 9 for the full story):

[Sample S2] Police swiftly acted on the provided descriptions, locating and arresting **Miller** within the hour.

The resulting semantics from FST are as follows: **Frame 1:** *Intentionally_act* → *Agent: police*, *Manner: swiftly*, *means: on the provided descriptions*, **Frame 2:** *Locating* → *Perceiver: Police*, *Sought entity: Miller*, **Frame 3:** *Arrest* → *Authorities: police*, *Suspect: Miller*, *Time: within the hour*. Establishing connections *between* the semantics from **Sample S1** to **Sample S2** requires connecting the dots better: Are the two *Arrest* frames identified in the adjacent samples referencing the same *arrest*? Are "*Police*" and "*Law Enforcement*" entities the same *authority* FE? The latest extraction lacks information about why *Miller* was arrested; there's no mention of the *Robbery* identified in the previous frame. We now propose an alternate Semantics framework that does not inherit these challenges.

4 The Generative Semantic Workspace

To motivate our approach, we first describe its underlying intuitive human-like workflow. Assume the Observer has encountered a ***situation*** S, and is reading X_S – the unrolling of a situation instance in S given by a sequence of text segments: $[C_1, \ldots, C_N] \in X_S$. For example, the Observer reads $C_1 =$ **Sample S1** and then reads $C_2 =$ **Sample S2**, and so on. As in Tab. 1, the Observer generates a ***Workspace instance*** $\mathcal{W}_1$ which comprises the actors, their roles, states, inter-actor relationships and unresolved questions. A network representation of $\mathcal{W}_1$ is shown in Fig. 2. Clearly we need a computational model – that we call the ***"Operator"*** – to codify this step of going from a text segment to the Workspace instance.

Next, the Observer reads **Sample S2** and generates another Workspace instance $\mathcal{W}_2$ *involving the same person*, *Miller* and crime, *robbery*. Clearly, the actors may have acquired new roles or states. For example, the question *"What led the law enforcement to apprehend johnathan miller?"* is unresolved given **Sample S1** and shows up as a valence node in $\mathcal{W}_1$ (see Fig. 2 (left, top)). However, upon reading **Sample S2**, we notice that the question is answered: "... *acted on the provided descriptions* ...". As a result, the question is now outdated and must be removed from the workspace instance (see Fig. 2, right top). This requires another computational unit – ***Reconciler*** – that takes the relevant parts of $\mathcal{W}_1$ and $\mathcal{W}_2$ and merges them into a ***Working memory*** instance $\mathcal{M}_2$ which will be then updated to create $\mathcal{M}_3$ by reconciling it with $\mathcal{W}_3$ and so on.

To formalize this intuitive model of the Observer, we denote the family of all possible workspace instances, including all possible working memory instances as the Workspace $\mathbb{W}$ (inspired by the concept of Global Workspace Theory (Baars, 1993)). A subspace specific to situation S is denoted $\mathbb{W}_S$. Then the **Operator** can be formally defined as $f_{\Phi_S} : \mathcal{X}_S \rightarrow \mathbb{W}_S$, and the **Reconciler**, $g_{\Psi_S} : \mathbb{W}_S \times \mathbb{W}_S \rightarrow \mathbb{W}_S$. Appendix B.1 explains

how LLMs (i.e. Φ_S, Ψ_S) make the dynamically created labels in workspace instances semantically consistent; that is, semantically similar context would lead to similar semantic labels.

4.1 Training the Operator model

Given an input sample C_n, such as **Sample S1**, the Operator model outputs a set of roles, states, and predicates for each actor that describe the inter-actor relationships, and questions about the Semantics that are yet to be answered going forward (see Tab.1 and Fig. 2). Formally, f_Φ is trained to output a Workspace instance W_n as sampled from a parameterized distribution across the workspace $\mathbb{W}$: i.e. $P_{\mathbb{W}}(\mathcal{W}_n|C_n;\Phi)$. Observer-centric Semantics discussed in Sec. 4 intuits that this distribution is best expressed as a CRF, where a set of Identifiers sampled from a Semantics distribution are in fact *themselves* representative of conditional distributions that lead to the generation of more granular Identifiers. In our setup, the CRF is comprised of conditional distributions estimated for each Semantic part from Tab. 1 – see Fig. 3. A sampled path through this CRF results in a Workspace instance.

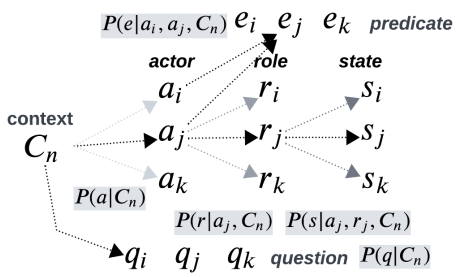

Figure 3: **Conditional Random Field model of the Workspace:** A workspace instance is constructed by sampling the CRF given an input text segment C_n.

4.2 Workspace instance as a Complex Network: A probabilistic interpretation

One can view the task of jointly estimating a workspace instance as the estimation of a *complex network*:

$$\mathcal{W} := \mathcal{W}(V, E),$$

where the nodes $v \in V$ and edges $e \in E$ correspond to the following lexicon map:
Node (v)**:** {Actor a, Role r, State s}; {Question q}
Edge (e)**:** Source Node $v_S \rightarrow$ Predicate $\rightarrow$ Target Node v_T.
Then, the Workspace instance itself can be viewed as a sample drawn from the Workspace:

$$\mathcal{W} \sim f_\Phi(C_n) \implies V_{\mathcal{W}}, E_{\mathcal{W}} \sim f_\Phi(C_n).$$

Assume conditional independence:

$$P(V_{\mathcal{W}}, E_{\mathcal{W}}|C_n) \approx \prod P(\{v_S, e, v_T\}|C_n).$$

$$P(\{v_S, e, v_T\}|C_n) = P(e|v_S, v_T, C_n) \times P(v_S, v_T|C_n).$$

Now, $P(v_S, v_T|C_i) = P(v_S|C_n) \times P(v_T|C_n)$. Each can be further decomposed:

$$\begin{aligned} P(v|C_n) &:= P(\{a, r, s\}|C_n) \\ &= P(s|a, r; C_n) \\ &\times P(r|a; C_n) \times P(a|C_n). \end{aligned}$$

4.3 Reconciler: Comparing Workspace Instances as Complex Networks

Reconciler Classification Labels	
Nodes, Edges	0: the existing workspace instance is sufficient 1: the new instance overwrites the workspace 2: both instances are important and unrelated
Questions	0: question remains unanswered and relevant 1: question is answered or irrelevant

Table 3: **Summary of classification tasks trained in the Reconciler:** Nodes and edges are classified pairwise. *Reconciliation* or REC applies on pairwise nodes and edges (0: keep previous, 1: replace, 2: keep both), and {question, node} and {question, edge} pairs are classified – *Question Resolution* or QR – as answered or irrelevant (1) or unanswered and relevant (0).

Consider two Workspace instances $\{\mathcal{W}_i, \mathcal{W}_{i+1}\} \in \mathbb{W}$: the Reconciler compares the two instances to construct an aggregated workspace instance $\mathcal{M} \in \mathbb{W}$. Since we represent the instances as actor-centric Complex Networks – $\mathcal{W}_i := \mathcal{W}_i(V, E), \mathcal{W}_{i+1} := \mathcal{W}_{i+1}(V, E)$,

$$\begin{aligned} &\mathcal{M} \sim g_\Psi(\mathcal{W}_i, \mathcal{W}_{i+1}) \\ \iff &\langle V_M, E_M \rangle \sim g_\Psi(\langle V_i, E_i \rangle, \langle V_{i+1}, E_{i+1} \rangle), \end{aligned}$$

the Reconciler can be formulated as a classification task, where every pair of nodes and edges are compared as follows:

$$d = g_\Psi(v_i, v_{i+1}), g_\Psi(e_i, e_{i+1}).$$

A summary of the possible decisions is presented in Tab. 3: (i) Retain the node/edge currently in $\mathcal{W}_i$ and discard the node/edge in $\mathcal{W}_{i+1}$ (as noise or infeasible); (ii) Discard the node/edge in W_i and incorporate the node/edge in $\mathcal{W}_{i+1}$; (iii) Keep both

$\mathcal{W}_i, \mathcal{W}_{i+1}$ nodes/edges. Questions are compared with both nodes and edges.

Despite the Reconciler being posed as a *local* node/edge-level model that doesn't consider the two Workspace instances in their entirety, the actor-centric nature allows modifications at a smaller neighborhood-level: *an actor's associated Semantics can only be influenced by the actors around them.* Empirically, such a Reconciler demonstrates high performance at aggregating Workspace instances: see classification performance in Tab. 8 and networks in App. Sec. G.

Actor-Centric Memory and Autoregressive Semantics: With the Operator and Reconciler trained, they can be paired in an auto-regressive setup: given a sequence of consecutive text segments: $\{C_1, \ldots, C_N\} \in \mathcal{X}$, the Operator outputs Workspace instances $\{\mathcal{W}_1, \ldots, \mathcal{W}_N\} \in \mathbb{W}$, and the Reconciler is deployed in an auto-regressive fashion to continually update a consensus Workspace instance $\mathcal{M}^*$ such that:

$$\mathcal{M}^*_{n+1} \leftarrow g_\Psi(\mathcal{M}^*_n, \mathcal{W}_n).$$

4.4 Practical Considerations

4.4.1 Data

News reports are a popular resource for sampling semantics-rich stories that belong to universally recognized situation patterns (in an effort to cater to a wider readership). We query GDELT (Leetaru and Schrodt, 2013), a Jigsaw-powered news-indexing platform, with Situation identifiers to retrieve a small set of situation-conditioned `en_US` articles. Tab. 4 presents statistics about the data. These situations were manually selected as an initial seed set – similar to FrameNet's early versions containing few frames (Lowe, 1997) – to assess the validity of the GSW framework (see FAQ B.2 for more on our choice of situations).

Situation Label	Documents	Sentences	Tokens
crime and justice	80	1209	100,635
fire fighting	79	1116	87,901
technology development	81	1334	122,493
healthcare	81	1259	117,962
economy	78	1264	110,605

Table 4: **Data Statistics:** Situation-specific news reports are sampled from GDELT. Each document (or article) is split into short contexts $C_1, \ldots, C_N$ (of 3 sentences). Just ~80 articles per situation is sufficient to finetune LoRA adapters to project *any* text segment onto the Semantic workspace.

4.4.2 Silver-Standard Annotations and Computational Requirements

GPT-4 (Achiam et al., 2023) is found to be excellent at generating Workspace instances by using carefully (Ekin, 2023) constructed multi-turn prompts (see Codebase) facilitated by Guidance[2] and given short contexts $\ldots, C_i, \ldots, C_j, \ldots$ (3 sentence window) drawn from GDELT news articles. However, smaller, open-source models such as LLaMA (Touvron et al., 2023) and Mistral (Jiang et al., 2023) models (including Chat-pretrained) struggle to output the semantics without finetuning. In order to construct open-source, hostable, but still powerful approximations of GPT-4 Operators locally, adapters such as LoRA (Hu et al., 2021) are now available as data- and parameter-efficient tunable modules (see PeFT (Li and Liang, 2021; Zhang et al., 2023)). That is:

$$f_{\hat{\Phi}}(C_i, \mathcal{T}) \sim f_\Phi(C_i, \mathcal{T}).$$

Different QLoRA (Dettmers et al., 2023) adapters are learned for each situation and can be inserted into a shared LLaMA backbone to match situation-specific input contexts (plug-and-play) (see Fig. 4). Each adapter only contains 1M trainable parameters (rank-2) and requires a surprisingly few number of data samples to be finetuned (see Tab. 4). FAQ B.4 details our training setup.

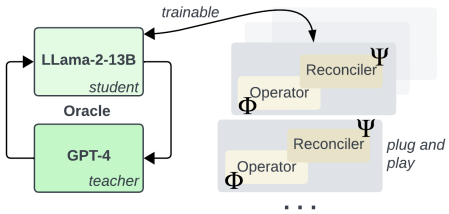

Figure 4: **Modular parameter sets for the Workspace model:** Every Operator and Reconciler model (specific to a situation) comprises $< 1M$ parameters – using PeFT and LoRA – and relies on the same shared LLM Oracle (LLaMA-2-13B).

5 Evaluation Methods

We evaluate the Operator and Reconciler across multiple situations listed in Tab. 4 using human evaluation via UpWork service as well as existing model baselines detailed below. An ablation study across situations and model architectures is provided in the Appendix (see Sec. E, B.3, F).

[2] www.github.com/guidance-ai/guidance

Operator: The workspace instances $\mathcal{W}_i$ generated from C_i are evaluated with respect to baselines by UpWork workers (also see App. Sec. D): **[UpWork Task 1]** Perform a side-by-side blind evaluation with existing workspace instance generations output by state-of-the-art (SOTA) workspace model *baselines*: (i) Frame Semantic Transformer (FST) (Chanin, 2023), (ii) BERT for Semantic Role Labeling (BertSRL) (Shi and Lin, 2019), and an (iii) Event Extraction (GLEN) (Zhan et al., 2023); **[UpWork Task 2]** Given questions generated by the Operator and GPT-4 from the same corpus, perform a blind rating of question quality on a scale from 1(worst) - 5(best).

Task	Baseline	Reconciler Labels
REC/NLI	Entailment	0 (Keep old)
	Contradiction	1 (Replace)
	Neutral	2 (Keep old and new)
QR/QA	0 (Unanswered)	0 (Unanswered & relevant)
	1 (Answered)	1 (Answered or irrelevant)

Table 5: **An approximate mapping from Baselines to Reconciler tasks:** Pre-existing NLI baseline models can be repurposed as a Reconciler and similarly a Question-Answering model can, to an extent, perform Question Reconciliation.

