Social Influence Dialogue Systems: A Survey of Datasets and Models For Social Influence Tasks

Kushal Chawla[1*] **Weiyan Shi**[2*] **Jingwen Zhang**[3**]
Gale Lucas[1**] **Zhou Yu**[2**] **Jonathan Gratch**[1**]
[1]University of Southern California [2]Columbia University
[3]University of California Davis
[1]{chawla, lucas, gratch}@ict.usc.edu [2]{ws2634, zy2461}@columbia.edu
[3]jwzzhang@ucdavis.edu

Abstract

Dialogue systems capable of social influence such as persuasion, negotiation, and therapy, are essential for extending the use of technology to numerous realistic scenarios. However, existing research primarily focuses on either task-oriented or open-domain scenarios, a categorization that has been inadequate for capturing influence skills systematically. There exists no formal definition or category for dialogue systems with these skills and data-driven efforts in this direction are highly limited. In this work, we formally define and introduce the category of *social influence dialogue systems* that influence users' cognitive and emotional responses, leading to changes in thoughts, opinions, and behaviors through natural conversations. We present a survey of various tasks, datasets, and methods, compiling the progress across seven diverse domains. We discuss the commonalities and differences between the examined systems, identify limitations, and recommend future directions. This study serves as a comprehensive reference for social influence dialogue systems to inspire more dedicated research and discussion in this emerging area.

1 Introduction

Consider a human user who signs up to interact with a persuasive dialogue system that motivates for engaging in physical exercise. The system: 1) uses *social cues like small talk and empathy*, useful for providing continued support, and 2) employs *persuasive strategies to convince* the user who, at least in the short-term, is reluctant to indulge in exercise. Does such a system fit the definition of a *task-oriented dialogue system* that are traditionally designed to assist users in completing their tasks such as restaurant or flight booking (Zhang et al., 2020c)? Although the system is task-oriented or goal-oriented per se, the task here goes beyond the traditional definition of assisting users, given the

possible misalignment between the goals of the system and the user. Clearly, this system is also not open-domain (Huang et al., 2020). Although conversations involve social open-ended interactions, there is still a concrete goal of persuading the user towards a healthier habit.

Scenarios similar to above are ubiquitous in everyday life, including games (Peskov et al., 2020), social platforms (Tan et al., 2016), and therapeutic interactions (Tanana et al., 2016). Dialogue systems for these applications require a core function in human communication, that is, *social influence* (Cialdini and Goldstein, 2004; Cialdini, 2009), which involves influencing users' cognitive and emotional responses, leading to changes in thoughts, opinions, and behaviors through natural conversations. This goes beyond what is captured by traditional task definitions in the dialogue community which primarily focus on task completion and social companionship.

Despite numerous independent efforts in identifying and analyzing various social influence scenarios, there is a lack of common understanding around social influence in AI research which inhibits a systematic study in this space. Further, data-driven efforts for dialogue systems in this space are highly limited. To this end, we introduce the category of *social influence dialogue systems* (Section 2), providing a comprehensive literature review and discussing future directions.

Developing these systems holds importance in AI research for multiple reasons. Tackling these tasks not only involves AI but also aspects of game theory, communication, linguistics, and social psychology, making them an ideal testbed for interdisciplinary AI research. Most importantly, they reflect AI's general ability to consider their partners' inputs, tailor the communication strategies, personalize the responses, and lead the conversation actively.

We design a taxonomy for existing social influ-

---

*Equal contribution, **Co-supervised project

ence dialogue datasets, studying their task structure (symmetric vs asymmetric) and context (local vs global). We also organize them by their domains: games, multi-issue bargaining, social good, e-commerce, therapy and support, argumentation, conversational recommendations, and miscellaneous tasks (Section 3). We further design a taxonomy of existing methods, assisting readers to comprehend the progress and reflect on future directions. We organize them based on the system strategy, language generation, partner model, architecture, learning process, and the use of pretrained language models (Section 4). Finally, we identify key challenges and provide recommendations for future work (Section 5).

Over the years, research in task-oriented and open-domain dialogues has benefited from a myriad of survey efforts (Huang et al., 2020; Zhang et al., 2020c; Ni et al., 2021). We instead focus on dialogue systems with social influence capabilities and present a thorough review across various domains. We hope that our work serves as a timely entry point for interested researchers to take this area further, inspiring dedicated effort and discussion on social influence in the dialogue community.

2 Social Influence Dialogue Systems

"Social influence is a fact of everyday life" (Gass, 2015). It is the *change in thoughts, feelings, attitudes, or behaviors resulting from interaction with an individual or a group* (Rashotte, 2007). Influence is measured by quantifiable proxies of the observed change, like the interest to indulge in physical exercise before or after the interaction with a system, or the final deal in a negotiation as opposed to one person taking it all. Social influence dialogue systems act interactively and influence their partners in decision-making and behavioral contexts (Zhang et al., 2020a; Lee et al., 2020). This calls for an active role by the system, distinguishing them from other well-studied scenarios, such as purely task-oriented, where systems passively assist their partners to complete tasks, and open-domain, that target social companionship. Key social influence tasks include persuasion (Wang et al., 2019), aiming to change users' attitudes or behaviors, and negotiation, aiming to change the users' perspective to achieve a common ground (Lewis et al., 2017).

Conceptual overview: Figure 1 distinguishes between the kinds of conversational content in social
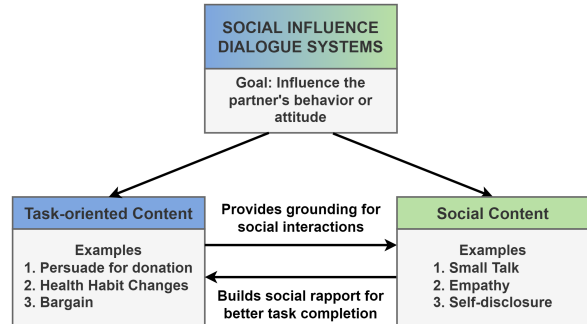

Figure 1: A conceptual overview.

influence interactions. The *task-oriented content* focuses on influencing for a domain-specific goal, like persuading for donation, bargaining with trade-offs, or encouraging healthier habits. These interactions may also contain *social content*, such as small talk, empathy, or self-disclosure. The task-oriented content provides a context for social interactions. Depending on the task, social content is optional, but if present, can in turn build rapport and enhance user-system relationship for improved task outcomes (Liao et al., 2021).

Connections with task-oriented and open-domain systems: Similar to a task-oriented or an open-domain scenario, social influence dialogue can also be seen as a sequential decision making process with the goal of maximizing the expected reward (Huang et al., 2020; Gao et al., 2018). Our proposed category is not meant to be disjoint from these traditional categories. However, it still uniquely brings together the tasks that capture social influence, *which is fundamentally absent from how we primarily define dialogue tasks in the community*. Defining a new category that captures social influence dialogue would foster a dedicated effort towards this important aspect of real-world conversations.

