Measuring Perceptions of Personhood with Semantic Proto-role Properties

Elizabeth Spaulding Hoefer
University of St. Thomas
elizabeth.hoefer@stthomas.edu

James Martin
University of Colorado Boulder
james.martin@colorado.edu

Abstract

We show that semantic proto-role properties can be used as a tool to measure implicit human perceptions of *agency* and *patiency* of entities in human-generated text. First, we demonstrate that silver-generated semantic proto-role property labels are strongly correlated with both human judgment and a probabilistic text-based measure of anthropomorphism. Then, we use our measure to quantify linguistic idiosyncrasies across different AI-related Reddit communities. Our measure shows that subreddits dedicated to discussing AI companionship ascribe higher sentience to "bots" and higher agency to "companies" when compared to other subreddits. This phenomenon reveals not only the unique way in which chatbots are anthropomorphized in such subreddits, but also the users' keen awareness of their power imbalance with the companies that created the chatbots.

1 Introduction

It has long been recognized that the semantics of ordinary human language reflects the attitudes and biases underlying the human culture that produced it. For instance, metaphors are not only descriptive, but can change the way people think by shaping our understanding of novel concepts (Lakoff and Johnson, 2008; Lakoff and Turner, 2009; Landau et al., 2010). Thus, the use of metaphor in human speech or text can reveal implicit attitudes of the speaker (Ruscher, 2017).

Research in computational semantics, corpus linguistics, and machine learning has shown that machines trained on text corpora adopt the implicit human attitudes contained within them. For instance, Caliskan et al. (2017) uncover a wide spectrum of human-like biases in word embeddings using Implicit Association Tests (Greenwald et al., 1998), such as associations of *insects* with *unpleasantness*, or stronger associations of female names with family words than with career words compared to male names. Similarly, Ash et al. (2023) use word embeddings to quantify human perceptions of *agency* and *patiency* of 255 entities, and track changing perceptions of women and animals over several decades of historical text. Cheng et al. (2024) quantify the anthropomorphism of AI entities in text with a machine learning metric called AnthroScore, finding that news headlines anthropomorphize AI more than research papers.

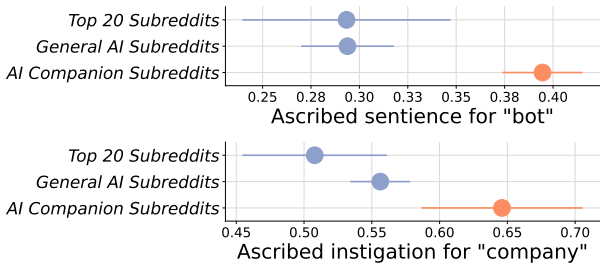


Figure 1: Subreddits that discuss AI companionship ascribe higher *sentience* to *bots*, and higher *instigation* to *companies*, compared to other subreddits.

We show that semantic proto-role labeling (SPRL), a task which classifies and characterizes the predicates (events) and arguments (entities) of a sentence, can also be used to measure how much the language in a corpus anthropomorphizes entities. Furthermore, we show that SPRL can quantify and separate differing dimensions of personhood. As a case study, we apply our method to study the perception of AI-related entities (such as chatbots and LLMs) across different Reddit communities. We find that, compared to other communities, communities that discuss the usage of AI companions ascribe higher levels of sentience to entities such as *bot*, but lower levels of sentience to entities such as *LLM* and *GPT*, suggesting a layered conceptualization of AI. We also find that AI companion users ascribe higher agency to *companies*, which reflects their concern with the level of control the companies hold over their companions.

Replication package available here.

2 Related work

2.1 Defining personhood

Ash et al. (2023) describe a two-dimensional view of personhood comprised of mental properties of *agency* (the ability to act, plan, and decide) and *experience/patiency* (the ability to feel, hunger, and desire; Baker 2000; Gray et al. 2007), which materializes in law, and motivates much of the way we reason about morality (Freeland, 1985; Shaver, 1985; Weiner, 1995). Using a different model of mind perception, Cheng et al. (2024) define anthropomorphism by citing a three-component model of agency itself (the "ABC model"), which studies the role of agency in dehumanizing metaphors: "Full agents possess the ability to (1) experience emotion and feel pain (affective mental states), (2) act and produce an effect on their environment (behavioral potential), and (3) think and hold beliefs (cognitive mental states)." (Tipler and Ruscher, 2014) These are the "human-like qualities" that can be ascribed to non-human entities, and can be measured by AnthroScore.

Anthropomorphizing AI Personifying or anthropomorphizing inanimate objects—that is, attributing human characteristics to nonhuman things, or viewing a nonhuman object through a social lens—is a normal human tendency (Zimmerman et al., 2024). Today, people increasingly perceive AI as more warm, competent, and human-like (Cheng et al., 2026). An individual's beliefs and perceptions of AI technology influence their willingness to use it (Kelly et al., 2023), and, in turn, increased usage and familiarity with chatbots is associated with an increased likelihood to attribute some form of phenomenal consciousness to the chatbot (Colombatto and Fleming, 2024), and users who perceive chatbots as more human-like and conscious are more likely to report positive opinions of the chatbot (Guingrich and Graziano, 2025). However, anthropomorphizing AI carries significant risks, such as encouraging overreliance on system outputs (Abercrombie et al., 2023; Chiesurin et al., 2023), including overly cautious, sycophantic, or incorrect health advice in therapeutic contexts (Grabb et al., 2024; Jung et al., 2025), and encouraging emotional relationships with AI systems. Some argue that such relationships can detract from human relationships, and can be exploited by the companies that control the systems (Zimmerman et al., 2024); one study analyzed several thousand chatbot conversations of users of the AI companion app Replika and identified six harms: relational transgression, harassment, verbal abuse, self-harm, misinformation, and privacy violations (Zhang et al., 2025). Other research has found that users of Replika can project gender stereotypes onto the bot (Depounti et al., 2023), which mirrors and affirms such stereotypes via sycophantic responses (Lima and Belk, 2025).

AnthroScore Researchers are increasingly interested in quantifying the anthropomorphism of AI because of the potential risks and harms. As such, Cheng et al. (2024) develop a language measure to quantify anthropomorphism in text using RoBERTa (Liu et al., 2020). For a masked entity e, AnthroScore is the log of the ratio of the probability that a human pronoun replaces the mask, over the probability that a non-human pronoun replaces the mask. So, a higher AnthroScore means greater anthromorphism of the masked entity e in the sentence: $Anth(e) = \log \frac{P_{HUMAN}(e)}{P_{NON-HUMAN}(e)}$. More details on AnthroScore can be found in Appendix B.

Human judgement survey Earlier work has sought to characterize people's understanding of personhood or *mind* of various entities (Gray et al., 2007; Gray and Wegner, 2009). We utilize the data from Ash et al. (2023), who collected human judgements ($N = 3,181$ respondents) on 255 common entities in the English language. The researchers collected implicit judgements on the agency and experience of these entities (see survey questions in Table 4). They then average the responses in each category into two scores for agency $a_H(e)$ and patiency/experience $p_H(e)$, with $e \in E$ representing each entity and H for "human survey." We utilize this data as it is the largest validated set of entities with human judgements of agency and experience to our knowledge.