Reconciler: We identify 2 tasks and baselines for the Reconciliation and Question Resolution tasks: [**B1**] *Natural Language Inference (NLI):* Given a pair of ordered text segments – the *premise* and the *hypothesis*, output either (a) *entailment:* the hypothesis is entailed by the premise; (b) *contradiction:* the hypothesis contradicts the premise; or (c) *neutral:* the hypothesis and premise can co-exist. A suitable mapping of labels from the Reconciliation task to NLI is shown in Tab. 5. Models: RoBERTa (Liu et al., 2019) and DeBERTa (He et al., 2020) fine-tuned on the SNLI (Bowman et al., 2015) dataset. [**B2**] *Question-Answering:* Given a *context, question* tuple, output an answer as a span within the input context. A question is *resolved* if QA returns a non-empty span. Models: RoBERTa (Liu et al., 2019) fine-tuned on SQuAD (Rajpurkar et al., 2016).

6 Results and Discussion

We present qualitative and quantitative results for the Operator and Reconciler models below. Several Workspace instances created as a result of parsing short stories are presented in App. Sec. G.

Ours (Baseline)			
Situation	vs. Zhan et al (GLEN)	vs. Shi & Lin (BERT-SRL)	vs. Chanin (FST)
crime & justice	**0.90** (0.10)	**0.96** (0.04)	**0.7** (0.3)
economy	**0.98** (0.02)	**0.96** (0.04)	**0.86** (0.14)
firefighting	**0.98** (0.02)	**0.98** (0.02)	**0.79** (0.21)
healthcare	**1.0** (0.00)	**0.96** (0.04)	**0.94** (0.06)
tech. dev.	**0.96** (0.04)	**0.96** (0.04)	**0.86** (0.14)

Table 6: **Operator Evaluations - Part 1 - Comparison with Existing Frameworks:** Given a short context, English-speaking annotators are shown the unlabeled outputs of the Operator and a baseline framework (GLEN, BERT-SRL, FST) and asked to select the one which best summarizes the semantics in the text. The Operator is preferred over baselines across situations.

Situation	GPT-4 (d*)	Operator (d)	| d - d* |
crime & justice	4.45 +/- 0.98	4.47 +/- 0.81	= 0.02
economy	4.84 +/- 0.50	4.74 +/- 0.72	= 0.10
firefighting	4.80 +/- 0.66	4.63 +/- 0.91	= 0.17
healthcare	4.88 +/- 0.47	4.53 +/- 1.14	= 0.35
tech. dev.	4.84 +/- 0.67	4.61 +/- 0.94	= 0.23
(Avg.: 4.68)	4.76	4.60	

Table 7: **Operator Evaluations - Part 2 - Quality of Generated Questions:** English-speaking annotators are asked to rate the quality of Operator-generated questions given a short context C_i on a scale from 1(worst)-5(best). Questions generated by GPT-4 for the *same* contexts are used as a baseline. The Operator model is observed to generate questions of similar quality to GPT-4.

6.1 Operator

We present the comparison with workspace instance baselines in Tab. 6. Operator outputs on a held-out test set are strongly preferred ($> 90\%$ outputs selected) over GLEN (Zhan et al., 2023), BertSRL (Shi and Lin, 2019) and Frame Semantic Transformer (Chanin, 2023) (>70%) for all situations. A comparison of Operator-generated questions with GPT-4 is presented in Tab. 7. Operator question quality is highly rated across all situations (4.6 average rating) and is close to the GPT-4 quality for each situation (< 0.35 avg. rating difference).

6.2 Reconciler

The classification results of the Reconciler are presented in Tab. 8. The LLaMA + LoRA Reconciler outperforms the DeBERTa (He et al., 2020) and RoBERTa (Liu et al., 2019) baselines by a wide margin with an improvement of $> 25\%$ at the task of Reconciliation and $> 50\%$ at Question Resolution on Accuracy (ACC) and F1 score (weighted)

	ACC. ↑			F1 ↑			
Situation	**Reconciliation**						
	He et al.	Liu et al.	**Reconciler**	He et al.	Liu et al.	**Reconciler**	***Sensitivity***
crime and justice	0.72	0.71	**0.81**	0.69	0.68	**0.81**	***0.91***
fire fighting	0.71	0.69	**0.89**	0.68	0.66	**0.89**	***0.99***
technology development	0.74	0.69	**0.87**	0.71	0.61	**0.87**	***0.92***
healthcare	0.74	0.70	**0.91**	0.71	0.65	**0.91**	***0.96***
economy	0.76	0.70	**0.88**	0.74	0.63	**0.88**	***0.96***
	Question Resolution						
	Liu et al. (Base)	Liu et al. (Large)	**Reconciler**	Liu et al. (Base)	Liu et al. (Large)	**Reconciler**	**-**
crime and justice	0.53	0.59	**0.92**	0.49	0.57	**0.92**	-
fire fighting	0.63	0.62	**0.97**	0.60	0.60	**0.98**	-
technology development	0.74	0.67	**0.99**	0.72	0.64	**0.99**	-
healthcare	0.69	0.68	**0.97**	0.67	0.66	**0.97**	-
economy	0.63	0.64	**0.94**	0.60	0.60	**0.94**	-

Table 8: **Reconciler performance on different situations across Reconciliation and Question Resolution Tasks and (pre-existing) Baselines:** The two (sub-)tasks – Reconciliation [REC] (3 classes), and Question Resolution [QR] (2 classes) – are described in Tab. 9. Baselines for the two subtasks include SOTA NLI (Bowman et al., 2015) and QA (Liu et al., 2019) pipelines – see Sec 5 – that are existing tasks that share similar goals. As seen across situations, the Reconciler outperforms the baselines handily (REC - $> 25\%$, QR - $> 50\%$).

on a held-out test set. Reconciliation is a 3-class classification task and Question Resolution is a binary classification task; the test set is made sure to be class-balanced. **Sensitivity** (see Tab. 8 - right) merges Class 1 (replace) and Class 2 (keep both) in the Reconciliation task. A high Sensitivity score indicates that the Reconciler model is capable of correctly deciding whether information from a new Operator instance is to be added to the consensus workspace instance $\mathcal{M}^*$ or ignored.

7 Concluding Remarks

An Observer-centric semantic representation scheme is an essential evolutionary construct that has allowed species to predict and dynamically adapt to an uncertain and constantly changing world, and the proposed computational approximation –*Generative Semantic Workspace*–promises to add similar abilities to NLU frameworks. GSW demonstrates considerable success at constructing consensus Workspace instances – colloquially referred to as "Working memory" – in which actor-centric Semantics are aggregated (by the *"Operator"*) and modified continually (by the *"Reconciler"*) given a stream of text segments. Evaluations using human annotators and baseline NLU models have revealed that the resulting Workspace instances convey a plot-like Semantic summary that is more comprehensive than its counterparts. Like a true Working memory, the consensus workspace instances generated by GSW, incrementally and in a compositional manner, stitch together the salient Semantic features of actors – the roles, states – and the inter-actor relationships (predicates). Unanswered questions provide a means of looking to the future to complete the missing valences of the past.

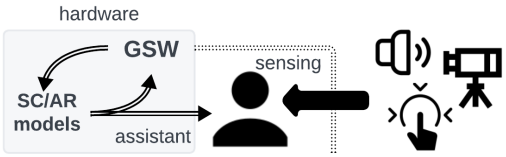

Figure 5: *A prospective use-case of GSW - Targeted information retrieval of Semantics in SC/AR settings.*

Among the several prospective real-world applications, the Complex Network representation of the Workspace instance may facilitate multi-hop actor-lexicon *traversals* – using such existing KG-QA methods as DeepPath (Xiong et al., 2017) – that can help uncover unrivaled, long-range Semantic connections across the Workspace: a novel use-case within Targeted Information Retrieval (or *Semantic Search*). More broadly, GSW can be envisioned to form the basis of a foundation model for augmented reality (AR) or spatial computing (SC) frameworks: a *real* Observer and a corresponding AI assistant encounter experiences in concert, which in turn enables the AI assistant to facilitate reasoning about Observer intentions and thus, predict, and even perhaps, *preempt* decision-making.

8 Limitations

8.1 Technical challenges in the GSW implementation

[*Coreference resolution*] Workspace instances require long-range coreference resolution to map entity mentions across multiple sentences to a shared actor (that constitutes a node). This module can at times pose inaccuracies. [*Data variety*] Training and testing of the Operator and Reconciler models are conducted on news media reports which are cleaner, linear, and structured when compared to other text datasets such as social media posts. [*Retrieval in the Reconciler:*] As the Workspace instance grows larger, the Reconciler has to infer more pairwise decisions between nodes and edges in the latest Operator output $\mathcal{W}_n$ and the consensus Workspace instance $\mathcal{M}_n^*$. We have implemented several heuristics to reduce the number of comparisons: for example, using actor mentions in $\mathcal{W}_n$ to identify a substantially smaller subgraph $\tilde{\mathcal{M}}_n^* \subset \mathcal{M}_n^*$ that is likely to be updated. Such heuristics may result in some Semantic parts becoming outdated if the subgraph $\tilde{\mathcal{M}}_n^*$ does not include all the nodes and edges that need updates. [*Recall vs. Precision:*] The Operator is trained with a carefully balanced negative sampling hyperparameter (40%), to avoid the generation of spurious lexicon samples – observed frequently in the generation of predicates – that comprise the Workspace instance: training data for negative samples are also retrieved from querying GPT-4 (see Codebase). As a result, the occasional input context may yield too many or too few Semantic parts.

8.2 Potential Risks and Ethical Considerations

Given that GSW uses a LLaMA-backbone, the same potential risks and ethical considerations apply as with any pretrained generative-style language model: that the Operator model (CRF of lexicon distributions) can reflect the bias in the finetuning dataset and as a result, can construct potentially harmful, violent, racist and sexist lexicon samples. But the bias can be more nuanced: roles attributed to actors (noun phrases) that are described in the training data in the same stereotypical roles, may be once again associated with the actor during inference *despite the input context suggesting an entirely different role*. Our experiments found very little evidence of this form of bias.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

Bernard J Baars. 1993. *A cognitive theory of consciousness*. Cambridge University Press.

Collin F Baker. 2017. Framenet: Frame semantic annotation in practice. *Handbook of Linguistic Annotation*, pages 771–811.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

David Chanin. 2023. Open-source frame semantic parsing. *arXiv preprint arXiv:2303.12788*.

William J. Clancey. 1997. *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge University Press, USA.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec*, volume 2, pages 837–840. Lisbon.

Sabit Ekin. 2023. Prompt engineering for chatgpt: A quick guide to techniques, tips, and best practices. *Authorea Preprints*.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Ray Jackendoff. 2017. In defense of theory. *Cognitive Science*, 41(S2):185–212.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Richard Johansson and Pierre Nugues. 2008. Dependency-based semantic role labeling of propbank. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 69–78.

Paul Kingsbury and Martha Palmer. 2002. From TreeBank to PropBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).

Saul A. Kripke. 1972. *Naming and Necessity*, pages 253–355. Springer Netherlands, Dordrecht.

Kalev Leetaru and Philip A. Schrodt. 2013. Gdelt: Global data on events, location, and tone. *ISA Annual Convention*.

Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Chenliang Li, Haiyang Xu, Junfeng Tian, Wei Wang, Ming Yan, Bin Bi, Jiabo Ye, Hehong Chen, Guohai Xu, Zheng Cao, Ji Zhang, Songfang Huang, Fei Huang, Jingren Zhou, and Luo Si. 2022. mplug: Effective and efficient vision-language learning by cross-modal skip-connections.

Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2020. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70.

Manling Li, Sha Li, Zhenhailong Wang, Lifu Huang, Kyunghyun Cho, Heng Ji, Jiawei Han, and Clare Voss. 2021. The future is not one-dimensional: Complex event schema induction by graph modeling for event prediction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5203–5215, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 1641–1651.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

John B Lowe. 1997. A frame-semantic approach to semantic annotation. In *Tagging Text with Lexical Semantics: Why, What, and How?*

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2795–2806, Online. Association for Computational Linguistics.

Andrew Mellor. 2017. The temporal event graph. *Journal of Complex Networks*, 6(4):639–659.

Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 104–111.

Joakim Nivre. 2010. Dependency parsing. *Language and Linguistics Compass*, 4(3):138–152.

Martha Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the generative lexicon conference*, pages 9–15. GenLex-09, Pisa, Italy.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2011. *Semantic role labeling*. Morgan & Claypool Publishers.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, and Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.

Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *Computational Linguistics and Intelligent Text Processing*, pages 100–111, Berlin, Heidelberg. Springer Berlin Heidelberg.

Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling. *ArXiv*, abs/1904.05255.

Elizabeth Spaulding, Kathryn Conger, Anatole Gershman, Rosario Uceda-Sosa, Susan Windisch Brown, James Pustejovsky, Peter Anick, and Martha Palmer. 2023. The DARPA Wikidata overlay: Wikidata as an ontology for natural language processing. In *Proceedings of the 19th Joint ACL-ISO Workshop on Interoperable Semantics (ISA-19)*, pages 1–10, Nancy, France. Association for Computational Linguistics.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Charles Sutton, Andrew McCallum, et al. 2012. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Luis von Ahn, Mihir Kedia, and Manuel Blum. 2006. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 75–78, New York, NY, USA. Association for Computing Machinery.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. *arXiv preprint arXiv:1909.03546*.

Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.

Wenhan Xiong, Thi-Lan-Giao Hoang, and William Yang Wang. 2017. Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Conference on Empirical Methods in Natural Language Processing*.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin Raffel, and Mohit Bansal. 2023. Ties-merging: Resolving interference when merging models.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2020. Xlnet: Generalized autoregressive pretraining for language understanding.

Qiusi Zhan, Sha Li, Kathryn Conger, Martha Palmer, Heng Ji, and Jiawei Han. 2023. Glen: General-purpose event detection for thousands of types. *arXiv preprint arXiv:2303.09093*.