Task-oriented scenarios focus on collaborative information exchange for a common goal of task completion. In social influence tasks, the goals of the system and the user can be different and even conflicting, leading to collaborative or non-collaborative interactions. Further, the goals can go beyond the current task (e.g. multiple therapy interactions, repeated negotiations), leading to social interactions for long-term relationships. If a scenario involves the system's goal to influence its partner, we consider it under social influence in this paper. For instance, He et al. (2018) studied buyer-seller price negotiations. The task of the buyer is to

negotiate for a reasonable price (arguably making it task-oriented), but achieving it requires social influence skills of engaging in trade-offs and building a rapport with the seller so as to reach an agreement. **Measures of Success**: The above discussion indicates that a comprehensive evaluation of social influence systems must draw from both task-oriented and open-domain dialogue research. Since there exist surveys that discuss the evaluation in these settings (Deriu et al., 2021; Li et al., 2021), we don't cover them here in detail. However, we define three essential axes for evaluation: 1) *Linguistic Performance*, or the system's linguistic sophistication based on automatic (e.g. perplexity, BLEU) and human (e.g. fluency, consistency, coherency) evaluation. 2) *Influence Outcome*, or the ability to influence defined by objective goals like the negotiated price or weight loss after therapy. 3) *Partner Perception*, or the subjective evaluation of the user, for instance, the user's satisfaction, likeness towards the system, and interest in interacting again. In a buyer-seller negotiation, if the seller hates the buyer in the end, no matter how favorable the deal is for the buyer, one might argue that this is still a failed negotiation for the buyer. Hence, we encourage future work to take all three dimensions into account collectively.

3 Social Influence Across Diverse Application Areas

We now illustrate social influence across numerous domains and application areas. In total, we curated 22 datasets from prior work that capture social influence in various forms, spanning 12 publication venues, 4 languages, and 7 application domains (see Appendix A for details on the compilation process). In general, the datasets capture the following information about an interaction: the *non-conversational context* for the participants (e.g. negotiation preferences or other role-specific information), the *conversation* between them, and *outcome assessment*. Optionally, some datasets also gather participant demographics and personality traits, utterance-level annotations, and subjective evaluations via post-surveys.

To understand the structural similarities and differences between these datasets, we design a taxonomy with two primary dimensions: **Task Structure** (*Symmetric* vs *Asymmetric*), and **Context Definition** (*Global* vs *Local*). **Task Structure** captures whether the participant roles are defined in a sym-

metric or an asymmetric manner. For instance, a typical multi-issue negotiation is symmetric, in the sense that both parties have their own preferences and goals based on which they actively try to reach a favorable agreement (Lewis et al., 2017). On the other hand, a counseling session between a therapist and a patient is asymmetric, where the therapist attempts to emotionally support the patient by employing social influence skills (Althoff et al., 2016). **Context Definition** relates to whether the input context before each interaction is defined globally or locally. For instance, the PersuasionForGood dataset globally defines the context of persuasion for charity donation, which is kept the same throughout (Wang et al., 2019). On the contrary, in a typical debate, although the rules are defined globally, the conversation topic and arguments are local and can vary for each conversation (Durmus and Cardie, 2019). We present this categorization in Table 1. We further categorize the datasets according to their **Domain**, **Source**, and the **# of parties**. We provide key statistics and the available metadata in Appendix B. We now briefly discuss the datasets in each domain.

Games: Strategy games involve social influence dynamics of trust and deception. Diplomacy captures deception in long-lasting relationships, where players forge and break alliances to dominate Europe (Peskov et al., 2020). Catan revolves around the trade of resources for acquiring roads, settlements, and cities (Asher et al., 2016; Boritchev and Amblard, 2021). The players have access to only a subset of resources that they would need, which encourages strategic influence and trade.

Multi-Issue Bargaining Tasks (MIBT): MIBT is a tractable closed-domain abstraction of a typical negotiation (Fershtman, 1990). It is based on a fixed set of issues each with a predefined priority for each player, which essentially governs the goals of the players. If the priorities of the players align, this leads to competitive negotiations, where each party attempts to convince their partner with trade-offs and persuasive arguments. If they don't, this allows cooperative interactions where the negotiators try to find optimal divisions that benefit everyone. DealOrNoDeal (Lewis et al., 2017) involves negotiations over three issues: *books*, *balls*, and *hats*. Other datasets define a more grounded scenario, such as symmetric CaSiNo (Chawla et al., 2021b) negotiations between two campsite neighbors and asymmetric JobInterview (Yamaguchi et al., 2021)

Name (Citation)	Domain	Source	Structure	Context	# of Parties
STAC (As16)	Games	Crowdsource	Symmetric	Global	Multiparty
Diplomacy (Pe20)	Games	Crowdsource	Asymmetric	Global	Multiparty
DinG (Bo21)	Games	University game night logs	Symmetric	Global	Multiparty
Tabletop (De15)	MIBT	Face-to-face, Wizard-of-Oz	Symmetric	Local	Bilateral
DealOrNoDeal (Le17)	MIBT	Crowdsource	Symmetric	Local	Bilateral
CaSiNo (Ch21)	MIBT	Crowdsource	Symmetric	Local	Bilateral
JobInterview (YaD21)	MIBT	Crowdsource	Asymmetric	Local	Bilateral
PersuasionforGood (Wa19)	Social Good	Crowdsource	Asymmetric	Global	Bilateral
CraigslistBargain (He18)	E-commerce	Crowdsource	Asymmetric	Local	Bilateral
AntiScam (Li20)	E-commerce	Crowdsource	Asymmetric	Global	Bilateral
MI (TaC16)	Therapy & Support	Psychotherapy session logs	Asymmetric	Global	Bilateral
SMS Counseling (Al16)	Therapy & Support	SMS chat logs	Asymmetric	Global	Bilateral
EmpatheticDialogues (Ra19)	Therapy & Support	Crowdsource	Asymmetric	Global	Bilateral
Hotline Counseling (De19)	Therapy & Support	Synthetic Transcripts	Asymmetric	Global	Bilateral
mPED (LiE21)	Therapy & Support	Physical activity clinical trials	Asymmetric	Global	Bilateral
Congressional Debates (Th06)	Argumentation	U.S. Congressional transcripts	Asymmetric	Local	Multiparty
Supreme Court (Da12)	Argumentation	Oyez.org transcripts	Asymmetric	Local	Multiparty
DeCour (Fo12)	Argumentation	Italian court hearings	Asymmetric	Local	Multiparty
ChangeMyView (Ta16)	Argumentation	Reddit	Asymmetric	Local	Multiparty
DDO Debates (Du19)	Argumentation	debate.org logs	Symmetric	Local	Bilateral
Court Debates (Ji20)	Argumentation	China Court transcripts	Asymmetric	Local	Multiparty
Target-Guided (Ta19)	Miscellaneous	Crowdsource	Symmetric	Local	Bilateral

Table 1: Categorization of social influence dialogue corpora. This list is non-exhaustive, and also covers the datasets that have enabled research into various sub-tasks and analyses that can eventually be useful for dialogue systems in respective domains. MIBT: Multi-Issue Bargaining Task. Key statistics and associated metadata are in Appendix 3.

negotiations between recruiters and applicants.