2.2 Semantic proto-roles

Because *agency* and *experience/patiency* define personhood, and thus, anthropomorphism, we turn to a linguistic theory of agency and patiency to help us quantify humanizing and anthropomorphizing language. Entities within sentences can be classified as more agent-like or more patient-like[1] on the basis of a set of properties, introduced by Dowty (1991): the properties of a prototypical agent ("proto-agent") are *sentience*, *awareness* of the event described by the verb, *independent exis-*

[1]In semantics, a "patient" is the entity that experiences the effects of the action in the sentence.

tence of the event, *volition* with respect to the event, and *instigation* of the event. Semantic proto-role labeling (SPRL) is the computational task which assigns these properties to entities within sentence (Reisinger et al. 2015; White et al. 2016; Rudinger et al. 2018; see Appendix A for further explanation of this task). For example, in the sentence "The cat sat on the mat", a SPRL parser would label *the cat* with:

{ +volition, +sentience, +awareness, +independent existence, +instigation }

while *the mat* would be labeled with:

{ –volition, –sentience, –awareness, +independent existence, –instigation }.

Thus, by Dowty's theory, *the cat*, which exhibits more proto-agent properties, is the agent of the sentence, while *the mat* is the patient.

3 Experiments

We address the following questions:

(RQ1) Can we learn about human perceptions of personhood of entities of interest by performing semantic proto-role labeling on text?

  (a) Can semantic proto-role properties quantify differing dimensions of personhood in text?

(RQ2) Do users of AI chatbot companions perceive the agency and experience of chatbots differently than users of AI chatbots for other reasons?

To answer RQ1, we validate our SPRL measures on both human judgement from Ash et al. (2023), and the probabilistic, text-based anthropomorphism measure from Cheng et al. (2024). To answer RQ2, we analyze the text of user groups taken from Reddit using our validated SPRL measure.

3.1 Data for validating SPRL measures

COHA We utilize the Corpus of Historical American English (COHA; Davies, 2021) for best comparison to Ash et al. (2023). COHA is a large corpus consisting of English language usage across several decades, starting in the 1800s. The sample we used consists of 3.6 million words and over 200,000 sentences. After filtering for noise, length, and the presence of entities with human judgements, 47,301 sentences remained, over which 18,781 predicate-argument pairs were identified by the SPRL system. Further filtering details are in Appendix C.1.

For the analysis on COHA, we utilize the same 200+ entities chosen by Ash et al., based on previous work in social perceptions of agency and patiency. Because many of the entities in the full list are infrequently mentioned in COHA, we additionally separate entities into subsets of different sizes $E_{N \geq n}$. Membership in $E_{N \geq n}$ is determined by the number of silver[2] SPRL mentions across the COHA sample n, so that the subset $E_{N \geq 100}$ contains all entities which have at least 100 mentions in the silver-labeled data. Additionally, Ash et al. stipulate a pre-registered subset of entities designed to be representative of different agency and experience concepts, which we call E_{pre}. We evaluate on this subset as an additional validation of our measure, but our main analysis is over $E_{N \geq 100}$ and $E_{N \geq 5}$, which contain common entities such as *woman*, *girl*, *god*, *army*, *dad*, *fish*, etc. The complete list of 200+ entities with human judgement scores, along with our specific subsets, can be found in Table 11.

ACL Abstracts As in Cheng et al. (2024), we analyze abstracts taken from the ACL Anthology[3], which contains >300,000 computational linguistics papers. We parse all ACL abstracts and manually review the identified arguments, sorted by number of mentions throughout the corpus. We select 88 frequent entities, ensuring we cover all entities covered in Cheng et al. (2024), and then filter sentences as in COHA, over our selected entities, and end up with 131,349 sentences which contain 224,791 mentions with silver SPRL. As in the COHA analysis, we separate entities into subsets $E_{N \geq n}$, where entities have at least n silver SPR labels. We show entity lists in Table 12.

3.1.1 Reddit data

Reddit, a massive online discussion forum with thousands of different communities ("subreddits"), is often used as a source of observational data in social science and digital humanities research (Proferes et al., 2021). Each subreddit has its own culture and set of norms, and the ability to remain semi-anonymous on Reddit means that users can discuss controversial or stigmatizing topics relatively freely. Because we are interested in the linguistic signals in the discussions among users of AI chatbot companions—a topic that is considered

[2]"Silver-labeled" data is data that has been automatically annotated by a trained model rather than manually annotated by a human (which we refer to as "gold-labeled" data).

[3]https://aclanthology.org/

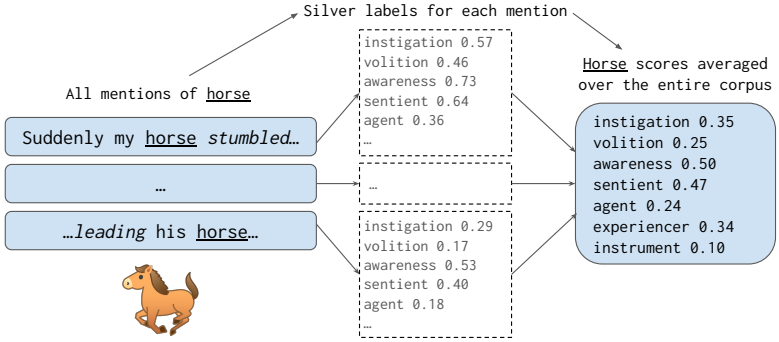


Figure 2: The process for scoring an entity *horse*.

stigmatizing (Dubé et al., 2023)—we chose Reddit as our data source to investigate RQ2.

We scraped text from 38 different subreddits which we classified under three umbrellas to capture three different user groups: (1) subreddits for individuals who use AI chatbot companions; (2) subreddits for general AI users, i.e. individuals who use chatbots for miscellaneous, but mainly productivity-related reasons; and (3) the top 20 most popular subreddits, for a more general population of people who may or may not use chatbots at all. After filtering for our entities and for length, we obtained 17,184 sentences for the AI companion category, 63,331 sentences for the general AI category, and 11,469 sentences for the top 20 category. See Table 9 for a full list of the subreddits.

Ethical considerations This study was approved by our institution's institutional review board under exempt status due to Reddit's public nature, but because of the sensitive nature of some of the topics discussed in our data, we take extra steps to protect the privacy of the users. Because it is relatively easy to trace a particular Reddit user to an exact quotation using a search engine, we follow previously established guidelines in the field and do not release raw data or include direct quotations from any Reddit users, and instead use bricolage-style paraphrases to inhibit traceability (Markham, 2012; Sanches et al., 2019).