Renrui Zhang, Jiaming Han, Chris Liu, Peng Gao, Aojun Zhou, Xiangfei Hu, Shilin Yan, Pan Lu, Hongsheng Li, and Yu Qiao. 2023. Llama-adapter: Efficient fine-tuning of language models with zero-init attention.

A Related Computational Models of Workspaces

Existing workspace modeling efforts have depended, to a large extent, on manually-annotated lexicon ontologies. Among the most popular efforts are PropBank (Kingsbury and Palmer, 2002) and FrameNet (Baker et al., 1998), which attempt to define a correspondence between (a) the syntactic "realizations" of the semantics *explicit* within the language structure of a sentence, and (b) a finite and discrete set of semantic "roles" (Levin, 1993). Both methods used expert linguists to define and annotate the role set and the role definitions. We detail the similarities and differences below:

PropBank: PropBank utilized a *bottom-up* approach: (1) Dependency Parse Trees (Nivre, 2010) were applied to a large text corpus to distill shared syntactic patterns ("Framesets") specific to each verb (a process known as "lexical sampling"). (2) For each Frameset, the corresponding sentences were manually annotated with an enumerated set of *arguments* `ARG:0, ..., ARG:N`. These arguments were later associated to verb-specific definitions using VerbNet (Schuler, 2005). The semantic roles are identified as corresponding *spans* within the sentence (commonly a `NP`, `NNP` subtree in the dependency parsing). For example, the sentence (A):

Officers captured **Sarah** at the **Sepulveda on-ramp** of the **405**.

would be annotated with the arguments:

Agent: **officers**, *Predicate*: **captured**, *Patient*: **Sarah**.

Perhaps the greatest benefit of PropBank was that its syntactic "grounding" made it possible for rule-based and early ML models (Johansson and Nugues, 2008; Shi and Lin, 2019) to *learn the task*

of distilling the semantics given a sentence, albeit within the confines of a *limited* ontology of > 3000 verbs and > 4000 Framesets

Event Databases: PropBank evolved in several directions, including efforts to unify it with related semantic lexicon such as VerbNet and FrameNet (Palmer, 2009; Shi and Mihalcea, 2005), or augment it via the DWD overlay (Spaulding et al., 2023) to WikiData (Vrandečić and Krötzsch, 2014). The latter of these efforts now manifests as "Event" databases (Liu et al., 2020; Xiang and Wang, 2019) such as the ACE (Doddington et al., 2004) and ERE (Aguilar et al., 2014) datasets, and led to the DARPA initiative of Event identification/extraction challenges. Events are best motivated by their related identification tasks: Given a sentence, identify the event(s) – from a set of *hundreds* of events in a pre-annotated schema (Zhan et al., 2023; Wadden et al., 2019; Lu et al., 2021) – that the sentence is referring to. For example, (A) would be annotated with the *Capture* event.

FrameNet: In contrast to PropBank and related Event ontologies, FrameNet[3] utilizes a *top-down* approach that is not tethered to the syntax structure. Rather, expert linguists aggregated roles (redefined as "Frame Elements" (FE)) from a large corpus of sentences, which are *known to co-exist* under a conceptual gestalt, or "Frame". Each frame additionally comprises a set of "Lexical Units" (LU) - valences (mostly verbs and nouns) whose occurrence in a sentence increases the likelihood of a frame. For example,

the *Frame*: **Taking Captive**

would contain the following frame elements and lexical units:

FE: *Agent*, *Captive*, *Cause*
LU: *capture.v*, *secure.v*

FrameNet (1000s of frames and $10,000$s of FE) is a substantially larger and more comprehensive ontology (Baker, 2017) compared to Propbank. When originally constructed, automated systems could not effectively identify the frames implied by a sentence; today, however, Transformer models (Vaswani et al., 2017; Devlin et al., 2019; Chanin, 2023) have demonstrated success at accurately modeling the sentence-to-frame mapping.

Despite the enormous success and wide adoption of PropBank, FrameNet, and their descendant works, the explicit, finite, and discrete lexicons they employ raise the question: *When is an explicit lexicon ontology complete?*. While FrameNet provides Frame-Frame precedence and subset relationships, these are coarse-grained and do not adequately answer the question: *How can we track the evolution of semantics across a stream of sentences?* - a key requirement for workspace models.

More recent work (Speer et al., 2017) has attempted to assemble a comparatively larger (and less stringent) open-schema semantic ontology of concepts using game-play based crowd-sourcing techniques (von Ahn et al., 2006). However, such efforts to scale manual annotation ultimately do not address how a complete ontology can be constructed. Event Graph Models (EGM) (Mellor, 2017) generate event networks to describe the dynamics of events in a text corpus, often using a combination of submodules such as Coreference Resolution (Ng and Cardie, 2002), Named Entity Recognition (NER) (Li et al., 2020) and Semantic Role Labeling (Palmer et al., 2011). Extensions (Li et al., 2021) generate the *most likely* event *template* sequences. These methods rely on predefined event schema to enumerate the set of possible events. However, while EGMs both track the evolution of semantics across sentences and offer an unsupervised approach to extending existing ontologies, these often marginalize across individual contexts in the training corpus, and generate the most likely event *schema* that follows a current event schema network. As a result, these works have yet to design methods to track the semantics across a specific document.

B Frequently Asked Questions

B.1 Why LLMs for Operator and Reconciler models?

A Workspace instance $\mathcal{W}_n$ is generated as a sample from the Operator f_Φ given an input C_n and the Reconciler g_Ψ compares a pair of Workspace instances $\{\mathcal{W}_n, \mathcal{W}_{n+1}\}$. In this setup, there is no explicit guarantee of stability.

There are two requirements to guarantee stability: (1) *Operator:* Given a pair of text contexts $\{C_i, C_j\} \in \mathcal{X}$ that are similar in semantic meaning but differ in syntax and grammar, the corresponding Workspace instances $\{\mathcal{W}_i, \mathcal{W}_j\}$ must have similar Semantic parts – actors, roles, states, predicates

[3] https://framenet.icsi.berkeley.edu/

and questions – that are inferred. I.e.:

$$f_\Phi(C_i, \mathcal{T}) \sim f_\Phi(C_j, \mathcal{T})$$
$$\forall \{C_i, C_j\} \in \mathcal{X}, \texttt{sem}(C_i) \approx \texttt{sem}(C_j),$$

where $\mathcal{T}$ is the Semantics extraction task (specified via prompting or finetuning); (2) *Reconciler:* Given two workspace instances $\{\mathcal{W}_i, \mathcal{W}_j\}$ that are similar in meaning but use different Identifiers, the Reconciler must be able to infer that these Identifiers refers to very similar Semantics distributions. I.e.:

$$g_{\Psi_S}(\mathcal{W}_i, \mathcal{T}) \sim g_{\Psi_S}(\mathcal{W}_j, \mathcal{T})$$
$$\forall \{\mathcal{W}_i, \mathcal{W}_j\} \in \mathbb{W}_S, \texttt{sem}(\mathcal{W}_i) \approx \texttt{sem}(\mathcal{W}_j).$$

Role of Large Language Models: Prior applications of LLMs – such as T5 (Raffel et al., 2019), XLNet (Yang et al., 2020), MPlug (Li et al., 2022), FST (Chanin, 2023), BERTopic (Grootendorst, 2022) – have effectively identified a distilled set of semantic labels – including topics, sentiments, and frames – across text segments $\ldots, C_i, \ldots, C_j, \ldots$ with similar meaning, despite significant variations in syntax and phrasing. Given its focus on identifying a distilled Semantics maps, GSW similarly employs LLMs (LLaMA specifically) to model the Operator and Reconciler and demonstrates through several experiments, that the instances are stable for short stories.

Summary of Subtasks	
Operator (Generation)	Actor / Role / State Identification, Question and Predicate Generation
Reconciler (Classification)	Node and Edge Reconciliation, Question Resolution

Table 9: **Summary of subtasks addressed by the Operator and Reconciler:** The Operator model generates a Workspace instance $\mathcal{W}_n$ containing roles, states, predicates and questions – by sampling a Conditional Random Field of Semantics – described in Sec. 2 – given an input text segment C_n. The Reconciler compares Workspace instances (a consensus *"Working memory"* $\mathcal{M}^*_n$ instance vs. latest $\mathcal{W}_n$) and outputs an updated consensus instance $\mathcal{M}^*_{n+1}$.

B.2 Why these situations?

The selected situations for our study – *crime and justice, firefighting, technology development, healthcare, and the economy* – were chosen for their frequent coverage in news, diverse roles, states, and predicates and varying durations of events. For instance, events that take place in a *Crime and Justice* situation range from the act of committing a crime all the way up to the delivery of justice, typically unfolding linearly over shorter timescales compared to, for example, the lengthy, cyclical process of *Technology Development*.

Furthermore, each situation can be roughly mapped to an entire set of conceptual gestalts – such as frames and events – that prior lexicon banks have annotated: for example, Crime and Justice can be associated to over 25 frames in FrameNet, 15 for Firefighting, 35 for Technology Development, 50 for Healthcare, and over 60 for the Economy. Frames are manually annotated conceptual gestalts that are identified from large text corpuses (see Sec. A). Therefore, a high coverage across the set of frames indicates that the corresponding Operator and Reconciler models are applicable to a wide range of situation instances. Furthermore, this mapping from situations to frames facilitates a fair comparison between GSW and frame-based Semantic frameworks. A similar correspondence is observed between situations and Event databases.

B.3 How do we extend the number of Situations?

Currently our situations act as predefined (i.e., not dynamically generated by GSW) conditional variables in the CRF model used for modeling Observer-centric semantics. Clearly, an overall awareness of the situation an observer is in influences the roles/states/interactions of the actors. There is, however, sufficient evidence in how humans and other species operate where the awareness of a situation itself can be predicted based on even higher-level cues. Based on these intuitions, we plan to pursue multiple avenues for making the AI observer more of a generalist and also enable it to adaptively create new situation labels based on experiences that do not sufficiently coincide with predictions based on existing situations based models. We articulate some of these approaches in the following.

First, our ablation studies indicate that learning distributions over the Semantics across one situation is easily able to adapt to unfamiliar situations and generate and reconcile Workspace instances that rival the situation-focused model 11. In future research we will use a fewer number of LoRA units, where each unit will focus on multiple situations, thereby merging different situations into a single composite label. Merging can also be conducted at the model level as recent efforts such as TIES (Ya-

dav et al., 2023) have demonstrated. This would result in a hierarchy much along the lines of how GDELT organizes its news articles into news topics and event databases are arranged into subclasses.

B.4 What are the computational requirements for training and inferencing?

LoRA units are augmented to a LLaMA backbone and are finetuned over 10 epochs ($\sim$ 72 hours training per model), batch size of 8, rank 2 (all layers and heads, totalling $1M$ parameters), dropout 0.05, weight scaling-parameter $\alpha = 32$, maximum window length $w = 1024$. We note that the use of LLaMA is freely available for research and requires a custom commercial license. Accordingly, our LoRA adapters + LLaMA models are also available under the same license.

Operator-to-Reconciler inference is performed sequentially (see Fig. 1) and the corresponding LoRA weights can be off-loaded and on-loaded onto the shared LLaMA pretrained backbone to save time cost. The Operator model processes multiple contexts $\ldots, C_i, \ldots, C_j, \ldots$ in batched mode, generating Workspace instances in parallel. Within each instance generation, since the lexicon CRF requires conditional generation, the roles corresponding to each actor are generated prior to generating the corresponding states and so forth.

B.5 How is spatial and temporal information encoded in the Workspace?

Existing lexicon frameworks such as PropBank and FrameNet encode the spatial and temporal information through explicitly defined lexicon Identifiers. In PropBank, verbs are annotated with a *temporal argument* `[ARG-TMP]`, which, when using BertSRL as the Text-to-Semantics model, maps to text spans such as "last week", "tomorrow", "at 10AM". Similarly `[ARG-LOC]` qualifies spans that convey location (local, city, state, country geographical identifiers). In FrameNet, corresponding frames (such as `Temporal Pattern`, `Location in Time`) and frame-specific frame elements (`location/place/region`, `time`) are temporal and spatial information-sensitive lexicon that Text-to-Semantics models such as Frame Semantic Transformers attach to the key text spans.

The Global Semantics Workspace has its own mechanisms to capture spatial and temporal information.

Capturing Spatial Information: At the sentence level, the Operator encodes spatial information as actors (constituting nodes in our complex network representation), often with accompanying roles and states that provide context-specific information.

Consider as an example the sentence in the caption of Fig. 10:

Miller allegedly brandished a weapon at a convenience store on Elm Street, demanding cash before fleeing the scene

the location *"elm street"* is instantiated as a node with roles *escape route* and *crime scene*, succinctly conveying both the crime's location and the suspect fleeing the scene. Similarly, often spatial information is embedded within questions asked by the operator, in Fig. 10 the question *"what evidence did the eyewitnesses provide to the police about the incident at the convenience store?"*, conveys the fact that the location of the crime was a convenience store. Further examples including Fig. 11 - actor *5th avenue*, Fig. 12 - actor *maple street and oak avenue*, Fig. 12 - actor *miami*, Fig. 25 - *stanford university* are presented in Sec. G.

Capturing Temporal Information: As with spatial information, the operator represents explicit temporal details as actors with associated roles and states. Furthermore, implicit temporal information is often encoded in the roles and states of other actors, suggesting the progression or completion of an action. For example, in the caption for Fig. 24:

the lira, hit a new record low against the U.S. dollar earlier in January, breaking 30 to the greenback for the first time. Analysts predict this will be the last hike for some time, especially with local elections approaching in March.

text spans like *"January"*, indicating the time frame of occurance of an event, are captured as Nodes in the network, and, more implicit temporal information, is conveyed through the States accompanying the action - in this case state *approaching* of actor *local elections*. Further examples including Fig. 9 - state *pending* of actor *arraignemnt*, Fig. 12 - actor *february*, Fig. 21 - actor *december* are presented in Sec. G.