Social Good: Social influence is critical for social good applications. The tactics must be personalized using knowledge that is both relevant and appealing. PersuasionForGood (Wang et al., 2019) involves asymmetric interactions led by a persuader who attempts to convince the other participant for charity donation by employing a variety of tactics. For instance, *Logical Appeal* uses reason and evidence to support the argument, while *Emotional Appeal* elicits specific emotions.

E-commerce: These tasks are typically asymmetric. A *buyer* influences the seller towards a reasonable price, while the *seller* tries to maximize their own profit. An effective system must combine price-related reasoning with language realization. CraigslistBargain (He et al., 2018) involves open-ended price negotiations with rich influence strategies like embellishments, side offers, emotional appeals, and using world knowledge. Another example is customer support interactions in AntiScam dataset (Li et al., 2020), where users defend themselves against attackers who try to steal sensitive personal information with convincing arguments.

Therapy & Support: Effective therapy using social influence aids in the treatment of mental disorders, and substance use disorders, along with changing undesirable behaviors like unhealthy diets. A counselor needs to be adaptive, personalized, should understand the core issues, and should facilitate a change in patient's perspective (Althoff et al., 2016). In SMS counseling, Althoff et al. (2016) found that linguistic influence like pushing the conversation in the desired direction is associated with perspective change. Similar scenarios were captured in other datasets as well (Demasi et al., 2019; Liang et al., 2021). Tanana et al. (2016) collected the Motivational Interviewing dataset where the goal is to elicit and explore the patient's own motivations for behavior change. EmpatheticDialogues (Rashkin et al., 2019) captured empathetic support interactions, which has been associated with rapport and better task outcomes (Kim et al., 2004; Norfolk et al., 2007; Fraser et al., 2018).

Argumentation: In addition to factuality and social proof, a convincing argument must also consider the intensity, valence, authoritativeness, and framing (Chaiken, 1987; Althoff et al., 2014). Tan et al. (2016) released the ChangeMyView logs from Reddit, involving discussions on numerous controversial topics. Other datasets include Debate Dot Org (DDO) debates on diverse topics (Durmus and Cardie, 2019), congressional proceedings (Thomas et al., 2006), and court hearings (Fornaciari and Poesio, 2012; D.-N.-M. et al., 2012; Ji et al., 2020).

Conversational Recommendation: Everyday scenarios naturally hold potential for influence via recommendations, for instance, a movie fan per-

suading their friends to watch a movie that they adore. Li et al. (2018) and Dodge et al. (2016) collected movie recommendation datasets. Instead of guiding the conversation towards a specific movie, the goal is simply to provide recommendations based on facts and personal experiences. Nevertheless, they still provide interesting examples of scenarios that can involve social influence.

Miscellaneous: The Target-Guided dataset (Tang et al., 2019) was constructed from the PersonaChat corpus (Zhang et al., 2018). Instead of being open-ended, the Target-Guided scenario defines a concrete goal of naturally guiding the conversation to a designated target subject, thereby, making it a social influence setting.

4 Methodological Progress

Having summarized the datasets that capture social influence, we now discuss the modeling approaches developed for social influence dialogue systems. Most domains have seen efforts in analyzing human dialogue behaviors and their impact on task outcomes. Examples include analyzing deception in games (Peskov et al., 2020), the impact of persuasive strategies and dialogue acts on charity donations (Wang et al., 2019), cooperative and non-cooperative strategies in MIBT (Chawla et al., 2021b), the use of emotion expression for predicting partner perceptions (Chawla et al., 2021a), and studying semantic categories of persuasive arguments on web forums (Egawa et al., 2019).

In addition, researchers have targeted various domain-specific subtasks that can be crucial for the eventual development of dialogue systems in this space. This involves research in lie detection methods (Yeh and Ku, 2021; Yu et al., 2015), discourse parsing (Shi and Huang, 2019; Ouyang et al., 2021), strategy prediction (Chawla et al., 2021b; Wang et al., 2019), breakdown detection (Yamaguchi et al., 2021), outcome prediction (Sinha and Dasgupta, 2021; Chawla et al., 2020; Dutt et al., 2020), and argument mining (Dutta et al., 2022).

Research that directly targets the development of dialogue systems in this space is still nascent. Among other challenges like limited cross-cultural diversity and relatively smaller dataset size, social influence dialogue settings pose a unique challenge: an average human often exhibits sub-optimal strategic behaviors in social influence tasks (Wunderle, 2007; Babcock and Laschever, 2009). This means that standard seq2seq approaches trained on these

collected datasets using supervised learning are fundamentally insufficient for developing effective dialogue systems with influence capabilities. Hence, prior work has put a special attention to the system strategy, employing different ways to model the strategy and language together.

We design a taxonomy of methods developed for social influence tasks, assisting readers to comprehend the progress and reflect on future directions. We organize them based on the system strategy, language generation, partner model, architecture, learning process, and the use of pretrained language models. We present annotations for all the surveyed methods in Table 2 and discuss the common categories in brief below.

4.1 Strategy Representation

Implicit: The most obvious way to represent the system strategy is *implicitly*, without any intended decoupling between system strategy and response realization. This corresponds to the usual sequence-to-sequence framework that has been a standard baseline for the methods developed in this space. An important example is the work by Lewis et al. (2017), who were one of the first works to train end-to-end dialogue models that exhibit social influence. The authors employed a neural network based on GRUs, one for encoding the negotiation context, one to encode the dialogue utterances, and two recurrent units to generate the output agreement in a bidirectional manner.

Latent vectors: Yarats and Lewis (2018) explored *latent vectors* to decouple utterance semantics from its linguistic aspects. Their hierarchical approach first constructs a latent vector from the input message, which is then used for response generation and planning. These latent vectors are trained to maximize the likelihood of future dialogue messages and actions, which enables the decoupling between semantics and realization.

Dialogue Acts (DAs): Dialogue Acts, such as greeting, offer propose, agreement, or disagreement, are effective at capturing a high-level structure of the dialogue flow in social influence settings, reducing the model strategy to first predicting the dialogue act for the next response. The use of DAs makes it convenient to apply reinforcement learning approaches (Zhang et al., 2020b; Yang et al., 2021), while also aiding in developing a modular dialogue system design (He et al., 2018).