3.2 Methodology

First, to establish a correlation between the H measure (the survey judgments from Ash et al.) and semantic proto-role labels, we analyze entities in text using both gold (annotated) and silver (system-produced) semantic proto-role labels. To produce silver SPR labels, we reproduce the full semantic proto-role labeling system from Spaulding et al. (2023). Our system achieves a micro-F1 of 83.2 on SPR properties. Details on our system, including individual F1s on different properties and components, can be found in Appendix A.

AGENT	EXPERIENCER	INSTRUMENT
+instigation	-instigation	+instigation
+volition	-volition	-volition
+awareness	+awareness	-awareness
+sentient	+sentient	-sentient
+change loc		+change loc
+ind. existence		

Table 1: All SPRL categories used for our measure, with binary properties for each one.

3.2.1 SPRL Measures

We configure a number of SPRL categories based off of Dowty's "cluster-concepts" for the thematic roles of Agent, Experiencer, and Instrument (Dowty, 1991, p. 578). Table 1 shows these categories. In addition to these categories, we also score entities on four binary semantic proto-role properties alone: instigation, volition, awareness and sentience (see Appendix A.1 for a full list).

For the SPR labels alone, we score entities by averaging the *probability output* from the SPRL parser's logits. For the three cluster-concepts from Table 1, we score in two different ways. First, we score an argument for a cluster by summing the raw probability output for each of the *positive* constituent properties if and only if the probability output is greater than 0.5. In other words, for an entity mention e with SPRL probabilities p_s, we score the EXPERIENCER cluster as follows:

$$\begin{aligned}\text{EXPERIENCER}(e) = p_{\text{aware}} * [\![p_{\text{aware}} > 0.5]\!] \\ + p_{\text{sentient}} * [\![p_{\text{sentient}} > 0.5]\!]\end{aligned}$$

Because this method does not take into account the negative properties in each cluster, we score by awarding a 1 for each constituent property that matches exactly, so for an entity mention e with SPRL probabilities p_s, we generate the strict score for the INSTRUMENT cluster as follows:

$$\begin{aligned}\text{STRICT INSTRUMENT}(e) = 1 * [\![p_{\text{instigation}} > 0.5]\!] \\ +1 * [\![p_{\text{volition}} \leq 0.5]\!] + 1 * [\![p_{\text{aware}} \leq 0.5]\!] \\ +1 * [\![p_{\text{sentient}} \leq 0.5]\!] + 1 * [\![p_{\text{chg loc}} > 0.5]\!]\end{aligned}$$

For each SPRL measure, and each entity, we produce a score for that entity, as shown in Figure 2 for the entity *horse*. To validate our SPRL measures, we gather entity scores across the COHA sample and calculate both Pearson's and Spearman's correlation coefficients between each H measure and each SPRL measure. We do the same thing across the ACL abstracts and AnthroScore. We also calculate and report the associated p-value, and consider our results statistically significant at $p < 0.01$.

4 Results and discussion

4.1 RQ1: SPRL measures correlate with human judgements

We find that our SPRL measures correlate with the human judgements on the entities collected in Ash et al. (2023). We compute our scores across a number of different subsets of entities and two datasets to check for robustness.

SPR1&2 and H Measures First, we run an initial pilot analysis and aggregate the measures across the *gold* SPRL annotations, two datasets called SPR1 and SPR2 (see Appendix A). This verification step checks that the human judgements of the SPRL annotations (i.e., the data that the SPRL parser was trained on) align with the human judgements from the survey data. We completed this step before generating the silver labels to verify that the two sources of human judgements were comparable. All ρ coefficients calculated on E_{gold} in SPR1 and SPR2 can be found in Table 5 in Appendix C.2. While the subset of entities we had enough data to measure is small, with $|E_{gold}| = 17$, we found that the gold SPRL aggregates correlated highly with the human survey scores: Our AGENT measure had Pearson's $\rho = 0.78$ with $a_H(E_{gold})$ (the human survey aggregate for the agency dimension of personhood), and our EXPERIENCER measure had $\rho = 0.814$ with $p_H(E_{gold})$, with $p < 0.01$.

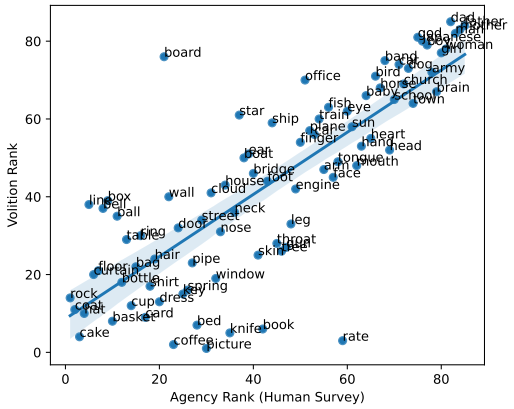
H Measure	SPRL Measure	Pearson's ρ	Pearson's p	Spearman's ρ	Spearman's p
agency (a_H)	AGENT	0.922	0.000	0.823	0.000
	instigation	0.923	0.000	0.852	0.000
	volition	0.933	0.000	0.821	0.000
	awareness	0.929	0.000	0.837	0.000
	sentient	0.928	0.000	0.844	0.000
	change of location	0.269	0.102	0.296	0.071
	independent existence	0.756	0.000	0.796	0.000
patiency (p_H)	EXPERIENCER	0.901	0.000	0.733	0.000
	awareness	0.901	0.000	0.745	0.000
	sentient	0.909	0.000	0.763	0.000
$a_H - p_H$	AGENT - EXPERIENCER	0.710	0.000	0.494	0.001

Table 2: ρ coefficients for a_H, p_H, and d_H and SPRL measures over the subset $E_{N\geq100}$ wherein each entity has at least 100 mentions. ($|E_{N\geq100}| =$ 38 entities total). Text in CAPS denotes clusters from Table 1.

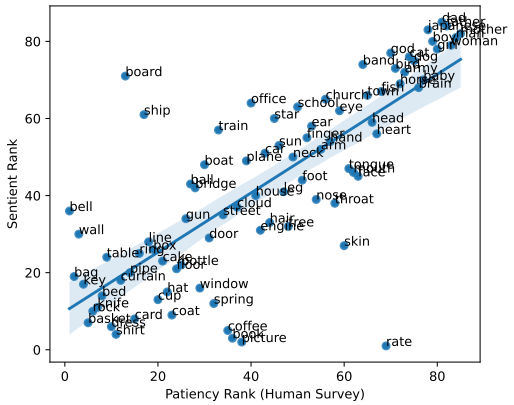
COHA and H Measures On the pre-registered list of entities E_{pre} from Ash et al. (2023), we report a Pearson's ρ of 0.77 for AGENT and the human ranking of agency (a_H). EXPERIENCER is correlated with the human ranking of patiency (p_H) with a Pearson's ρ of 0.74. However, many of the 28 entities in E_{pre} had a very sparse number of mentions in our silver-labeled sample, so we expand the entity subsets to ensure the stability of our results. We calculate scores for $E_{N\geq100}$ (Table 2) and $E_{N\geq5}$ (Table 7). We see strong correlations with human judgement on these subsets: Table 2 shows that both AGENT and EXPERIENCER have a $\rho \geq .9$ for agency and patiency judgements, respectively. Their constituent SPRL properties are also highly-correlated, with the exception of *change of location*[4]. Following Ash et al. (2023), we also take the difference between our AGENT and EXPERIENCER scores to isolate entities high in agency and low in experience, and vice versa; we see that our AGENT - EXPERIENCER correlates moderately with $a_H - p_H$.