At the Reconciler-level, temporal and spatial reasoning is paramount to deciding whether nodes and edges from the latest Operator workspace instance $\mathcal{W}_n$ should replace, augment to, or be rejected in favor of the instance in the Working memory $\mathcal{M}_n^*$. Future work will attempt to highlight these time and space features embedded in the Workspace instances and project the various events and inter-

actor interactions described in the Workspace onto a shared timeline and virtual space.

C Convergence Plots

Convergence loss curves for all situations are reported in Fig. 6.

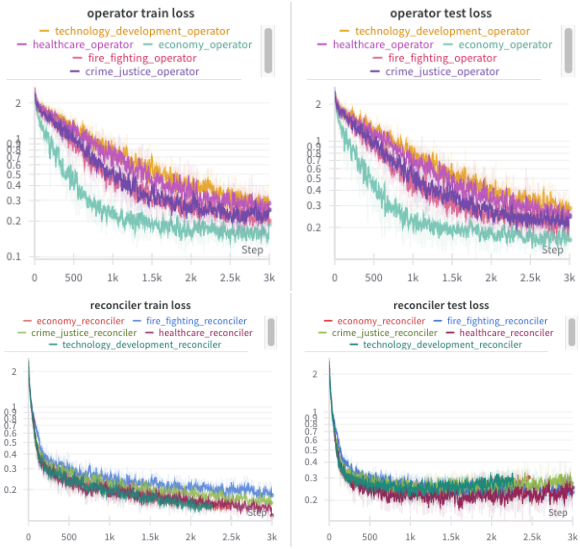

Figure 6: **Operator and Reconciler convergence plots across situations**: Each operator/reconciler model uses shared LLaMA2 (backbone) + QLoRA weights – *plug-and-play setup*. Strong train and test loss convergence implies task learning and good generalizing potential.

D Annotator Guidelines

Annotators who exceeded $\$50K$ in total gross pay were recruited from UpWork, a talent resource. These candidates were first interviewed in a 10-minute session to verify that they were proficient in English and those that had prior experience in annotating large-scale AI/ML data – listed as a verified skill on the platform – were selected to move on to the next round. Following this, they were given a set of 10 task prototype examples and 10 unanswered labeling tasks. Those that got 9 out of the 10 annotations right moved on to the first round of labeling. Each task was labeled twice – by the *annotator* and a *verifier* – to ensure quality of the results. Annotators were paid $\$5/40$ samples which was estimated to take them about 30 minutes at most, or at the rate of $\$10/$hour, which was confirmed to exceed the federal minimum wage where the annotators were situated. The raw annotations are to be released as part of the codebase. Raw annotator guidelines are presented in Figs. 7,8.

Base Model	Adapters	Precision	Parameters
LLaMA-7b	LoRA	16 bit	1M
Mistral-7b	QLoRA	4 bit	800k
LLaMA-13b	**QLoRA**	**4 bit**	**1.6M**
Llama-XXb Chat	-	-	-

Table 10: **Model Architecture Search:** LLaMA-13b trained with 4-bit QLoRA was found to output significantly better *Crime and Justice* situation-specific workspace instances. Finetuned T5 (Raffel et al., 2019), Mistral (Jiang et al., 2023) (7B) models performed worse. LLaMA 70B models performed comparably to the 13B parameter model. LLaMA-Chat models (7B, 13B, 70B) demonstrate poor quality outputs with no further finetuning.

E Ablation

A.1 Generalizing Operator and Reconciler models to other situations: The *Crime and Justice* Operator and Reconciler models are evaluated on text segments associated to *other situations*. We observe very little "situation-bias" in the lexicon (specifically the roles, states and questions) generated by the Operator – example shown in App. Sec. F. Additionally, the *Crime and Justice* Reconciler shows strong classification performance *across all situations*, and even outperforms the situation-specific models in some cases (see Tab. 11, in bold).

A.2 Can LLaMA-based Chat models perform the role of the Operator and the Reconciler without any custom training? In Tab. 10, we show the alternate model architectures and PeFT adapters considered for training the Operator and Reconciler models. We manually evaluated the quality of the outputs of the trained models and the prompted Chat versions, and these performed significantly worse than the architecture we have used in the main text (highlighted in green).

F Situation transfer from one situation to another

The following Workspace instance is generated by the Crime and Justice-trained LoRA model when applied on the following post from a *Fire-Fighting* situation:

After a gender reveal party gone wrong, 10 units of the LAFD were sent to handle the fire at the scene. The gender reveal party caused over a million acres to the engulfed in flames in South Savannah, causing billions of dollars in damages and loss of property.

Situation		Reconciliation			Question Resolution	
Train	Test	ACC.↑	F1↑	Sensitivity↑	ACC.↑	F1↑
crime and justice	*crime and justice*	0.81	0.81	0.91	0.92	0.92
	fire fighting	0.83	0.83	0.95	0.95	0.94
	technology development	0.87	**0.87**	**0.94**	0.99	0.99
	healthcare	0.85	0.85	0.92	0.92	0.92
	economy	0.80	0.81	0.87	**0.99**	**0.99**

Table 11: **Generalization of one Reconciler to other Workspaces**: A Reconciler trained on the *Crime and Justice* situation is shown to generalize across the other situations. Bold indicates cases where classification performance of the *Crime and Justice* trained reconciler is better than the performance of reconcilers trained on each specific situation (compare to Tab. 8).

Nodes(actor, role, states):
"actor": "10 units", "role": "firefighting force", "state": "deployed",
"actor": "flames", "role": "cause of destruction", "state": "engulfing",
"actor": "party", "role": "damage causer", "state": "caused damage",
"actor": "property", "role": "loss victim", "state": "damaged",
"actor": "south savannah", "role": "affected area", "state": "engulfed",

Edges(label, source, target):
"label": "caused loss of", "source": "flames", "target": "property", "attributes": "in south savannah"
"label": "engulfed in", "source": "flames", "target": "south savannah", "attributes": "in south savannah"
"label": "caused loss of", "source": "party", "target": "property", "attributes": "in south savannah"
"label": "caused fire in", "source": "party", "target": "south savannah", "attributes": "south savannah"

Questions:
"who was responsible for causing the fire at the gender reveal party in south savannah?",
"how did the gender reveal party lead to the destruction of over a million acres in south savannah?",
"when and where did the gender reveal party take place?", "where did the 10 units of the lafd go after responding to the fire?",
"what measures are being taken to prevent such incidents in the future?",
"why did the 10 units of the lafd respond to the fire at the scene?"

G Sample Workspace Instances

In the next few pages, we present 4 workspace instances for each situation in Tab. 4 generated by the Workspace (Operator+Reconciler) model at the end of a given passage (mentioned in the caption). These figures can be considered as additional results supporting the one instance shown in Fig. 2.

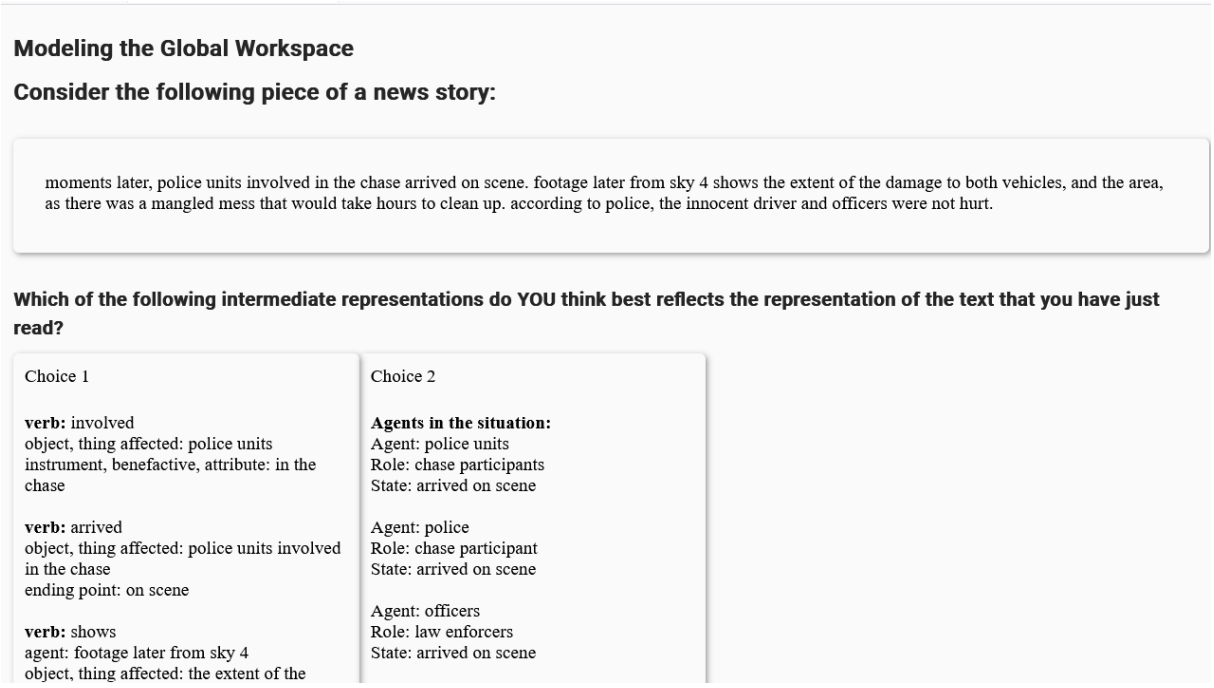

Figure 7: **Annotator instructions for UpWork Task 1:** Annotators are asked to compare the outputs of the Operator to the Semantic map output by a baseline framework (either GLEN, BertSRL, FST) given a shared input text context. During annotation, one random baseline map and the Operator output are presented in random order and the annotator is asked to pick the representation of the Semantics that best reflects the information in the context.

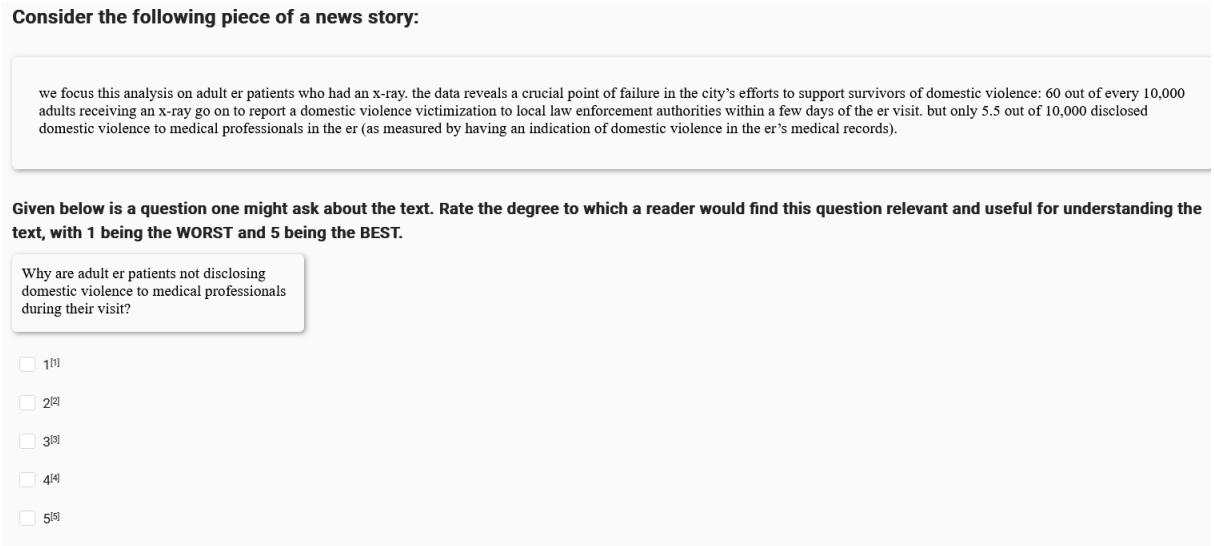

Figure 8: **Annotator instructions for UpWork Task 2:** Given a text segment, annotators are faced with the task of comparing the generated questions from GSW (Operator model) to the questions generated from GPT-4 (annotators are unaware during annotation which question is generated by the GSW model). Annotators find it extremely difficult to identify the source model.