Semantic Strategies: The structural properties ex-

Method	Domain	Strategy	NLG	Partner Model	Architecture	Learning	PLM
Le17	MIBT	Implicit	Generation	Simulated User	Enc-Dec	RL	
Ya18	MIBT	Latent Vectors	Generation	Simulated User	Hierarchical	RL	
Zh20	MIBT	DAs	Generation	DA Look-Ahead	Modular	RL	
He18	E-Com, MIBT	DA	Templates + Retrieval	Implicit	Modular	RL	
Ya21	E-Com	DAs	Templates + Retrieval	DA Look-Ahead	Modular	RL	
Zh19	E-Com, Social Good	DAs + Semantic	Generation	Implcit	Hierarchical	SL	
Jo20	E-Com	DAs + Semantic	Generation	Implcit	Hierarchical	SL	
Li20	E-Com, Social Good	DAs	Generation	Implicit	Dec-only	SL	GPT
Li21	Therapy	Implicit	Generation	Implicit	Enc-Dec	SL	
Jh21	Therapy	Emotion Labels	Retrieval, Generation	Simulated User	Modular	RL	BERT, GPT
Ha21	Argumentation	DAs	Rule-based	Implicit	Modular	SL	
Wu21	Social Good	Implicit	Generation	Implicit	Dec-only	SL	GPT2
Ta19	Misc	Keywords	Retrieval	Implicit	Modular	SL	

Table 2: Categorization of methods (non-exhaustive) for social influence dialogue. We only cover papers that explicitly design a dialogue system. NLG: Natural Language Generation, PLM: Pretrained Language Model, MIBT: Multi-Issue Bargaining Task, E-Com: E-Commerce, DA: Dialogue Act, Enc: Encoder, Dec: Decoder, SL: Supervised Learning, RL: Reinforcement Learning. Methods that use RL usually apply it in conjunction with SL.

pressed by DAs are insufficient for capturing semantics like emotion, small talk, and appeal. To better incorporate them, researchers have relied on additional utterance-level annotations grounded in prior theories in social influence contexts (Wang et al., 2019; Chawla et al., 2021b). These strategies have been used in conjunction with DAs (Zhou et al., 2019; Joshi et al., 2020).

4.2 Language Generation

An important aspect of the system design is an effective way to realize the language, that is, to generate the next response so that it portrays the desired strategic behaviors. Borrowing from task-oriented and open-domain research, existing dialogue models for social influence use a variety of methods to generate the final system response.

Templates and retrieval methods: Predefined templates and response retrieval from the training data simplify the generation pipeline, improving controllability and modularity. He et al. (2018) used templates in their generator which are later filled by retrieving similar responses from the data. This allowed the authors to explore supervised and reinforcement learning at the level of DAs for the influence strategy of the system.

Conditional Generation: Text generation methods result in more diverse responses, but negatively impact the controllability and interpretability. Prior work relies on autoregressive text generation conditioned on the dialogue history, non-conversational context, and additional annotations. These are either encoder-decoder networks (Lewis et al., 2017; Li et al., 2020; Joshi et al., 2020) or use a decoder-only design (Li et al., 2020). A useful future direction is to combine generation with retrieval for

knowledge-grounded settings like argumentation. Similar methods have been explored for other NLP tasks like open-domain question answering and question generation (Lewis et al., 2020).

4.3 Partner Modeling

Partner modeling refers to inferring the mental states of the partner based on the conversation. For example, understanding the cause that the persuadee cares about in the PersuasionForGood context, or inferring the priorities of the partner in DealOrNoDeal negotiations. Building an accurate partner model is essential in social influence settings for guiding the decision-making of the system (Baarslag et al., 2013; Zhang et al., 2020b). Hence, we discuss various ways in which prior work tackles partner modeling.

Implicit: A majority of the efforts do not explicitly model the behavior of the partner but instead, this behavior implicitly guides the next response of the sequence-to-sequence dialogue system pipeline.

Simulated User: Lewis et al. (2017) trained a simulated user on the available data in a supervised manner. This was then used to further train the dialogue system. Instead of inferring mental states explicitly, this takes a more behavioral approach of estimating the future actions of the partner and using these for training via reinforcement learning.

Dialogue Act Look-Ahead: With a similar idea, Zhang et al. (2020b) proposed OPPA model with a *look-ahead* based partner modeling strategy at the level of DAs. At each step, OPPA first estimates the user's future DA, which is then used to select the next DA of the system. The authors found significant improvements on the DealOrNoDeal task. Yang et al. (2021) used a similar method

for buyer-seller negotiations. Taking a different approach, Chawla et al. (2022) trained a ranking model to directly predict the hidden preferences of the partner in a multi-issue negotiation. Instead of predicting future actions, these methods assume that the partner's behavior can be explained by their context and goals in the dialogue. However, this approach is yet to be used in an end-to-end system.

4.4 Training

Architecture Choices: One crucial aspect is the *architecture design*: End-to-end (Lewis et al., 2017; Radford et al., 2019) vs Modular (He et al., 2018). While end-to-end methods improve the diversity and need less manual effort, a modularized design enhances controllability and explainability. Perhaps, this is why modular methods are popular in large-scale models (Hadfi et al., 2021). Improving the control of desired variables such as topics, strategy, or emotion in the end-to-end methods is an open area of research and is yet to be explored for social influence dialogue systems.

Supervised Learning (SL) and Reinforcement Learning (RL): Zhou et al. (2019) used SL to train a hierarchical encoder-decoder for generating the next response and used Finite State Transducers (FSTs) to encode the historic sequence of DAs and persuasive strategies into the model, showing improvements in negotiation and persuasion tasks. The performance was later improved by Joshi et al. (2020), who replaced FSTs with Graph Neural Networks to better model the interdependencies. Others have relied on RL to explicitly optimize the model on task-specific objective outcomes. While SL trains the model to mimic the average human behavior, RL techniques, such as those based on REINFORCE (Williams, 1992), allow the system to explore its own strategies in the wild while being guided by one or more overall reward metrics. Lewis et al. (2017) used RL in negotiations, with the final points scored in the agreed deal as the reward. More recent work employed RL to incorporate simplistic partner models into the decision-making process of the dialogue system, showing improvements in negotiation tasks (Zhang et al., 2020b; Yang et al., 2021).

Multi-tasking and Pretraining: Limited efforts have also explored multi-tasking and pretrained language models for social influence dialogue systems, which provide promising ways to deal with the challenge of insufficient training data. Liu (2021)

trained a sequence-to-sequence transformer on a mix of Cornell Movie Dialogue corpus (Danescu-Niculescu-Mizil and Lee, 2011) and psychotherapy data. Li et al. (2020) fine-tuned the GPT model (Radford et al., 2018), while employing multi-tasking to incorporate intents and slots for both the human and the system. Wu et al. (2021) recently introduced ARDM which uses GPT2 (Radford et al., 2019) to separately encode the utterances of the human and the dialogue system, reducing the reliance on additional annotations.

5 Discussion and Recommendations

Past few years have seen an exciting progress in social influence dialogue systems. However, building sophisticated and practically useful systems remains a challenging endeavor. Several limitations still exist that must be addressed. To guide future work, we now discuss the key challenges and provide our recommendations.