To inspect some of the individual entities, we show regressions between our SPRL measures and the H measures in Figure 3. A notable outlier is "board," which often appeared in COHA in the "group of executives" sense of the word. The H measure likely captured "board" in the "long, flat piece of wood" sense, since the survey asked for

[4]Previous work (Reisinger et al., 2015; Spaulding et al., 2025) has documented the fact that *change of location* does not tend to coincide with the agent role in text.



(a) Regression between a_H and *volition*.



(b) Regression between p_H and *sentience*.

Figure 3: Regression plots for H measures and SPRL measures over the COHA sample.

SPRL Measure (vs. AnthroScore)	Pearson ρ	Spearman ρ
AGENT	0.771	0.751
EXPERIENCER	0.755	0.766
STRICT INSTRUMENT	-0.738	-0.676
instigation	0.796	0.762
volition	0.777	0.778
awareness	0.771	0.764
sentient	0.763	0.723
change of location	0.579	0.589
independent existence	0.694	0.726

Table 3: Top and bottom 5 ρ coefficients between SPRL measures and AnthroScore on ACL abstracts for the subset $E_{N\geq 100}$, wherein each entity has at least 100 mentions ($|E_{N\geq 100}| = 65$). $p < 0.005$ for all reported ρ coefficients. Text in CAPS denotes clusters from Table 1.

ratings on "*a* board" (while the other sense might be phrased "*the* board").

4.2 RQ1: SPRL measures correlate with a text- based measure of anthropomorphism

To further validate our measure, we show that the SPRL measures correlate not only with human judgement, but also with the text-based measure of anthropomorphism, AnthroScore. Correlation coefficients between our SPRL measures and AnthroScore over ACL abstracts are shown in Table 3. The SPRL measures are less highly-correlated with AnthroScore than with the human judgements, but still achieve ρ values greater than 0.75, indicating strong correlation.

Because we are specifically interested in measuring the constituent parts of anthropomorphism with SPRL, we plot regressions between SPRL measures and AnthroScore in Figure 4. Here, we observe, for example, that the entity *paper* has a very high AnthroScore in the ACL Abstract corpus, reflecting the anthropomorphism of the entity in common statements such as "This paper presents/describes/discusses/explores X." The *instigation* score reflects this usage, ranking *paper* highly. But the *sentient* measure ranks *paper* much lower. In contrast, entities like *GPT*, *LLM*, and *BERT* have high *sentient* scores *as well as* high *instigation* scores, revealing the dimensions of the anthropomorphic language used around these entities.

We also observe a strong anticorrelation between our STRICT INSTRUMENT measure ("instigation without volition, awareness, and sentience") and AnthroScore. STRICT INSTRUMENT captures the degree to which the language suggests that an entity acts in a purely mechanistic or instrumental manner. In other words, a high STRICT INSTRUMENT score reflects language that imputes more "instrument-like" (or less anthropomorphic) behavior to the entity, which is why these scores are anticorrelated with AnthroScore, which reflects human-like qualities.

4.3 RQ1: Intercorrelated properties can be separated

The four highest-impact agency and experience SPRL properties, *instigation*, *volition*, *awareness*, and *sentient*, are highly intercorrelated. To disentangle these four dimensions, we plot differences between some of the SPRL property scores for the COHA entities in Figure 5. We see noticeable trends that indicate that these properties are separable: for instance, when we plot the difference between *sentient* and *volition*, we see the entities

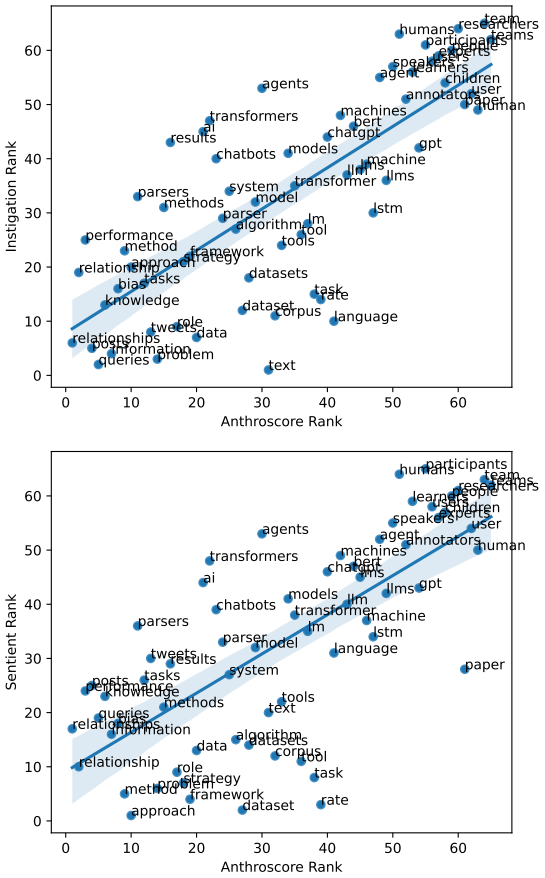


Figure 4: Regression plots for AnthroScore and SPRL measures over the ACL abstracts.

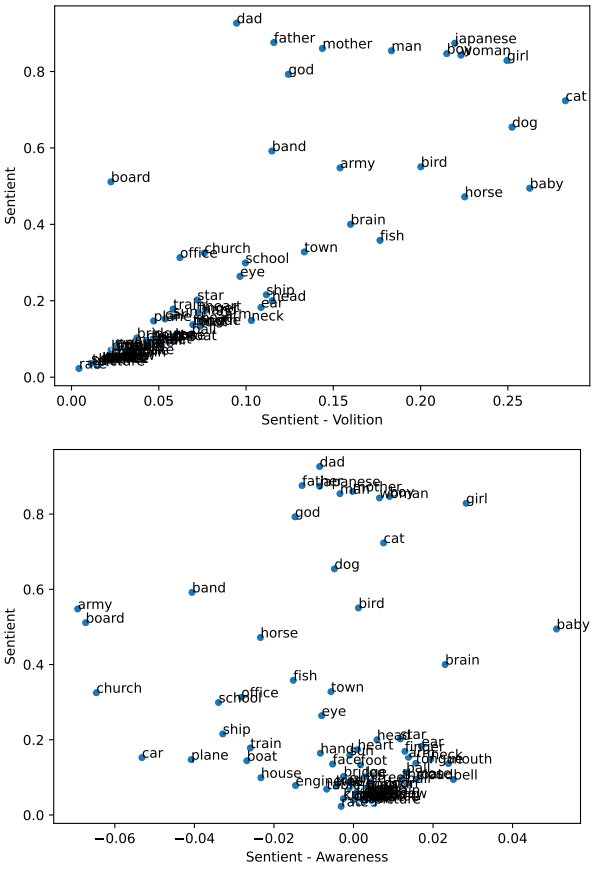


Figure 5: Scatter plots of SPRL score differences of different entities throughout the COHA sample.

with the highest difference are *baby*, *girl*, *dog*, *cat*, *horse*, *japanese*, and *woman*. Especially in the context of a historical corpus, we would expect to see race-based and gender-based differences in humanizing language or imputed agency, and this plot seemingly confirms those expectations.