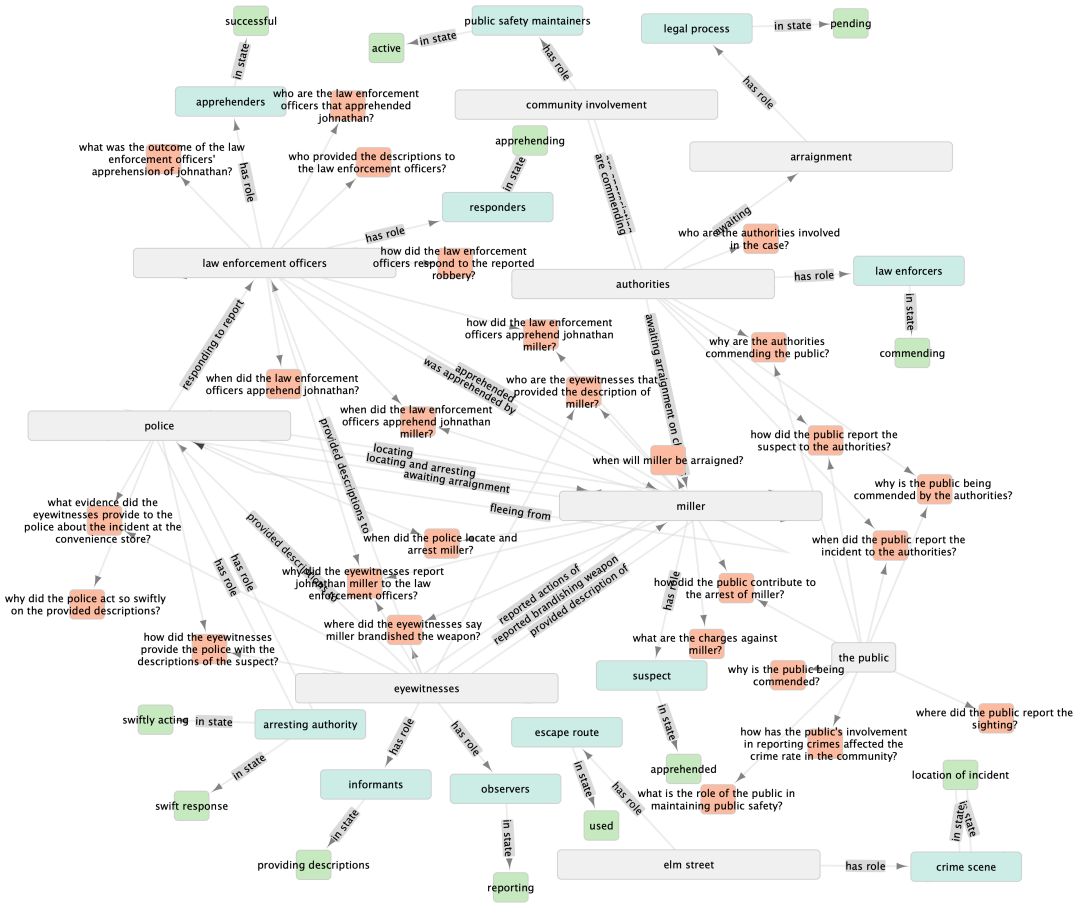

Figure 9: **Crime and Justice -** *Yesterday, in a swift response to a reported robbery, law enforcement officers apprehended Johnathan Miller, a 32-year-old resident of Greenview Avenue, in the downtown area. According to eyewitnesses, Miller allegedly brandished a weapon at a convenience store on Elm Street, demanding cash before fleeing the scene. Police swiftly acted on the provided descriptions, locating and arresting Miller within the hour. He is currently detained at the city's central precinct, awaiting arraignment on charges of armed robbery. Authorities are commending the public for their quick reporting, emphasizing the importance of community involvement in maintaining public safety.*

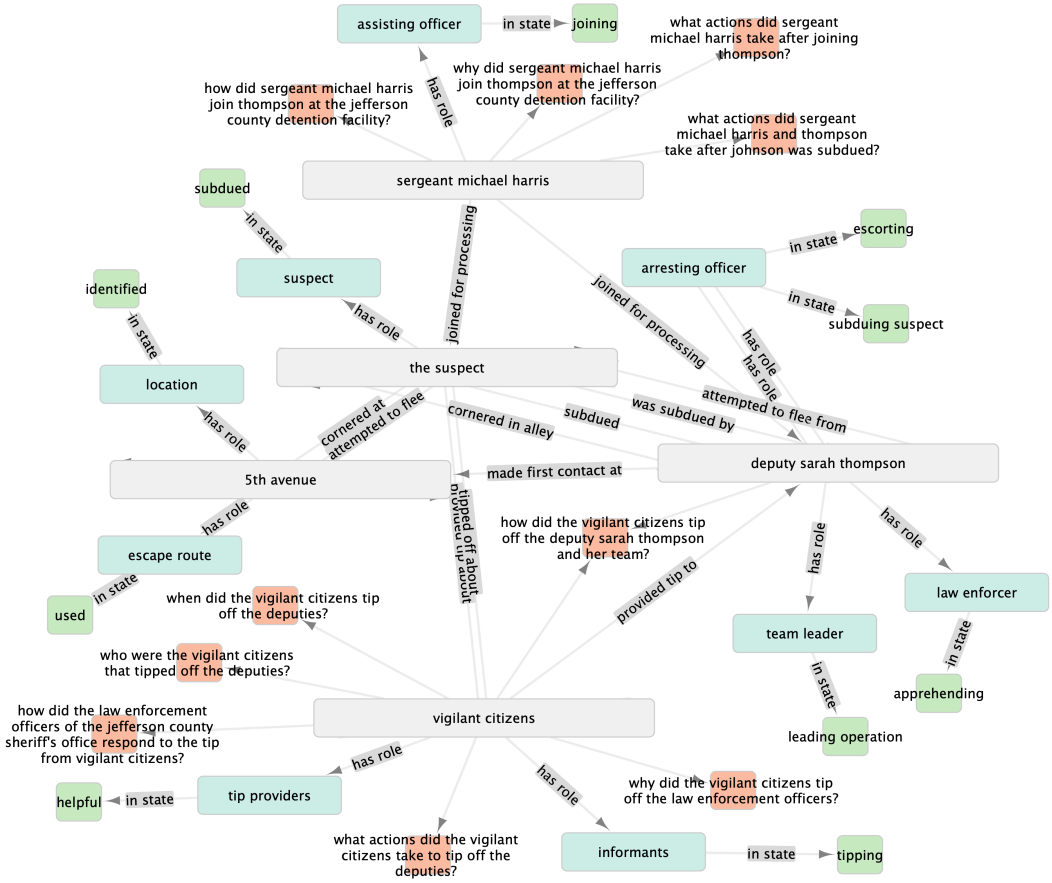

Figure 10: **Crime and Justice -** *In a swift operation conducted by the law enforcement officers of the Jefferson County Sheriff's Office on Tuesday, the notorious fugitive, Mark Johnson, was apprehended in the downtown area of Birmingham, Alabama. Acting on a tip from vigilant citizens, Deputy Sarah Thompson and her team made first contact with the suspect, cornering Johnson in an alley off 5th Avenue South. Johnson was handcuffed as he attempted to flee. After being subdued, Thompson escorted Johnson to the patrol car, in which Sergeant Michael Harris joined him to the Jefferson County Detention Facility for processing.*

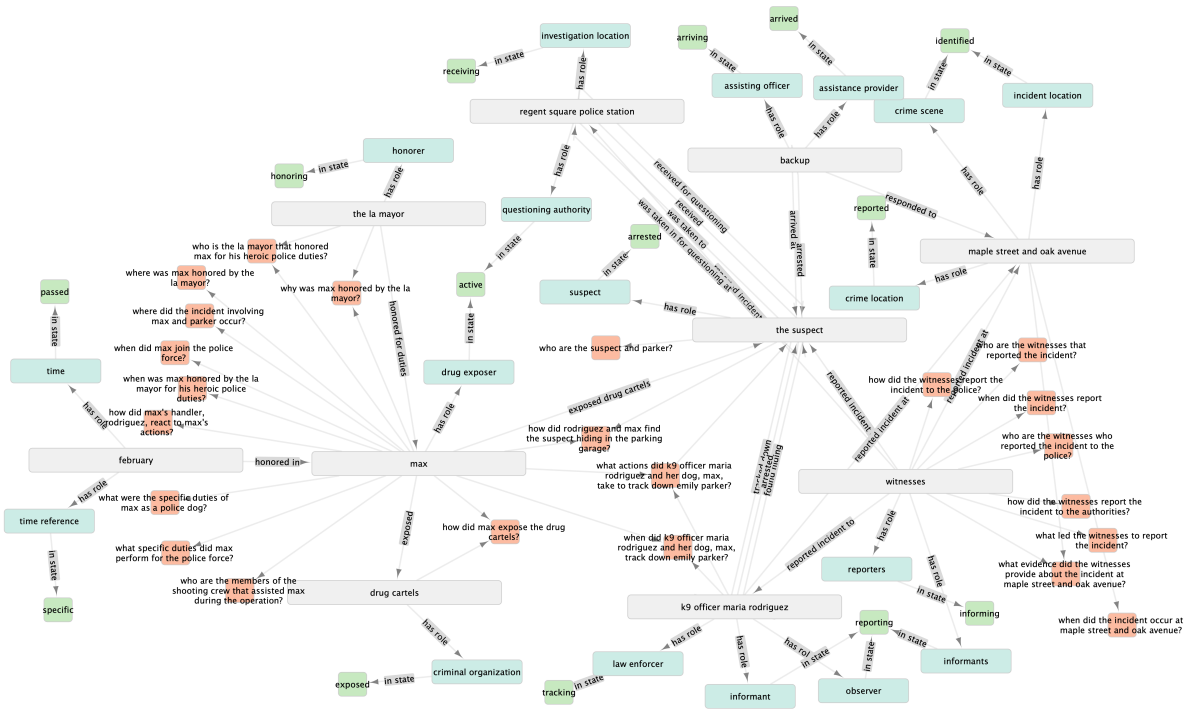

Figure 11: **Crime and Justice -** *In a city hit-and-run case, K9 officer Maria Rodriguez and her dog, Max, tracked down Emily Parker, 28, who fled the scene. Witnesses reported the incident at Maple Street and Oak Avenue. Rodriguez and Max found the suspect hiding in a parking garage. Backup arrived, and Parker was arrested and taken in for questioning at Regent Square police station. Max was honored by the LA mayor for his heroic police duties in exposing drug cartels in February.*

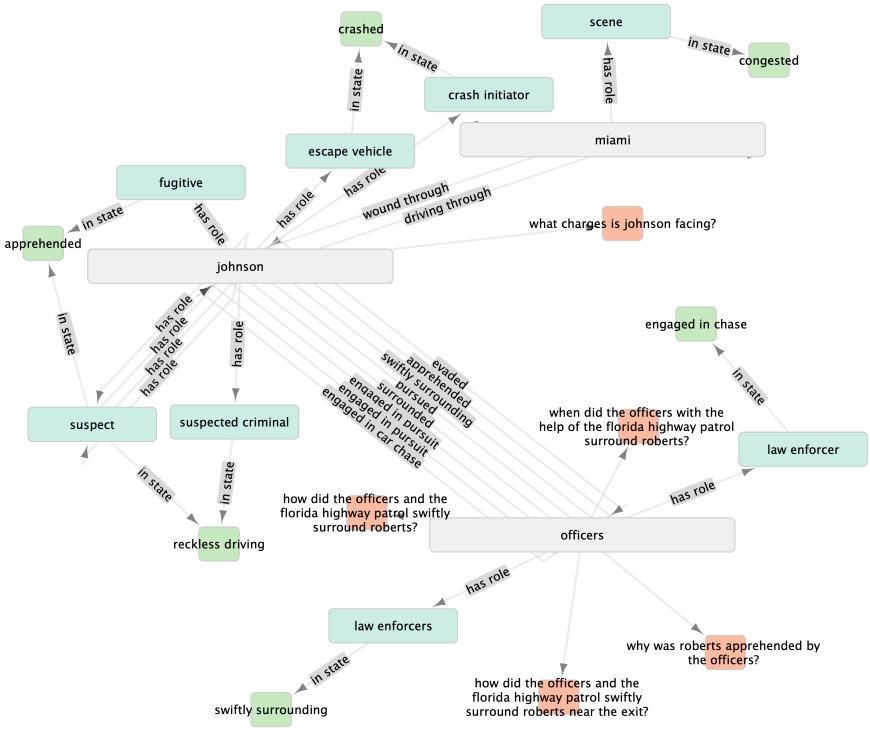

Figure 12: **Crime and Justice -** *During rush hour on Interstate 95 in Miami, officers from the 23rd Precinct engaged in a harrowing car chase with Jason Roberts, 35, suspected of reckless driving and evading law enforcement. The pursuit wound through downtown Miami's crowded streets, with Johnson in a white Benz. Eventually, Roberts's vehicle careened off the interstate near exit 7B, where officers with the help of the Florida Highway Patrol swiftly surrounded him. Roberts was apprehended and taken to the Apalachian court for processing, facing charges of reckless endangerment and fleeing the scene.*

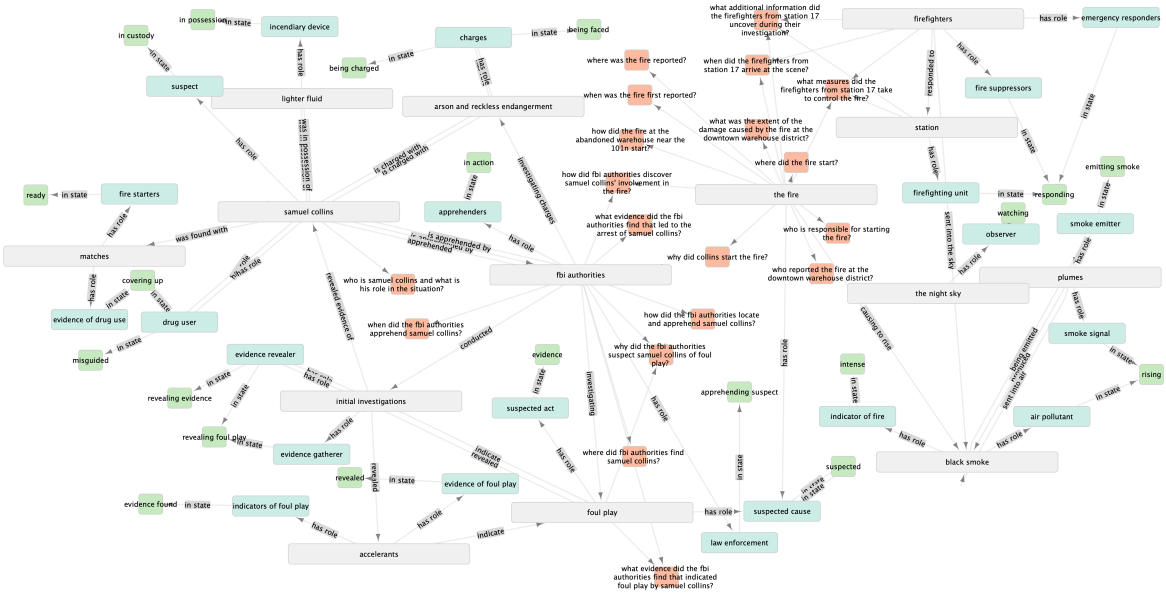

Figure 13: **Fire Fighting -** *In a swift response to a blaze reported at the downtown warehouse district, firefighters from Station 17 battled a ferocious fire that erupted at an abandoned warehouse near the 101N late last night. The fire, suspected to be arson, engulfed the entire structure on 5th Street, sending plumes of black smoke into the night sky. Initial investigations have revealed evidence of accelerants, indicating foul play. FBI authorities apprehended Samuel Collins, 24, discovered under a moving van parked nearby, with matches and lighter fluid in his possession. Collins is believed to have started the fire in a misguided attempt to cover up evidence of drug use, and faces charges of arson and reckless endangerment. The fire is 90% contained.*