Need for unifying the efforts: One challenge in this space has been the lack of large-scale datasets for model training. Social influence tasks are complex for crowdsourcing workers to understand and to participate in. Hence, prior work used extensive instructions and tutorials, making the study expensive and time consuming (Wang et al., 2019; Chawla et al., 2021b). To address this, we recommend the researchers to aim for a more unified view of the efforts in social influence.

First, this would encourage researchers to adopt the best practices from other social influence scenarios. For instance, most datasets miss out on user attributes like demographics and personality, which are crucial in social influence scenarios (Stuhlmacher and Walters, 1999; Bogaert et al., 2008). Most datasets also ignore the partner perception after the interaction is over. This can result in misleading conclusions about the model performance, where the models perform well objectively, but hurt the relationship with their partners, and thus, negatively impacting practical utility (Aydoğan et al., 2020).

Secondly, a holistic outlook will promote transfer learning and domain adaptation. Our taxonomy for datasets (Table 1) governs the way systems must be modeled and trained. Task structure is crucial to understand whether the model can learn from the utterances of all parties or just one. Further, understanding the context definition guides how it must be encoded. Hence, one interesting future

direction is joint training on datasets with similar structure and context definition.

Finally, progress in task-oriented and open-domain systems can inspire more unified modeling for social influence tasks involving multiple skills in the same interaction (e.g. a combination of negotiation and persuasion tactics as common in realistic scenarios). Roller et al. (2020) blend various open-domain tasks to address multiple challenges together (e.g., persona-based, knowledge-enriched, etc.). Hosseini-Asl et al. (2020) concatenate structured and unstructured data in task-oriented dialogues, and unify task-oriented dialogue system building to be a single sequence generation problem. Future work should explore similar unified approaches for social influence settings as well, especially since these tasks follow a common conceptual foundation (Figure 1), with similar evaluation and theoretical principles (Cialdini, 2009).

To encourage this unified view, we encapsulate our insights from this survey effort in a theoretical framework, which is presented in Appendix C. The framework covers key components for designing a social influence dialogue task, including system attributes, target audience, underlying modeling techniques, and evaluation mechanisms.

Theory integration: Most modeling efforts are based on crowdsourced datasets. Since crowdsourcing workers may not exhibit optimal strategies, supervised training on these datasets is fundamentally insufficient to build an effective system for applications like pedagogy (teaching social skills to students). Unfortunately, this holds regardless of how system strategy and partner model are designed. Further, using RL to optimize on objective rewards is also not expected to be enough to reliably learn complex influence capabilities, especially when the reward is restrictive.

To address this, we recommend to tap into the vast amount of research effort in social sciences and psychology on building theories for social influence (Cameron, 2009; Giles, 2016; Lewicki et al., 2016; Cialdini and Goldstein, 2004). Instead of solely relying on the collected data, future work should consider leveraging fundamentals from this research to guide the dialogue policy. Previous works have studied resistance to social influence (Knowles and Linn, 2004; Dal Cin et al., 2004; Petty and Cacioppo, 1977; Ahluwalia, 2000). Rucker et al. (2004) found that people resist persuasion differently depending on their beliefs, suggest-

ing personalizing the social influence process. One can also employ the politeness theory (Brown and Levinson, 1978) and model the participants' *face* acts to better understand users in social influence contexts (Dutt et al., 2020).

Task Evaluation: Another key limitation of existing work is the lack of a comprehensive evaluation. Prior work majorly focused on objective metrics which only provides a limited view of the model performance. A comprehensive evaluation is challenging since it must consider partner perception along with objective outcomes. Building user simulators could potentially alleviate this problem (Li et al., 2016; Jain et al., 2018; Shi et al., 2019). Most existing simulators are developed for task-oriented systems which follow a certain agenda. Future research should study how to use partner modeling to build social influence user simulators for more efficient and accurate task evaluation (He et al., 2018; Yang et al., 2020). For instance, one could potentially design different user personalities and simulate the change in user's beliefs, opinions, and attitudes accordingly (Yang et al., 2021).

Multimodal systems: Being a core function of human communication, social influence occurs not just through text, but through all possible modalities. Schulman and Bickmore (2009) showed that embodied agents achieve better persuasion results than text-only agents. Other studies have recognized the importance of emotion in social influence tasks (Asai et al., 2020; Chawla et al., 2021a). Nguyen et al. (2021) proposed a speech dataset in debates and study the influence of spoken tactics on persuasiveness across genders. Given these findings, we encourage interdisciplinary efforts in the future to explore the developement of multimodal social influence agents.

Knowledge-enriched systems: Social influence tasks often involve constantly-changing world knowledge such as organization facts and news. Often, the system's internal state (e.g., the change of task setting from one set of products to a different set) needs to be updated. Retraining the entire system is costly to maintain after the initial development. Recent work has proposed to augment the dialogue system with internet-search ability to generate more factual and updated responses in open-domain dialogues (Komeili et al., 2021). Future efforts in this direction will benefit social influence dialogue systems as well.

6   Conclusions

We introduced the category of social influence dialogue systems that aim to influence their partners through dialogue. We presented a survey of the recent prior work in this space, compiling datasets and methods across diverse application domains. We pointed out key limitations in existing methodologies and proposed promising directions for designing more sophisticated systems in the future. Our survey reveals that although substantial progress has been made, this is still an emerging research area. We hope our work inspires more dedicated interdisciplinary effort and discussion, which is necessary for making progress in this space.

7   Broader Impact and Ethical Considerations

Social influence is ubiquitous in everyday life. Research on how we use influence in all aspects of our lives spans a number of fields, including social psychology, communication, consumer behavior, behavioral change, and behavioral economics. This research has led to crucial findings about the strategies of social influence and how they impact our decision-making. Over the past few decades, research has accumulated and demonstrated the effectiveness of using various strategies across contexts and domains. Prominent examples include core principles of social influence by Cialdini from social psychology: reciprocity, commitment and consistency, social proof, liking and attractiveness, authority, and scarcity (Cialdini, 2009). Further, communication strategies used in persuasion and general social influence contexts include credibility appeals, two-sided argumentation, emotional tactics, and appeals to social norms, among others (Cameron, 2009; O'keefe, 2015).

First, the well-studied principles in social influence research can guide the development of effective dialogue systems with influence capabilities. In fact, many of the strategies found in the datasets developed for social influence tasks (Section 3) directly map to the principles laid out by Cialdini, for instance, credibility and emotional appeal in PersuasionForGood dataset (Wang et al., 2019) and reciprocity observed in CaSiNo negotiation dataset (Chawla et al., 2021b). Second, research in social influence dialogue systems provides novel datasets on human-human and human-machine communication, and therefore, holds a great potential to advance theories of human cogni-

tion and influence processes (Gratch et al., 2015). The datasets and subsequent analyses can further contribute new theoretical insights to social influence research.