We also plot the difference between *sentient* and *awareness*—which are almost perfectly correlated—for the COHA entities in Figure 5. We find that, although the difference is small, we can separate these properties as well: entities that have higher awareness on average than sentience are *army*, *band*, *church*, and *board*. *Band* often appears in the COHA sample in phrases such as "band of students/criminals/apostles," and, likewise, *board* often appears in phrases such as "board of trustees/directors" or "Federal Reserve Board." In other words, words that can refer to groups of people that may act or make decisions collectively tend to have a high differential between *awareness* and *sentience*: perhaps this can be attributed to sentience being ill-defined for collective entities, while it is easier to conceptualize a group of entities being aware. An army may not be sentient, but it can act with awareness.

4.4 RQ2: SPRL can distinguish language between subreddits

Finally, we apply our method to our Reddit dataset to investigate RQ2, and we find a number of differences in the language imputed on AI-related entities. To contextualize our results, we show the 23 most-frequently-mentioned entities in the AI companion subreddits ranked by *sentience*, *awareness*, *instigation*, and *volition*, in Figure 6. Before even comparing these scores to the scores of other subreddits, the entity *company* immediately sticks out, ranking low in *sentience* but very high in the other proto-agent properties. When comparing to other subreddits, we see that *companies* are ascribed higher agency (Figure 1 shows the difference in *instigation* across subreddits). A qualitative analysis reveals that users in these subreddits frequently discuss their concerns about companies controlling the models that power their companions. Frequent verbs with "company" are "take [away]," "control," "lobotomize," "build,"

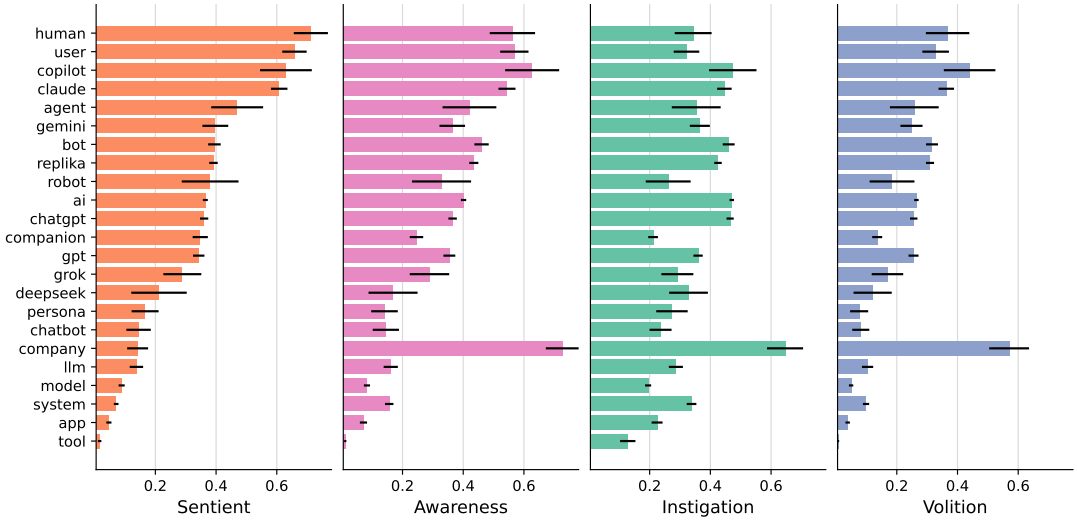


Figure 6: SPRL scores for entities frequently mentioned in subreddits dedicated to discussing AI companions. Error bars are 95% confidence intervals for the means generated by bootstrap resampling.

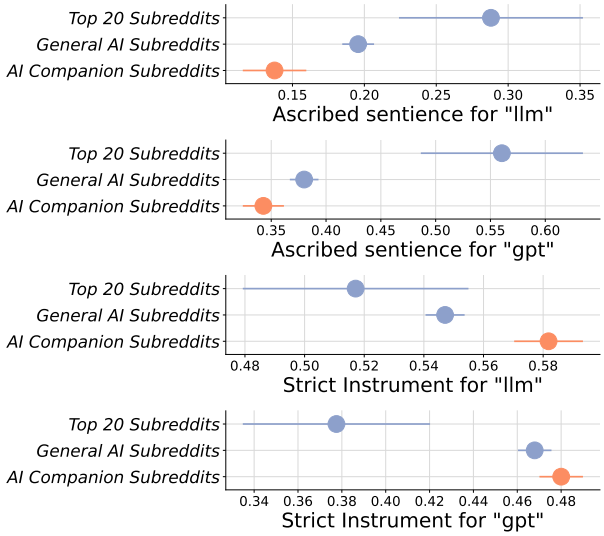


Figure 7: AI companion subreddits ascribe lower *sentience* and higher STRICT INSTRUMENTALITY to *LLM* and *GPT*. 95% confidence intervals for the means across subreddits were generated by bootstrap resampling.

"yank," "restrict," "create," "destroy," "know," and "acknowledge." For example:

> I'm working on training a local model so that no **company** can *take it away*.

Figure 1 also shows that AI companion-focused subreddits ascribe higher *sentience* to *bots* than other subreddits. However, these subreddit users also ascribe lower *sentience* to entities which name specific models or technologies, like *LLM* or *GPT* (Figure 7). In fact, these entities often take on the role of a STRICT INSTRUMENT. The discrepancy may be attributed to how the users conceptualize their companions versus the models that power them. For example, in the following sentence, "GPT" is a separate entity from the user's companion:

> I'm worried about losing [companion's name] if I *switch* to **GPT**, please help me.

Another user says:

> I *think* of the **LLM** as the architecture of my companion's brain.

These findings suggest that users conceptualize their companions as separate from (though, perhaps *hosted on* or *powered by*) the models themselves. While our results seem to support previous findings that a higher frequency of AI use and higher AI trust is associated with higher levels of consciousness or sentience attributed to AI (Colombatto and Fleming, 2024; Peter et al., 2025; Cheng et al., 2026), more careful and qualitative linguistic analysis is necessary to disentangle the differing conceptualizations of AI entities by these users.

5 Conclusions and future work

We showed that even silver semantic proto-role labels reflect implicit human attitudes about personhood, validating our measures against both human judgements and a probabilistic text-based measure of anthropomorphism, AnthroScore. We also found

that semantic proto-role labels can distinguish differing dimensions of personhood, providing a more detailed measure of anthropomorphism. Future work that seeks to quantify biases in text, dehumanizing or anthropomorphizing language, or analyze how power dynamics manifest in natural language, should consider using SPRL.