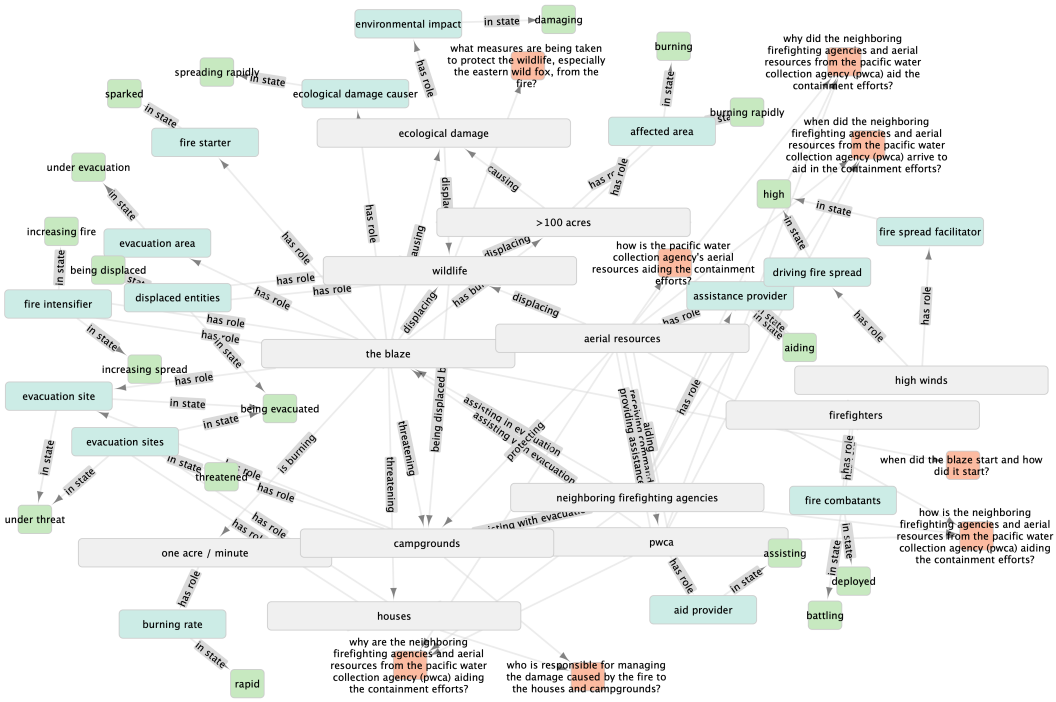

Figure 14: **Fire Fighting -** *Firefighters from the Redwood County Fire Department are battling a forest fire sparked by lightning in the eastern part of the Redwood National Forest. The blaze is spreading rapidly (estimated to burn through one acre / minute) due to high winds and dry conditions. Evacuations are underway at the nearby Cahuenga community, as the fire threatens houses and campgrounds. Despite efforts to contain it, the fire continues to spread (having burned >100 acres), causing ecological damage and displacing wildlife, especially the Eastern wild fox. Assistance from neighboring firefighting agencies and aerial resources from the Pacific Water Collection Agency (PWCA) is aiding containment efforts.*

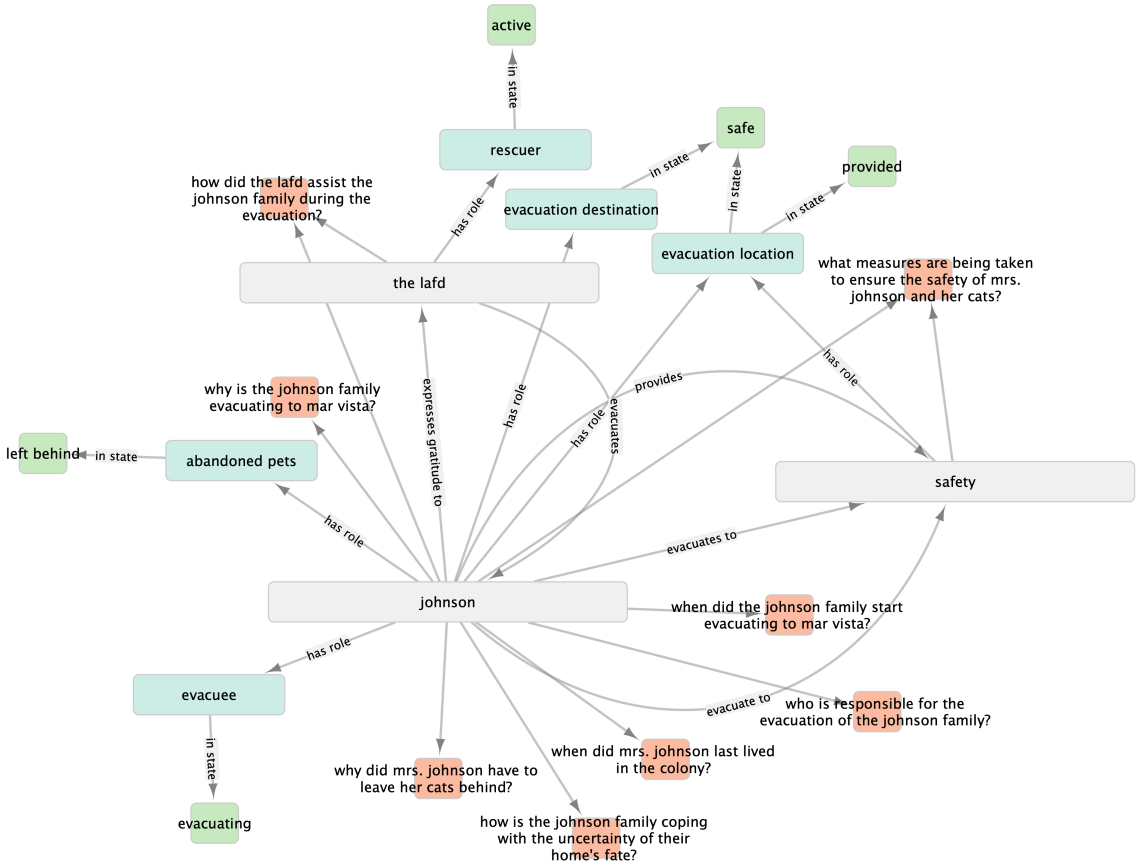

Figure 15: **Fire Fighting -** *In the Redwood National Forest, an electric fire sparked a larger wildfire that now has spread to over 50 acres and threatens over a hundred homes and indigeneous wildlife. Many families including the Johnson family anxiously awaits news of their home's fate as they evacuate to safety to Mar Vista. Mrs. Johnson was visibly distraught: 'I am grateful for the LAFD for evacuating me. I hope the colony remains intact. I have lived there for over 60 years and I had to leave my 3 cats behind.' Respite may take a while however with dry conditions expected from the El Nino weather patterns over the next two weeks.*

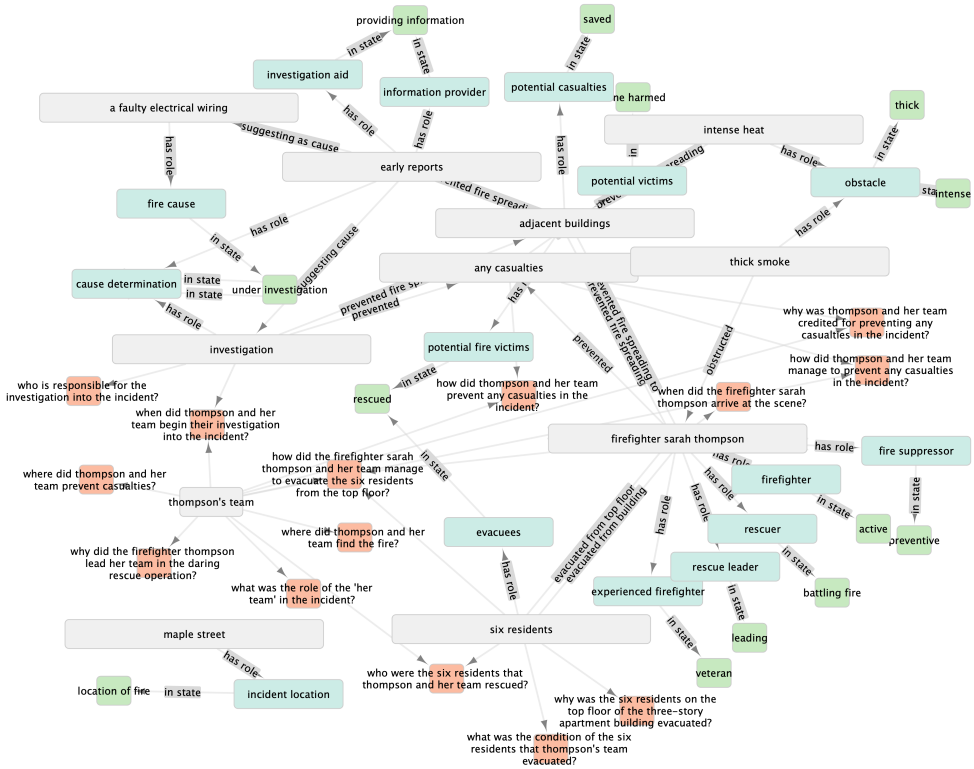

Figure 16: **Fire Fighting -** *In downtown Greenville, firefighter Sarah Thompson battled a fierce blaze that engulfed a three-story apartment building on Maple Street late last night. Thompson, a veteran of the Greenville Fire Department, led her team in a daring rescue operation, evacuating six residents trapped on the top floor. Despite intense heat and thick smoke, Thompson's swift actions prevented the fire from spreading to adjacent buildings. The cause of the fire is under investigation, with early reports suggesting a faulty electrical wiring. Thompson and her team are credited with preventing any casualties in the incident.*

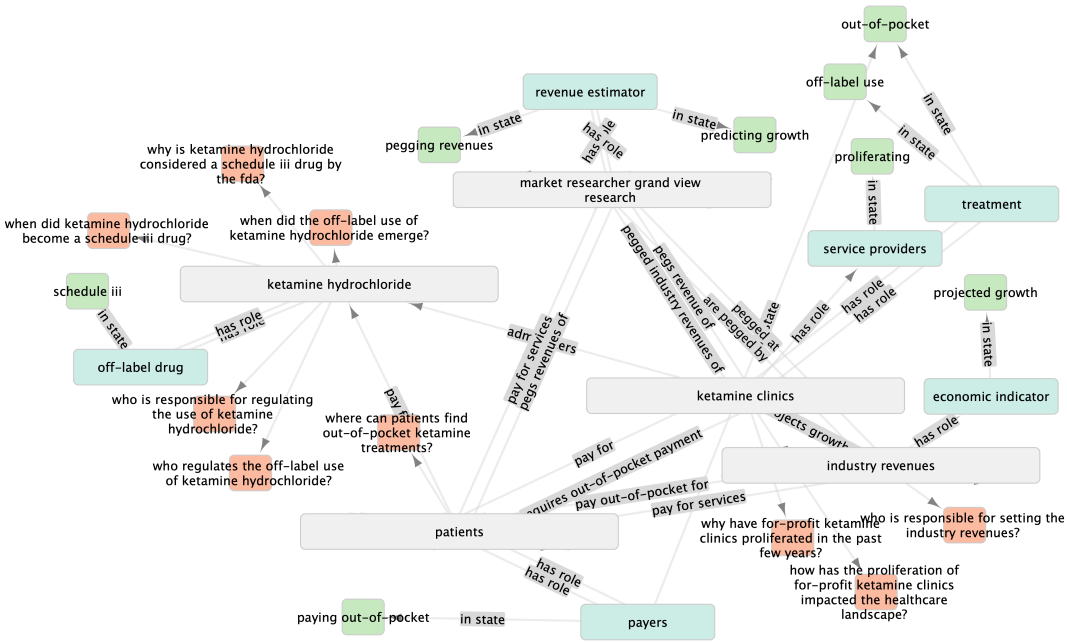

Figure 17: **Healthcare -** *For-profit ketamine clinics have proliferated over the past few years, offering infusions for a wide array of mental health issues, including obsessive-compulsive disorder, depression, and anxiety. Although the off-label use of ketamine hydrochloride, a Schedule III drug approved by the FDA as an anesthetic in 1970, was considered radical just a decade ago, now between 500 and 750 ketamine clinics have cropped up across the nation. Market researcher Grand View Research pegged industry revenues at $3.1 billion in 2022, and projects them to more than double to $6.9 billion by 2030. Most insurance doesn't cover ketamine for mental health, so patients must pay out-of-pocket.*