Although dialogue systems have already been used in a number of applications involving chatbots and AI assistants, advancements in social influence dialogue systems can help to bridge the gap between our existing task definitions and a number of other real-world applications. For instance, realistic customer support interactions often involve active behaviors from both the support agent and the user where the agent uses social cues for improved customer satisfaction and retention, while the user attempts to address their queries. These settings naturally involve aspects of social influence, unlike traditional task-oriented definitions where the dialogue system plays a passive role to assist the human users. As discussed earlier, social influence dialogue systems can positively help to advance other areas as well. In therapy domain, these systems can assist in various psychological treatments such as by increasing the willingness to disclose (Lucas et al., 2014). In pedagogy, they can help to make social skills training more accessible (Johnson et al., 2019).

While we think about these applications, it is crucial to also lay out proper ethical guidelines to avoid any misuse of these systems. Primary concerns are around the use of deception (e.g. in Diplomacy and other negotiation tasks), emotional appeals (e.g. in persuasion), and behavior change (e.g. in conversational recommendations).

To mitigate possible misuse scenarios or unintended harms, we now lay out a few ethical guidelines which also apply to dialogue research in general. First, rigorous attempts must be made to ensure that the data collection, design processes, and evaluations, strictly abide by the guidelines and regulations laid out by the relevant Institutional Review Board (IRB). Second, the research team needs to develop a thorough plan to monitor and understand the behaviors of the developed systems before deployment. This includes identifying the goals of the dialogue system, identifying potential toxic language use, and any discriminatory behaviors. Third, investment into improved data collection practices, along with explainable and controllable dialogue systems can help identify these issues early on and allow manipulation to avoid them. Fourth, we argue that transparency is the key.

All stakeholders must be made aware of the goals and design objectives of the system, along with any known misbehaviors or potential risks. The users must also be informed of any data collected during the deployment phase. Lastly, we believe that continuous monitoring of dialogue systems is necessary to ensure that the system performs consistently and does not diverge to unexpected conditions that may incur offensive or discriminative actions. We hope that our work promotes a more systematic study of social influence dialogue systems, which in turn will help to tackle the ethical concerns in a more principled way.

8 Limitations

Literature Search: We presented a survey of efforts in social influence dialogue systems. Although every attempt was made to provide the readers with a comprehensive overview of the research in this space, our work does not claim exhaustiveness in the covered literature and it is likely that we missed out on other relevant research in this space.
Intention for influence: The datasets and tasks covered in this literature review focus on scenarios where social influence is *intentional by design*. However, social influence can also be *unintentional*, that is, interactions between humans and machines can have unintended influence on the attitudes, behaviors, or feelings of the human user (Gass, 2015). For instance, changes in topic preferences after interacting with a system on a variety of topics, or incorporating biases after interacting with a biased system. As we continue to make an unprecedented progress towards AI systems that interact with humans via natural means of communication, we must also take into account the unintended influence on the users of the underlying technology. We hope that our work motivates researchers to study these effects methodically in the future.

Acknowledgments

References

Rohini Ahluwalia. 2000. Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research*, 27(2):217–232.

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, 4:463–476.

Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 12–21.

Sara Asai, Koichiro Yoshino, Seitaro Shinagawa, Sakriani Sakti, and Satoshi Nakamura. 2020. Emotional speech corpus for persuasive dialogue system. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 491–497.

Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the stac corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727.

Reyhan Aydoğan, Tim Baarslag, Katsuhide Fujita, Johnathan Mell, Jonathan Gratch, Dave de Jonge, Yasser Mohammad, Shinji Nakadai, Satoshi Morinaga, Hirotaka Osawa, et al. 2020. Challenges and main results of the automated negotiating agents competition (anac) 2019. In *Multi-agent systems and agreement technologies*, pages 366–381. Springer.

Tim Baarslag, Mark Hendrikx, Koen Hindriks, and Catholijn Jonker. 2013. Predicting the performance of opponent models in automated negotiation. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 2, pages 59–66. IEEE.

Linda Babcock and Sara Laschever. 2009. *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Sandy Bogaert, Christophe Boone, and Carolyn Declerck. 2008. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, 47(3):453–480.

Maria Boritchev and Maxime Amblard. 2021. Ding–a corpus of transcriptions of real-life, oral, spontaneous multi-party dialogues between french-speaking players of catan. In *Journées du GdR LIFT*.

Penelope Brown and Stephen C Levinson. 1978. Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction*, pages 56–311. Cambridge University Press.

Kenzie A Cameron. 2009. A practitioner's guide to persuasion: An overview of 15 selected persuasion theories, models and frameworks. *Patient education and counseling*, 74(3):309–317.

Shelly Chaiken. 1987. The heuristic model of persuasion. In *Social influence: the ontario symposium*, volume 5, pages 3–39.

Kushal Chawla, Rene Clever, Jaysa Ramirez, Gale Lucas, and Jonathan Gratch. 2021a. Towards emotion-aware agents for negotiation dialogues. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE.

Kushal Chawla, Gale Lucas, Jonathan May, and Jonathan Gratch. 2020. Exploring early prediction of buyer-seller negotiation outcomes. *arXiv preprint arXiv:2004.02363*.

Kushal Chawla, Gale M Lucas, Jonathan May, and Jonathan Gratch. 2022. Opponent modeling in negotiation dialogues by related data adaptation. *arXiv preprint arXiv:2205.00344*.

Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale Lucas, Jonathan May, and Jonathan Gratch. 2021b. Casino: A corpus of campsite negotiation dialogues for automatic negotiation systems. *arXiv preprint arXiv:2103.15721*.

R. B. Cialdini. 2009. *Influence: Science and Practice*. fifth ed. Pearson/Allyn & Bacon, Boston, MA.

Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annual review of psychology*, 55(1):591–621.

Cristian D.-N.-M., Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pages 699–708.

Sonya Dal Cin, Mark P Zanna, and Geoffrey T Fong. 2004. Narrative persuasion and overcoming resistance. *Resistance and persuasion*, 2:175–191.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 76–87.

Orianna Demasi, Marti A Hearst, and Benjamin Recht. 2019. Towards augmenting crisis counselor training by improving message retrieval. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 1–11.

Jan Deriu, Alvaro Rodrigo, Arantxa Otegi, Guillermo Echegoyen, Sophie Rosset, Eneko Agirre, and Mark Cieliebak. 2021. Survey on evaluation methods for dialogue systems. *Artificial Intelligence Review*, 54(1):755–810.

David DeVault, Johnathan Mell, and Jonathan Gratch. 2015. Toward natural turn-taking in a virtual human negotiation agent. In *AAAI Spring Symposia*. Citeseer.

Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander H Miller, Arthur Szlam, and Jason Weston. 2016. Evaluating prerequisite qualities for learning end-to-end dialog systems. In *4th International Conference on Learning Representations, ICLR 2016*.

Sebastian Duerr and Peter A Gloor. 2021. Persuasive natural language generation–a literature review. *arXiv preprint arXiv:2101.05786*.