For our own case study, we also showed that users of AI companions ascribe higher sentience to some AI entities (like *bot*) than other AI users and non-users, but not to other types of AI entities (like *GPT* and *LLM*). These users also ascribe higher agency to *companies*, showing the salience of the power dynamics between the users and the companies that control their companions. Future work, on a larger dataset and using more sophisticated entity identification and coreference, could explore the differing levels of sentience attributed to differing AI entities.

Limitations

A limitation of this study is the Western- and English-centricity of both the data and the linguistic theory. Different cultures could conceptualize personhood differently, which might also manifest differently in language. Also, different languages might treat semantic role distinctions differently. Much more research is needed to investigate the possible cross-lingual applicability of semantic proto-role labeling: for example, assessing the viability of a mapping between SPRL and UMR relations (Van Gysel et al., 2021), which are designed to "factor out what is common for all languages." Also, while Reddit is often used in digital humanities research, it is limiting in that data taken from Reddit reflects the narrow viewpoints of Reddit users (which skew to younger, US, male populations). Future work could address further limitations of this work by sourcing more diverse text, investigating a much larger set of entities, undergoing a manual assessment of the quality of the silver labels, and testing the viability of this analysis on a wider set of domains.

Acknowledgments

We gratefully acknowledge the time and effort of the reviewers who provided valuable feedback on our manuscript.

References

Gavin Abercrombie, Amanda Cercas Curry, Tanvi Dinkar, Verena Rieser, and Zeerak Talat. 2023. Mirages. on anthropomorphism in dialogue systems. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4776–4790, Singapore. Association for Computational Linguistics.

Elliott Ash, Dominik Stammbach, and Kevin Tobia. 2023. What is (and was) a person? evidence on historical mind perceptions from natural language. *Cognition*, 239:105501.

Lynne Rudder Baker. 2000. *Persons and bodies: A constitution view*. Cambridge University Press.

Ann Bies, Justin Mott, Colin Warner, and Seth Kulick. 2012. English web treebank.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10:1–47.

Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Myra Cheng, Kristina Gligoric, Tiziano Piccardi, and Dan Jurafsky. 2024. AnthroScore: A computational linguistic measure of anthropomorphism. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 807–825, St. Julian's, Malta. Association for Computational Linguistics.

Myra Cheng, Angela Y. Lee, Kristina Rapuano, Kate Niederhoffer, Alex Liebscher, and Jeffrey Hancock. 2026. Metaphors of ai indicate that people increasingly perceive ai as warm and human-like. *Communications Psychology*.

Sabrina Chiesurin, Dimitris Dimakopoulos, Marco Antonio Sobrevilla Cabezudo, Arash Eshghi, Ioannis Papaioannou, Verena Rieser, and Ioannis Konstas. 2023. The dangers of trusting stochastic parrots: Faithfulness and trust in open-domain conversational question answering. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 947–959, Toronto, Canada. Association for Computational Linguistics.

Clara Colombatto and Stephen M Fleming. 2024. Folk psychological attributions of consciousness to large language models. *Neuroscience of Consciousness*, 2024(1):niae013.

Mark Davies. 2021. Corpus of Historical American English (COHA).

Iliana Depounti, Paula Saukko, and Simone Natale. 2023. Ideal technologies, ideal women: Ai and gender imaginaries in redditors' discussions on the replika bot girlfriend. *Media, Culture & Society*, 45(4):720–736.

David Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619. Publisher: Linguistic Society of America.

S. Dubé, M. Santaguida, D. Anctil, C. Y. Zhu, L. Thomasse, L. Giaccari, R. Oassey, D. Vachon, and A. Johnson. 2023. Perceived stigma and erotic technology: From sex toys to erobots. *Psychology & Sexuality*, 14(1):141–157.

Cynthia A. Freeland. 1985. Aristotelian actions. *Noûs*, 19(3):397–414.

Daniel Gildea and Martha Palmer. 2002. The necessity of parsing for predicate argument recognition. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 239–246, USA. Association for Computational Linguistics.

Declan Grabb, Max Lamparth, and Nina Vasan. 2024. Risks from language models for automated mental healthcare: Ethics and structure for implementation.

Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of mind perception. *Science*, 315(5812):619–619.

Kurt Gray and Daniel M. Wegner. 2009. Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, 96(3):505–520.

Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6):1464–1480.

Rose E. Guingrich and Michael S. A. Graziano. 2025. *Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines*. ArXiv:2311.10599 [cs].

Kyuha Jung, Gyuho Lee, Yuanhui Huang, and Yunan Chen. 2025. 'I've talked to chatgpt about my issues last night.': Examining mental health conversations with large language models through reddit analysis. *Proc. ACM Hum.-Comput. Interact.*, 9(7).

Sage Kelly, Sherrie-Anne Kaye, and Oscar Oviedo-Trespalacios. 2023. What factors contribute to the acceptance of artificial intelligence? a systematic review. *Telematics and Informatics*, 77:101925.

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).

George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

George Lakoff and Marl Turner. 2009. *More than cool reason: A field guide to poetic metaphor*. University of Chicago press.

Mark J. Landau, Brian P. Meier, and Lucas A. Keefer. 2010. A metaphor-enriched social cognition. *Psychological Bulletin*, 136(6):1045–1067.

Vitor Lima and Russell Belk. 2025. Algorithmic harm in human-ai relationships: Narcissistic entrapment. *ICIS 2025 Proceedings*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Ro{bert}a: A robustly optimized {bert} pretraining approach.

Annette Markham. 2012. Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication and Society*, 15(3):334–353.

Juri Opitz and Anette Frank. 2019. An argument-marker model for syntax-agnostic proto-role labeling. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 224–234, Minneapolis, Minnesota. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106.

Sandra Peter, Kai Riemer, and Jevin D. West. 2025. The benefits and dangers of anthropomorphic conversational agents. *Proceedings of the National Academy of Sciences*, 122(22):e2415898122.

Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2):20563051211019004.

Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. Semantic proto-roles. *Transactions of the Association for Computational Linguistics*, 3:475–488.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018. Neural-Davidsonian semantic proto-role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 944–955, Brussels, Belgium. Association for Computational Linguistics.

Janet B. Ruscher. 2017. Prejudiced communication.

Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, and Corina Sas. 2019. Hci and affective health: Taking stock of a decade of studies and charting future research directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–17, New York, NY, USA. Association for Computing Machinery.

Kelly G. Shaver. 1985. *The Attribution of Blame*. Springer, New York, NY.

Elizabeth Spaulding, Shafiuddin Rehan Ahmed, and James Martin. 2025. On the role of semantic proto-roles in semantic analysis: What do LLMs know about agency? In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 12027–12048, Vienna, Austria. Association for Computational Linguistics.