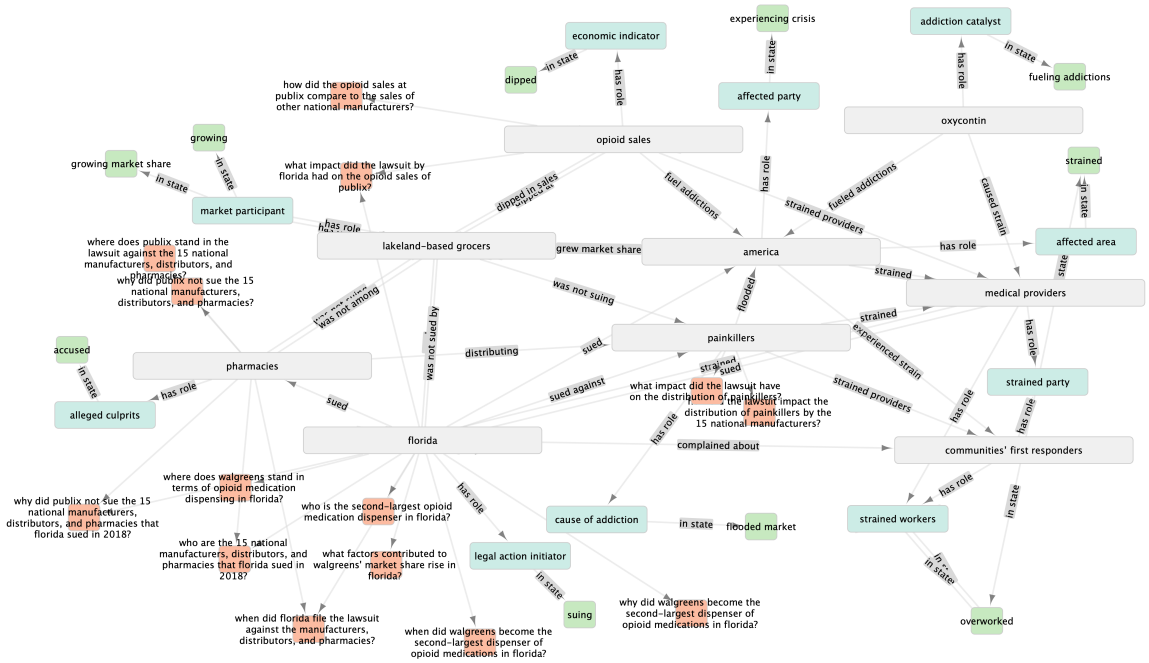

Figure 18: **Healthcare -** *"The Lakeland-based grocer's sales of oxycodone climbed from 26 million pills per year in 2011 to 43.5 million in 2019, the data shows. The increase in sales, which far outpaced the chain's addition of stores in Florida, saw its market share rise to 14%, enough to overtake CVS to become Florida's second-largest dispenser of all opioid medications, behind only Walgreens, which dispensed 28% of opioids in the state in 2019. The analysis excludes drugs like methadone prescribed for addiction treatment. Opioid sales at Publix dipped slightly in 2018 and 2019, the last two years with available data. Even as its market share grew, however, Publix was not among the 15 national manufacturers, distributors, and pharmacies that Florida sued in 2018. That lawsuit claimed other pharmacies had flooded America with painkillers such as OxyContin, fueling debilitating addictions that strained communities' first responders and medical providers.*

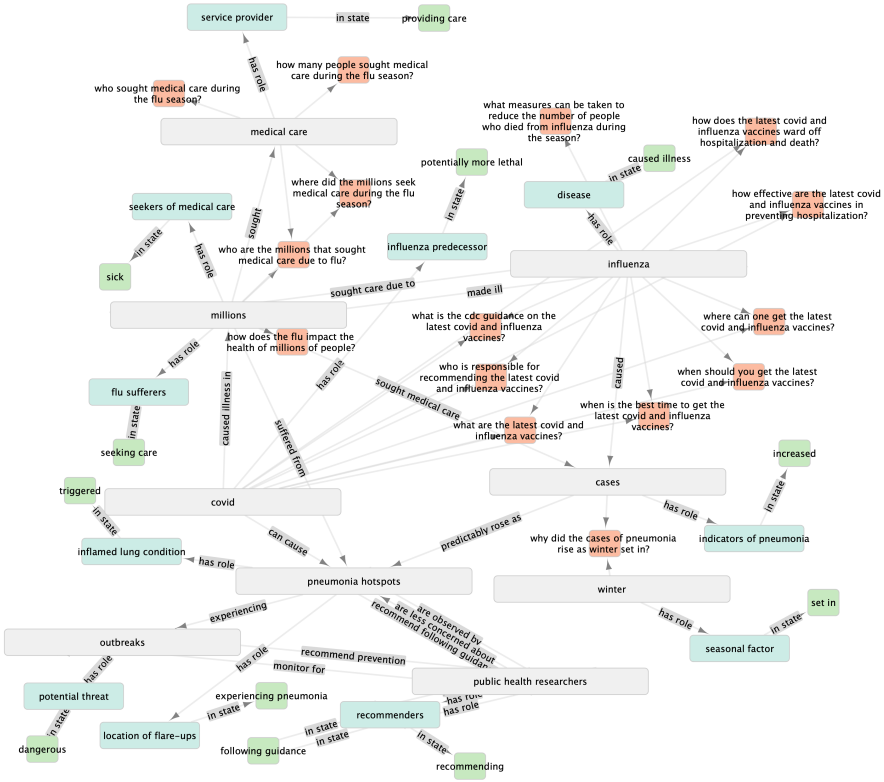

Figure 19: **Healthcare -** *Between 9,400 and 28,000 people died from influenza from Oct. 1 to Jan. 6, estimates the Centers for Disease Control and Prevention, and millions felt so ill from the flu that they sought medical care. Cases of pneumonia — a serious condition marked by inflamed lungs that can be triggered by the flu, covid, or other infections — also predictably rose as winter set in. Researchers are now less concerned about flare-ups of pneumonia in China, Denmark, and France in November and December, because they fit cyclical patterns of the pneumonia-causing bacteria Mycoplasma pneumoniae rather than outbreaks of a dangerous new bug. Public health researchers recommend following the CDC guidance on getting the latest covid and influenza vaccines to ward off hospitalization and death from the diseases and reduce chances of getting sick.*

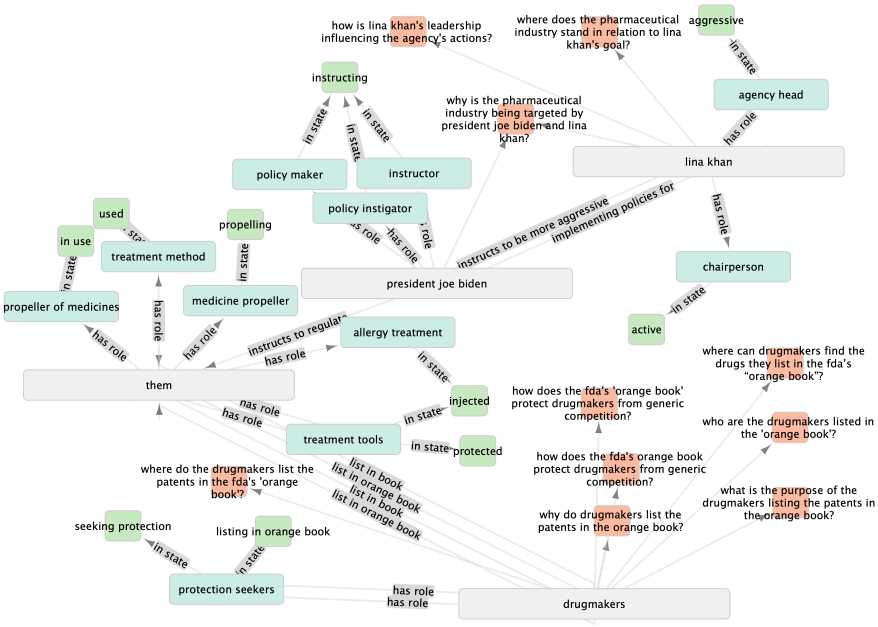

Figure 20: **Healthcare -** *President Joe Biden has instructed his Federal Trade Commission to be more aggressive in reining in the pharmaceutical industry. Under its chairperson, Lina Khan, the agency is aggressively testing the limits of its powers in pursuit of that goal. The targeted patents cover devices that propel medicines for asthma and emphysema into the lungs or inject epinephrine to treat a severe allergic attack. Drugmakers list them in the FDA's "Orange Book," which can afford the products greater protection from generic competition.*

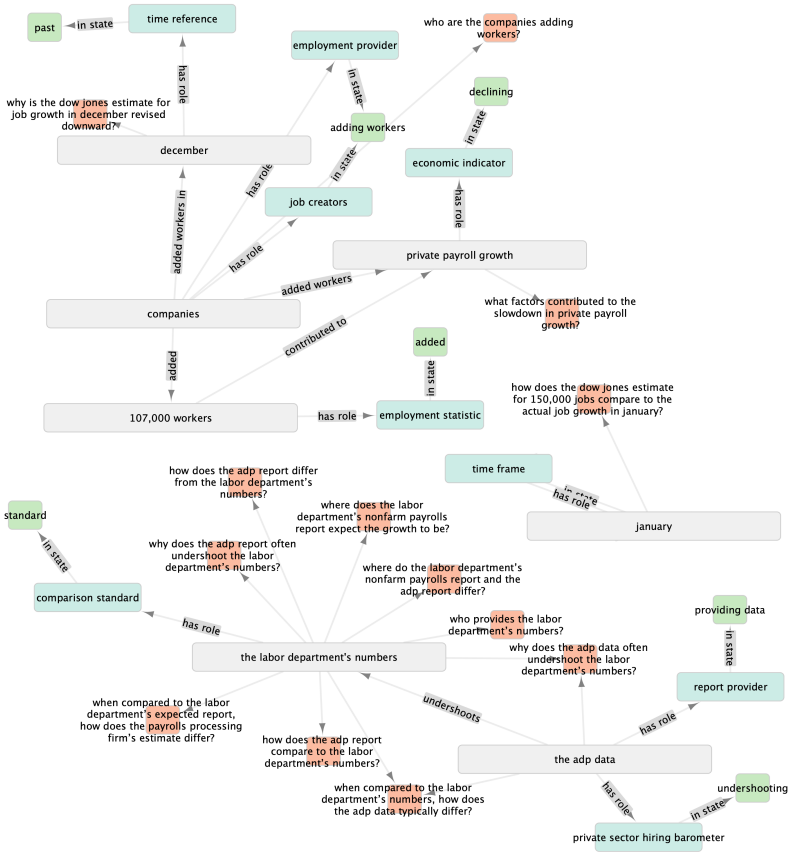

Figure 21: **Economy -** *Private payroll growth declined sharply in January, a possible sign that the U.S. labor market is heading for a slowdown this year. Companies added 107,000 workers in the first month of 2024, off from the downwardly revised 158,000 in December and below the Dow Jones estimate for 150,000, according to the payrolls processing firm. The release comes two days ahead of the Labor Department's nonfarm payrolls report, which is expected to show growth of 185,000, against the 216,000 increase in December. While the ADP data can provide a barometer for private sector hiring, the two reports often differ, with ADP often undershooting the Labor Department's numbers.*

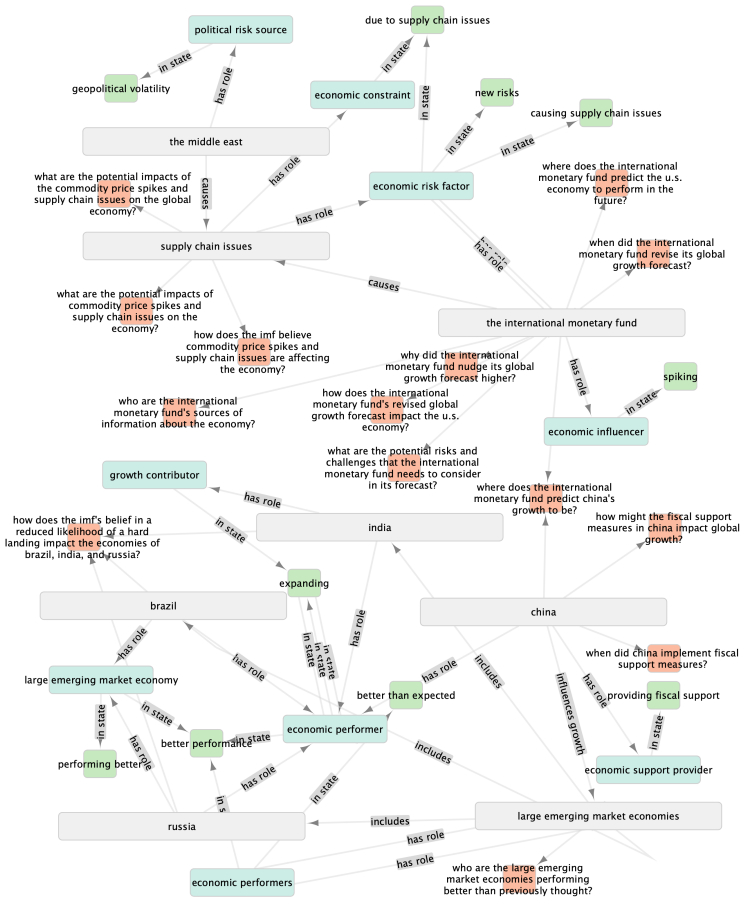

Figure 22: **Economy -** *The International Monetary Fund on Tuesday nudged its global growth forecast higher, citing the unexpected strength of the U.S. economy and fiscal support measures in China. It now sees global growth in 2024 at 3.1%, up 0.2 percentage point from its prior October projection, followed by 3.2% expansion in 2025. Large emerging market economies including Brazil, India and Russia have also performed better than previously thought. The IMF believes there is now a reduced likelihood of a so-called hard landing, an economic contraction following a period of strong growth, despite new risks from commodity price spikes and supply chain issues due to geopolitical volatility in the Middle East.*