Esin Durmus and Claire Cardie. 2019. A corpus for modeling user and language effects in argumentation on online debating. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Ritam Dutt, Rishabh Joshi, and Carolyn Penstein Rosé. 2020. Keeping up appearances: Computational modeling of face acts in persuasion oriented discussions. *arXiv preprint arXiv:2009.10815*.

Subhabrata Dutta, Jeevesh Juneja, Dipankar Das, and Tanmoy Chakraborty. 2022. Can unsupervised knowledge transfer from social discussions help argument mining? *arXiv preprint arXiv:2203.12881*.

Ryo Egawa, Gaku Morio, and Katsuhide Fujita. 2019. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 422–428.

Chaim Fershtman. 1990. The importance of the agenda in bargaining. *Games and Economic Behavior*, 2(3):224–238.

Tommaso Fornaciari and Massimo Poesio. 2012. Decour: a corpus of deceptive statements in italian courts. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1585–1590.

Jamie Fraser, Ioannis Papaioannou, and Oliver Lemon. 2018. Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 179–184.

Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural approaches to conversational ai. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1371–1374.

Robert H Gass. 2015. Social influence, sociology of. In *International Encyclopedia of the Social and Behavioral Sciences*, pages 348–354. Elsevier.

Howard Giles. 2016. *Communication accommodation theory: Negotiating personal relationships and social identities across contexts*. Cambridge University Press.

Jonathan Gratch, David DeVault, Gale M Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents*, pages 201–215. Springer.

Rafik Hadfi, Jawad Haqbeen, Sofia Sahab, and Takayuki Ito. 2021. Argumentative conversational agents for online discussions. *Journal of Systems Science and Systems Engineering*, 30(4):450–464.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.

Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. 2020. A simple language model for task-oriented dialogue. *Advances in Neural Information Processing Systems*, 33:20179–20191.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems (TOIS)*, 38(3):1–32.

Alankar Jain, Florian Pecune, Yoichi Matsuyama, and Justine Cassell. 2018. A user simulator architecture for socially-aware conversational agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 133–140.

Jiun-Hao Jhan, Chao-Peng Liu, Shyh-Kang Jeng, and Hung-Yi Lee. 2021. Cheerbots: Chatbots toward empathy and emotionusing reinforcement learning. *arXiv preprint arXiv:2110.03949*.

Changzhen Ji, Xin Zhou, Yating Zhang, Xiaozhong Liu, Changlong Sun, Conghui Zhu, and Tiejun Zhao. 2020. Cross copy network for dialogue generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1900–1910.

Emmanuel Johnson, Gale Lucas, Peter Kim, and Jonathan Gratch. 2019. Intelligent tutoring system for negotiation skills training. In *International Conference on Artificial Intelligence in Education*, pages 122–127. Springer.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2020. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In *International Conference on Learning Representations*.

Sung Soo Kim, Stan Kaplowitz, and Mark V Johnston. 2004. The effects of physician empathy on patient satisfaction and compliance. *Evaluation & the health professions*, 27(3):237–251.

Eric S Knowles and Jay A Linn. 2004. *Resistance and persuasion*. Psychology Press.

Mojtaba Komeili, Kurt Shuster, and Jason Weston. 2021. Internet-augmented dialogue generation. *arXiv preprint arXiv:2107.07566*.

Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–27.

Roy J Lewicki, Bruce Barry, and David M Saunders. 2016. *Essentials of negotiation*. McGraw-Hill.

Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning of negotiation dialogues. In *EMNLP*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. *Advances in neural information processing systems*, 31.

Xinmeng Li, Wansen Wu, Long Qin, and Quanjun Yin. 2021. How to evaluate your dialogue models: A review of approaches. *arXiv preprint arXiv:2108.01369*.

Xiujun Li, Zachary C Lipton, Bhuwan Dhingra, Lihong Li, Jianfeng Gao, and Yun-Nung Chen. 2016. A user simulator for task-completion dialogues. *arXiv preprint arXiv:1612.05688*.

Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. 2020. End-to-end trainable non-collaborative dialog system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8293–8302.

Kai-Hui Liang, Patrick L Lange, Yoo Jung Oh, Jingwen Zhang, Yoshimi Fukuoka, and Zhou Yu. 2021. Evaluation of in-person counseling strategies to develop physical activity chatbot for women. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 32–44.

Wang Liao, Jingwen Zhang, Yoo Jung Oh, and Nicholas A Palomares. 2021. Linguistic accommodation enhances compliance to charity donation: The role of interpersonal communication processes in mediated compliance-gaining conversations. *Journal of Computer-Mediated Communication*, 26(3):167–185.

Houjun Liu. 2021. Towards automated psychotherapy via language modeling. *arXiv preprint arXiv:2104.10661*.

Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.

Huyen Nguyen, Ralph Vente, David Lupea, Sarah Ita Levitan, and Julia Hirschberg. 2021. Acoustic-prosodic, lexical and demographic cues to persuasiveness in competitive debate speeches. *Proc. Interspeech 2021*, pages 1034–1038.

Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, Vinay Adiga, and Erik Cambria. 2021. Recent advances in deep learning based dialogue systems: A systematic survey. *arXiv preprint arXiv:2105.04387*.

Tim Norfolk, Kamal Birdi, and Deirdre Walsh. 2007. The role of empathy in establishing rapport in the consultation: a new model. *Medical education*, 41(7):690–697.

Daniel J O'keefe. 2015. *Persuasion: Theory and research*. Sage Publications.

Siru Ouyang, Zhuosheng Zhang, and Hai Zhao. 2021. Dialogue graph modeling for conversational machine reading. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3158–3169.

Denis Peskov, Benny Cheng, Ahmed Elgohary, Joe Barrow, Cristian Danescu-Niculescu-Mizil, and Jordan Boyd-Graber. 2020. It takes two to lie: One to lie, and one to listen. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3811–3854.

Richard E Petty and John T Cacioppo. 1977. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and social Psychology*, 35(9):645.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381.

Lisa Rashotte. 2007. Social influence. *The Blackwell encyclopedia of sociology*.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M Smith, et al. 2020. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*.

Derek D Rucker, Zakary L Tormala, and Richard E Petty. 2004. Individual differences in resistance to persuasion: The role of beliefs and meta-beliefs. *Resistance and persuasion*, page 83.

Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology*, pages 1–8.

Weiyan Shi, Kun Qian, Xuewei Wang, and Zhou Yu. 2019. How to build user simulators to train rl-based dialog systems. *arXiv preprint arXiv:1909.01388*.

Zhouxing Shi and Minlie Huang. 2019. A deep sequential model for discourse parsing on multi-party dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7007–7014.

Manjira Sinha and Tirthankar Dasgupta. 2021. Predicting success of a persuasion through joint modeling of utterance categorization. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3423–3427.

Alice F Stuhlmacher and Amy E Walters. 1999. Gender differences in negotiation outcome: A meta-analysis. *Personnel Psychology*, 52(3):653–677.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624.