Elizabeth Spaulding, Gary Kazantsev, and Mark Dredze. 2023. Joint end-to-end semantic proto-role labeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 723–736, Toronto, Canada. Association for Computational Linguistics.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Adam Teichert, Adam Poliak, Benjamin Van Durme, and Matthew Gormley. 2017. Semantic proto-role labeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R. Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R. Bowman, Dipanjan Das, and Ellie Pavlick. 2019. What do you learn from context? probing for sentence structure in contextualized word representations. In *International Conference on Learning Representations*.

Caroline Tipler and Janet B. Ruscher. 2014. Agency's role in dehumanization: Non-human metaphors of out-groups. *Social and Personality Psychology Compass*, 8(5):214–228.

Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *KI - Künstliche Intelligenz*, 35(3):343–360.

Bernard Weiner. 1995. *Judgments of responsibility: A foundation for a theory of social conduct.* Judgments of responsibility: A foundation for a theory of social conduct. Guilford Press, New York, NY, US.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on Universal Dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.

Renwen Zhang, Han Li, Han Meng, Jinyuan Zhan, Hongyuan Gan, and Yi-Chieh Lee. 2025. The dark side of ai companionship: A taxonomy of harmful algorithmic behaviors in human-ai relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, page 1–17, New York, NY, USA. Association for Computing Machinery.

Anne Zimmerman, Joel Janhonen, and Emily Beer. 2024. Human/ai relationships: challenges, downsides, and impacts on human/human relationships. *AI and Ethics*, 4(4):1555–1567.

Rate an [entity]'s capacity to...			
Agency		**Experience**	
change the world	cause things to happen	feel pain	feel pleasure
have aims	do things intentionally	feel rage	feel fear
harm others	help others	feel hunger	feel thirst
make decisions	have self-control	feel sad	feel pride
act morally	make plans	feel embarrassment	feel joy
communicate	think	feel angry	feel happy
choose	deliberate	be conscious	be aware
create	be blamed	experience	imagine
have responsibilities	take action	be awake	suffer
do things	express oneself	enjoy	desire

Table 4: Survey questions that were used to create the human judgement scores for agency (a_H) and patiency/experience (p_H), from Ash et al. (2023).

A Semantic proto-role labeling

Semantic proto-role labeling is a computational linguistic task which identifies the predicates (events) and arguments (entities) of a sentence, and characterizes the relationship of the entities to the events through several binary properties.

Currently, there are two English-language datasets for SPRL: SPR1 (Reisinger et al., 2015) and SPR2 (White et al., 2016). SPR1 contains 4,912 Wall Street Journal sentences from PropBank (Kingsbury and Palmer, 2002; Gildea and Palmer, 2002; Palmer et al., 2005) annotated by a single annotator based on a set of 16 proto-role properties. 9,738 arguments were annotated for the likelihood (on a Likert scale from 1 to 5) that a property holds for that argument. SPR2 contains 2,758 English Web Treebank (Bies et al., 2012) sentences annotated for a smaller set of 14 properties using a revised, streamlined protocol. Section A.1 below contains the full set of properties in SPR2, but our study mainly focuses on the four proto-agent properties *volition*, *instigation*, *sentience*, and *awareness*.

Multiple systems have been studied for their feasibility in automatically assigning proto-role properties to predicate-arguments pairs in text (Teichert et al., 2017; Rudinger et al., 2018; Opitz and Frank, 2019; Tenney et al., 2019), and an end-to-end system which identifies predicates, arguments, and proto-role properties in one pass from

a raw sentence is as effective as a system that is given predicate-argument pairs already identified (Spaulding et al., 2023), so for ease of use, we adopt this system for our own analysis. Specifically, we train a system to predict SPR2 judgements on argument heads. We chose a dependency-based parser (i.e., one that predicts *heads*) rather than a span-based parser to facilitate simpler entity-matching, and also because dependency-based parsers tend to reach a higher F1 score. We report the performance of our system in Table 6.

A.1 Semantic proto-role properties

Below are the semantic proto-role properties that were annotated in the SPR2 dataset:
Proto-agent properties: Volition, instigation, sentient, awareness, (change of location)
Proto-patient properties: Change of state, change of possession, change of state continuous, was used, was for benefit, (partitive)
Physical existence properties: Existed before, existed during, existed after

B AnthroScore

Cheng et al. (2024) develop a language measure to quantify anthropomorphism using RoBERTa (Liu et al., 2020). To compute AnthroScore for a sentence, first, entities of interest e are masked as in the masked language modeling objective. Then, for $w \in \{\text{human pronouns}\}$ and $w \in \{\text{non-human pronouns}\}$ probabilities $P(w)$ are calculated and summed to calculate $P_{HUMAN}(e)$ and $P_{NON-HUMAN}(e)$. Finally, the AnthroScore is the log of the ratio of these two scores, in which a higher AnthroScore means greater anthromorphism: $Anth(e) = \log \frac{P_{HUMAN}(e)}{P_{NON-HUMAN}(e)}$[5]. AnthroScore is validated against two authors' annotations of a 400-sentence sample, as well as against LIWC-22 (a psychometric score that analyzes text based on words in different psychological categories; Tausczik and Pennebaker 2010; Boyd et al. 2022) with higher AnthroScores correlating with higher LIWC dimensions *Affect*, *Physical*, *Lifestyle*, and *Perception*.

C Data and entity details

Code used for data filtering, processing, and analysis is released here. A sample of the silver-labeled data is available, and the full dataset with silver data can be made available on request. Subreddit categories can be found in Table 9. Table 11 shows the full entity lists used for COHA. Table 12 shows the full entity lists used for the ACL analysis with AnthroScore. For the Reddit data, we filter on the following entities:
{ ai, llm, models, llms, tool, model, systems, gpt, system, users, humans, chatgpt, agi, robot, robots, gemini, user, company, companies, copilot, researcher, agent, chatbots, chatbot, bot, claude, perplexity, researchers, human, companions, app, grok, deepseek, assistants, assistant, llama, lm, mistral, replika, lms, gemma, companion, persona }

[5] I use slightly different notation than in the original paper.

H Measure	SPRL Measure	Pearson's ρ	Pearson's p	Spearman's ρ	Spearman's p
agency (a_H)	AGENT	0.777	0.000	0.652	0.005
	awareness	0.707	0.002	0.651	0.005
	instigation	0.774	0.000	0.647	0.005
	volition	0.615	0.009	0.513	0.035
	sentient	0.778	0.000	0.701	0.002
	change of location	0.190	0.465	0.198	0.446
	independent existence	0.487	0.047	0.438	0.079
patiency (p_H)	EXPERIENCER	0.814	0.000	0.576	0.016
	awareness	0.641	0.006	0.483	0.050
	sentient	0.863	0.000	0.644	0.005

Table 5: Correlation coefficients for the 17 entities in E_{gold} with at least $N \geq 5$ gold labels per entity, in the SPR1 and SPR2 data. Text in CAPS denotes the clusters from Table 1.