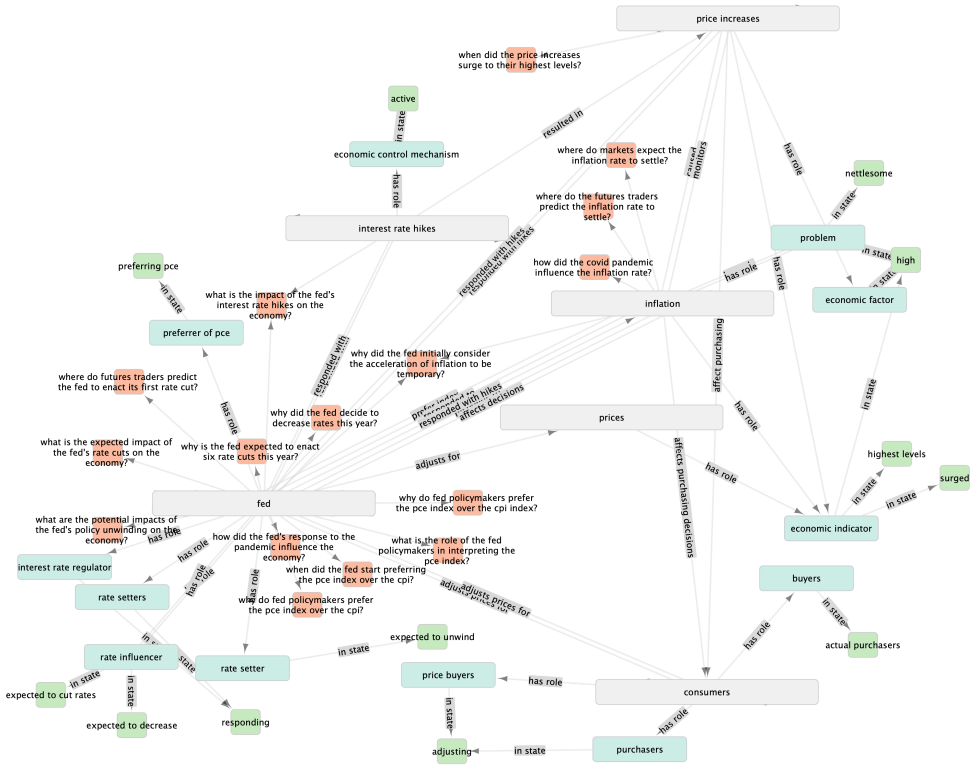

Figure 23: **Economy -** *While the public more closely follows the Labor Department's consumer price index, Fed policymakers prefer the PCE because it adjusts for shifts in what consumers actually buy, while the CPI measures prices in the marketplace. Inflation has been a nettlesome problem since the early days of the Covid pandemic, when price increases surged to their highest levels since the early 1980s. The Fed initially expected the acceleration to be temporary, then responded with a series of interest rate hikes that took its benchmark rate to its highest in more than 22 years. Now, with the inflation rate cooling markets largely expect the Fed to start unwinding its policy tightening. As of Friday morning, futures traders were assigning about a 53% chance the Fed will enact its first rate cut this cycle in March, according to CME Group data. Pricing points to six quarter-percentage point decreases this year.*

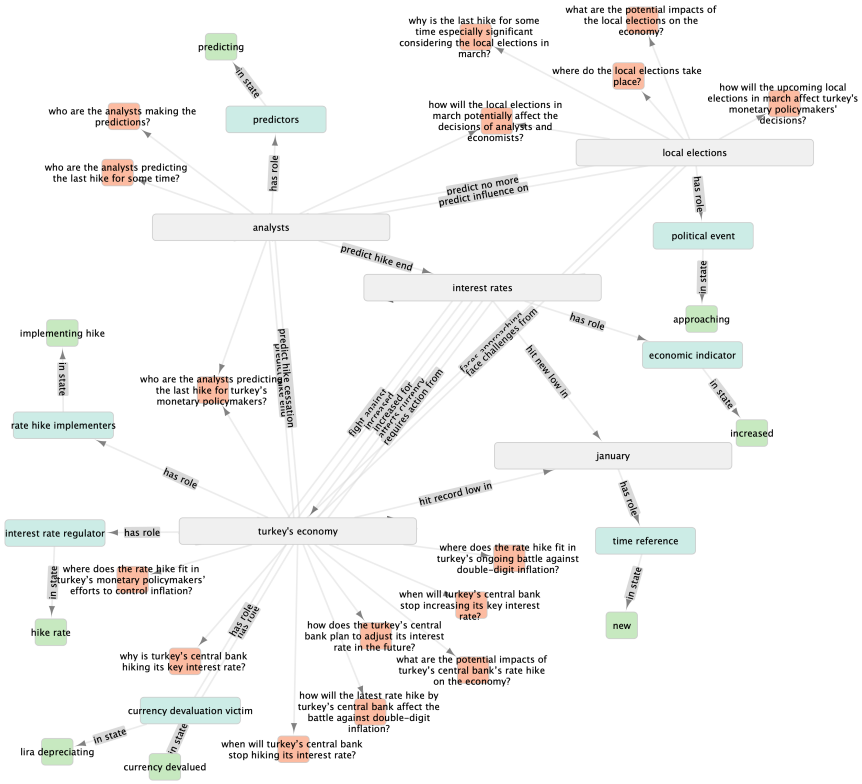

Figure 24: **Economy -** *Turkey's central bank on Thursday hiked its key interest rate by another 250 basis points to 45%. The hike to the benchmark one-week repo rate was in line with economists' expectations. It comes amid an ongoing battle against double-digit inflation for Turkey's monetary policymakers, with the rate hike the latest step in that effort. Inflation in Turkey increased to 64.8% year-on-year in December, up from 62% in November, and the country's currency, the lira, hit a new record low against the U.S. dollar earlier in January, breaking 30 to the greenback for the first time. Analysts predict this will be the last hike for some time, especially with local elections approaching in March.*

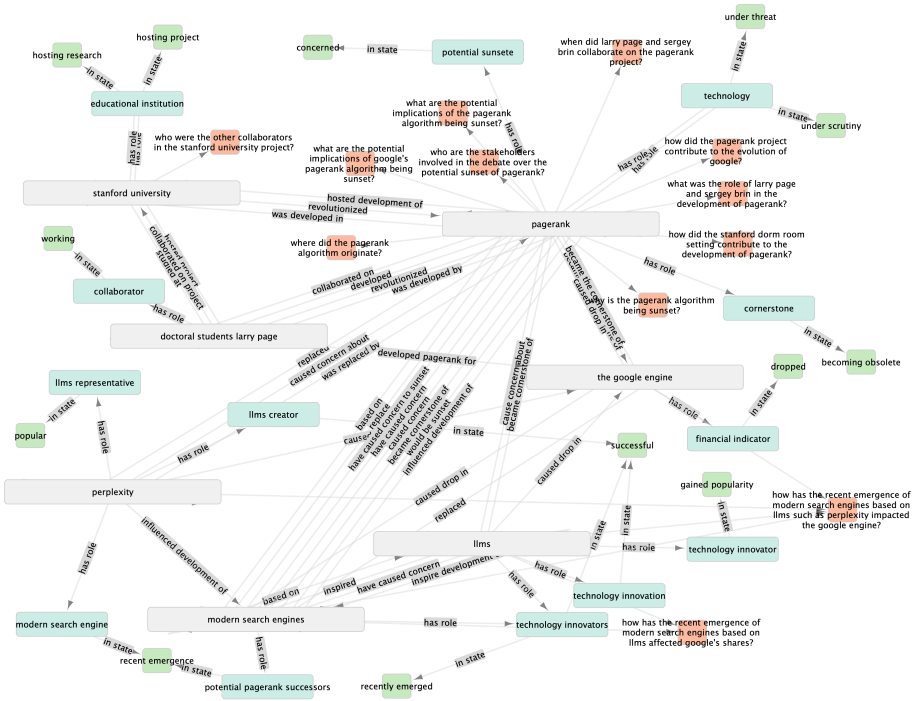

Figure 25: **Technology Development -** *In the late 1990s at Stanford University, doctoral students Larry Page and Sergey Brin collaborated on a groundbreaking project called PageRank to revolutionize internet search. Developed in a Stanford dorm room, PageRank became the cornerstone of the Google engine. While initially successful, the recent emergence of modern search engines based on LLMs such as Perplexity, have caused concern that the PageRank algorithm would be sunset. Google shares have dropped 5% since LLMs gained popularity in early 2020.*

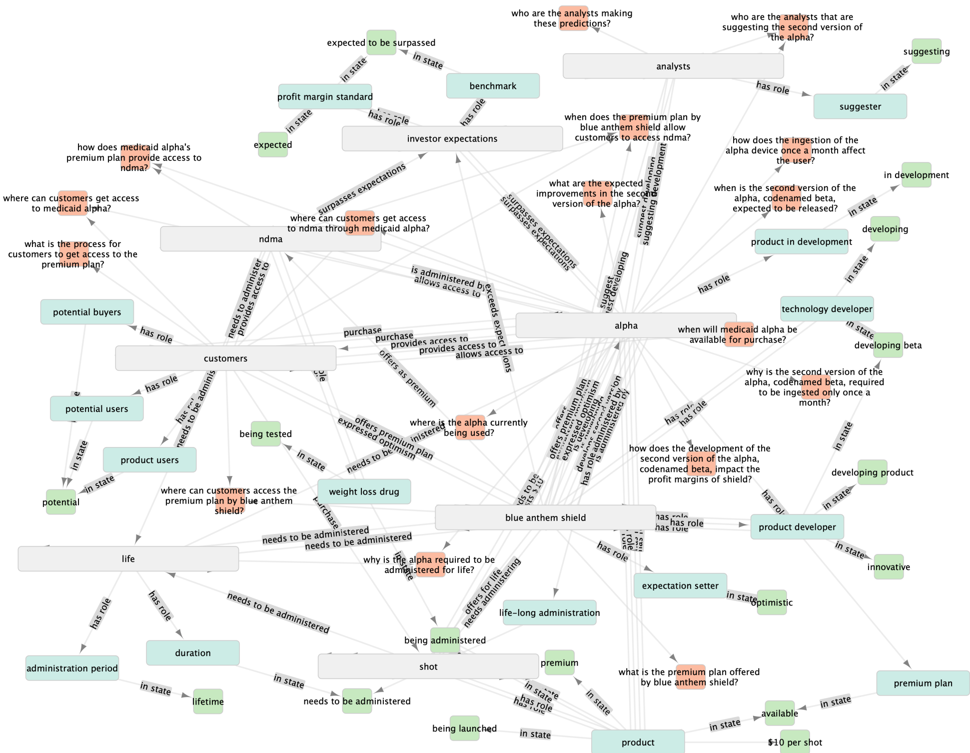

Figure 26: **Technology Development -** *Medicaid Alpha, the premium plan by Blue Anthem Shield, allows customers to get access to NDMA, the cutting-edge weight loss drug taking the US and UK by storm. Alpha costs $10 per shot and needs to be administered for life. Shield CEO Lua Li expressed optimism at the annual Shareholder's meeting, stating that the Q2-Q3 profit margins should surpass investor expectations by over 5%. Analysts suggest Shield is developing a second version of the Alpha, codenamed Beta, that is required to be ingested only once a month.*

Figure 27: **Technology Development -** *Apple released its new iPhone 20, the latest phone model rumored to beat market expectations for demand by late 2034. The latest offering boasts a significantly improved battery life of 25 hours, free text/call service, and an all-glass finish. It is also the first phone to be entirely recyclable. The latest phone will replace the old iPhone 19 at the same price of $999. Analysts expect the phone to sell widely in consumer-driven US/China markets. Yaccarino hinted that the profit margins would be lower on this version, but the demand is expected to exceed that of the 19 by over 20%.*

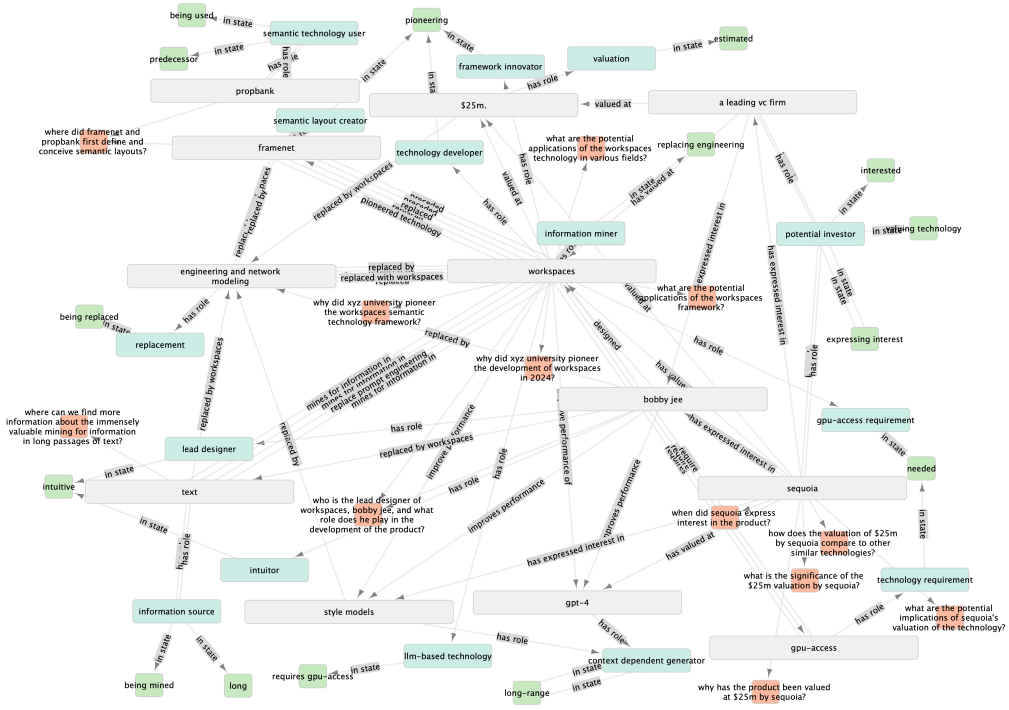

Figure 28: **Technology Development -** *Workspaces, a novel semantic technology framework pioneered at XYZ university in 2024 promises to revolutionize the semantic layouts defined and conceived first by FrameNet and PropBank in the mid-1900s. Workspaces are an LLM-based technology that requires GPU-access to run, and have proved immensely valuable at mining for information in long passages of text - a replacement to prompt engineering and network modeling. Lead designer of Workspaces, Bobby Jee, intuited that the Workspace models may significantly improve the performance of GPT-4 style models by over 30% at long-range context dependent generation tasks. Sequoia, a leading VC firm, has expressed interest in the product, and has valued the technology at $25M.*