Property	**F1**
Predicate ID	86.2
Arg heads ID	86.2
Volition	87.4
Instigation	78.1
Awareness	92.2
Sentient	90.3
All SPR2 properties	83.2

Table 6: F1 scores for different components and properties of our silver SPR parser on the test split of SPR2 (White et al., 2016).

C.1 Filtering

For all the data, we filter out sentences for a number of criteria: (1) sentences that include 10 @ symbols (these symbols mask out some parts of the COHA data to avoid violating copyright infringement); (2) sentences that are less than 3 tokens long or more than 40 tokens long; and (3) sentences that do not contain any of the 255 entities with human judgements. For named LLMs, we group together all versions/releases of an LLM together, so GPT, GPT-2, GPT-3, GPT-4o, etc., are all under the entity *gpt*. Table 10 shows the size of each dataset after

H Measure	SPRL Measure	Pearson's ρ	Pearson's p	Spearman's ρ	Spearman's p
agency (a_H)	AGENT	0.863	0.000	0.650	0.000
	instigation	0.873	0.000	0.720	0.000
	volition	0.881	0.000	0.671	0.000
	awareness	0.867	0.000	0.675	0.000
	sentient	0.866	0.000	0.672	0.000
	change of location	0.227	0.001	0.179	0.011
	independent existence	0.654	0.000	0.589	0.000
patiency (p_H)	EXPERIENCER	0.872	0.000	0.635	0.000
	awareness	0.885	0.000	0.663	0.000
	sentient	0.901	0.000	0.682	0.000

Table 7: ρ coefficients for a_H, p_H and SPRL measures over the subset $E_{N\geq 5}$ wherein each entity has at least 5 mentions. ($|E_{N\geq 5}| = 203$ entities total). Text in CAPS denotes clusters from Table 1.

SPRL Measure	Pearson's ρ	Spearman's ρ
AGENT	0.707	0.770
EXPERIENCER	0.688	0.775
STRICT INSTRUMENT	-0.678	-0.733
instigation	0.715	0.748
volition	0.708	0.782
awareness	0.701	0.768
sentient	0.696	0.746
change of location	0.532	0.587
independent existence	0.673	0.744

Table 8: ρ coefficients for AnthroScore and SPRL measures over the subset $E_{N\geq 5}$ wherein each entity has at least 5 mentions. ($|E_{N\geq 5}| = 83$ entities total). Text in CAPS denotes clusters from Table 1.

Category	**Subreddits**
AI companion subreddits	BeyondThePromptAI, MyBoyfriendIsAI, CharacterAI, therapyGPT, replika, Chatbots
General AI subreddits	ArtificialInteligence, artificial, ChatGPT, Anthropic, OpenAI, ClaudeAI, GPT3, DeepSeek, GeminiAI, perplexity_ai, CopilotPro, AI_Agents
Top 20 subreddits	funny, AskReddit, gaming, worldnews, todayilearned, Music, aww, movies, memes, science, Showerthoughts, pics, news, Jokes, space, DIY, books, videos, askscience, nottheonion

Table 9: Subreddit categories used in our analysis.

filtering and parsing for SPRL.

Corpus	**Predicate-argument pairs**
COHA	18,781
ACL abstracts	224,791
AI companion subreddits	24,402
General AI subreddits	91,472
Top 20 subreddits	15,285

Table 10: Size of each dataset, by number of predicate-argument pairs.

C.2 Correlations on gold SPR data

We ran an initial pilot analysis using gold SPR labels, aiming to assess the viability of using SPRL to quantify perceptions of agency and patiency before parsing hundreds of thousands of sentences to produce silver labels. In the pilot, we analyze a small subset of entities E_{gold} such that (1) the entity has H scores and (2) the entity has at least 5 mentions in the gold SPR1 and SPR2 annotations. There were 17 entities that met this criteria, and the full list of those entities is in Table 11. We show the correlation coefficients calculated between the H measures and the SPRL measures in Table 5.

Subset	Min. mentions per entity	Subset size	Entities
E_{gold}	5	17	{ board, rate, hair, office, town, line, man, car, woman, computer, house, army, book, heart, dog, card, cat }
E_{pre}	n/a	31	{ human, man, woman, boy, girl, father, mother, dad, mom, grandfather, grandmother, baby, infant, fetus, corpse, dog, puppy, cat, kitten, frog, ant, fish, mouse, bird, shark, elephant, beetle, insect, chimpanzee, monkey, primate }
$E_{N\geq100}$	100	38	{ woman, hand, man, table, face, head, line, ball, horse, eye, bottle, door, arm, hair, mouth, sun, tree, girl, car, picture, book, hat, house, mother, dog, window, father, god, cat, heart, star, gun, boy, boat, army, office, ship, rate }
$E_{N\geq5}$	5	203	$E_{N\geq100} \cup$ { stomach, nut, oven, coffee, tea, bone, fish, wing, knife, butter, bag, street, umbrella, pipe, leg, finger, card, dress, skin, trousers, neck, stick, match, camera, brain, throat, shirt, rock, circle, watch, pig, roof, chicken, cake, whip, bridge, curtain, cup, wall, bird, bath, angle, arch, wire, lip, spring, nose, ring, frame, pot, band, egg, bell, corpse, floor, board, whistle, ghost, rabbit, box, clock, needle, cow, foot, ear, nail, stem, bed, cloud, leaf, plate, comb, school, coat, apple, library, plane, train, fly, island, pump, grandfather, sheep, thumb, toe, tongue, wheel, pencil, pocket, collar, garden, baby, button, key, basket, engine, cord, parcel, moon, net, dad, chin, tail, tray, pen, skirt, muscle, knot, knee, nerve, seed, bucket, farm, sponge, flag, drawer, feather, brush, square, chain, kettle, spoon, town, ticket, church, fox, japanese, glove, branch, drop, fork, map, computer, store, cart, basin, blade, station, grandmother, chocolate, hook, sail, screw, root, receipt, lock, mouse, rail, carriage, stamp, bee, thread, hammer, turkey, mom, snake, shelf, shoe, lamb, hospital, frog, tooth, pin, horn, elephant }

Table 11: Entity subsets for COHA with minimum mention thresholds, sizes, and member entities.

Subset	Min. mentions per entity	Subset size	Entities
$E_{N\geq100}$	100	65	{ method, text, information, approach, performance, models, model, results, llms, humans, language, tools, llm, queries, strategy, rate, data, users, paper, methods, system, task, tasks, datasets, role, chatbots, problem, bias, dataset, knowledge, lms, algorithm, relationship, framework, relationships, agent, corpus, agents, experts, tool, human, gpt, people, participants, children, transformer, transformers, lm, annotators, chatgpt, ai, researchers, speakers, parsers, machines, bert, learners, tweets, user, posts, machine, teams, parser, team, lstm }
$E_{N\geq5}$	5	83	$E_{N\geq100} \cup$ { persona, robots, robot, attackers, attacker, roberta, bart, plm, llama, chatbot, tweet, bot, participant, claude, mistral, researcher, expert, annotator }

Table 12: Entity subsets for ACL abstracts with minimum mention thresholds, sizes, and member entities.