Michael Tanana, Kevin A Hallgren, Zac E Imel, David C Atkins, and Vivek Srikumar. 2016. A comparison of natural language processing methods for automated coding of motivational interviewing. *Journal of substance abuse treatment*, 65:43–50.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5624–5634.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327–335.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256.

Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2021. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1292–1301.

William Wunderle. 2007. How to negotiate in the middle east. *Military review*, 87(2):33.

Atsuki Yamaguchi, Kosui Iwasa, and Katsuhide Fujita. 2021. Dialogue act-based breakdown detection in negotiation dialogues. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 745–757.

Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2020. Improving dialog systems for negotiation with personality modeling. *arXiv preprint arXiv:2010.09954*.

Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2021. Improving dialog systems for negotiation with personality modeling. 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 681–693.

Denis Yarats and Mike Lewis. 2018. Hierarchical text generation and planning for strategic dialogue. In *International Conference on Machine Learning*, pages 5591–5599. PMLR.

Min-Hsuan Yeh and Lun-Wei Ku. 2021. Lying through one's teeth: A study on verbal leakage cues. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4504–4510.

Dian Yu, Yulia Tyshchuk, Heng Ji, and William Wallace. 2015. Detecting deceptive groups using conversations and network analysis. In *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL-IJCNLP 2015*, pages 857–866. Association for Computational Linguistics (ACL).

Jingwen Zhang, Yoo Jung Oh, Patrick Lange, Zhou Yu, Yoshimi Fukuoka, et al. 2020a. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet. *Journal of medical Internet research*, 22(9):e22845.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213.

Zheng Zhang, Lizi Liao, Xiaoyan Zhu, Tat-Seng Chua, Zitao Liu, Yan Huang, and Minlie Huang. 2020b. Learning goal-oriented dialogue policy with opposite agent awareness. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 122–132.

Zheng Zhang, Ryuichi Takanobu, Qi Zhu, MinLie Huang, and XiaoYan Zhu. 2020c. Recent advances and challenges in task-oriented dialog systems. *Science China Technological Sciences*, pages 1–17.

Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. 2019. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *International Conference on Learning Representations*.

A Literature Compilation

In this section, we provide details about how the literature was curated for our survey. We hope this helps the overall reproducibility and also guides similar studies in the future. We followed a simple two-stage process. First, we compiled the relevant datasets that capture various forms of social influence across diverse domains (presented in Section 3) and then, we compiled the techniques developed on these datasets (presented in Section 4).

Step I - Datasets: Our objective was to gather datasets that (by design) capture forms of social influence. We primarily focused on dialogue interactions but include the datasets based on transcripts from multimodal interactions as well. Given the large breadth of research in this space across a number of domains, our collection is not exhaustive but is rather restricted to the following sources.

We surveyed the past 6 years of *ACL conference proceedings. We then covered several online repositories of dialogue data to capture datasets published at other venues. This includes ParlAI[1], Huggingface[2], NLP-Progress[3], and Convokit[4]. Further, we revisited several recent surveys in dialogue systems and Natural Language Generation (NLG) research (Huang et al., 2020; Zhang et al., 2020c; Ni et al., 2021; Duerr and Gloor, 2021). Datasets that were categorized as task-oriented or open-domain in these surveys but also contain some aspects of social influence have been included in our discussion. As discussed in Section 4, we also include the datasets that have not been directly used for designing dialogue systems but rather for various Natural Language Understanding (NLU) subtasks that can be crucial for the eventual development of dialogue systems in this space. Finally, we also reviewed the citation graphs of the collected papers from Google Scholar. Overall, we ended up with 22 dataset papers, spanning 12 publication venues, 4 languages, and 7 application domains.

Step II - Methods: Compiling the methodological progress was based on the models developed on the curated datasets. For this purpose, we simply reviewed the citations of all the dataset papers using Google Scholar.

B Datasets

A comprehensive list of the available datasets for investigating social influence in dialogues is provided in Table 3. For each dataset, we mention the application domain, source, key statistics, as well as the available metadata and annotations apart from the conversation logs.

C Five Stages for Designing Social Influence Dialogue Systems

We develop a five-stage framework to summarize our recommendations for future work. These stages cover key decisions in the design of a dialogue system in this space, encouraging a holistic understanding of the system characteristics, target audience, underlying modeling techniques, and evaluation mechanisms. These steps are inspired by a behavior change model in healthcare research (Zhang et al., 2020a). We adapt this model to make it suitable for general social influence tasks in NLP. We present these steps in Figure 2.

---

[1] https://github.com/facebookresearch/ParlAI
[2] https://huggingface.co/docs/datasets/index
[3] http://nlpprogress.com/english/dialogue.html
[4] https://convokit.cornell.edu/documentation/datasets.html

Name (Citation)	Domain	Source	Key Statistics	Metadata & Annotations
STAC (Asher et al., 2016)	Games	Crowdsource	Dialogues: 1081 Turns/Dialogue: 8.5 Tokens/Turn: 4.2	Dialogue Acts; Discourse Structures
Diplomacy (Peskov et al., 2020)	Games	Crowdsource	Games: 12 Messages/Game: 1440.8 Words/Message: 20.79[+]	Intended and perceived truthfulness; Participant demographics and self-assessment of lying abilities; Ground-truth betrayals
DinG (Boritchev and Amblard, 2021)	Games	University game night logs	Games: 10 Turns/Game: 2357.5	Annotated question-answer pairs
Tabletop (DeVault et al., 2015)	MIBT	Face-to-face, Wizard-of-Oz	Face-to-face Dialogues: 89 Wizard-of-Oz Dialogues: 30	Participant demographics; Subjective questionnaire-based assessment
DealOrNoDeal (Lewis et al., 2017)	MIBT	Crowdsource	Dialogues: 5808 Turns/Dialogue: 6.6 Tokens/Turn: 7.6	—
CaSiNo (Chawla et al., 2021b)	MIBT	Crowdsource	Dialogues: 1030 Utterances/Dialogue: 11.6 Tokens/Utterance: 22	Participant demographics and personality traits; Outcome satisfaction; Partner perception; Strategy Annotations
JobInterview (Yamaguchi et al., 2021)	MIBT	Crowdsource	Dialogues: 2639 Turns/Dialogue: 12.7 Words/Turn: 6.1	Dialogue acts
PersuasionforGood (Wang et al., 2019)	Social Good	Crowdsource	Dialogues: 1017 Turns/Dialogue: 10.4 Words/Turn: 19.4	Participant sociodemographics, personality, and engagement in the conversation; Strategy annotations; Dialogue Acts
CraigslistBargain (He et al., 2018)	E-commerce	Crowdsource	Dialogues: 6682 Turns/Dialogue: 9.2 Tokens/Turn: 15.5	Dialogue Acts
AntiScam (Li et al., 2020)	E-commerce	Crowdsource	Dialogues: 220 Turns/Dialogue: 12.4 Words/Turn: 11.1	Dialogue Acts; Semantic Slots
Motivational Interviewing (Tanana et al., 2016)	Therapy & Support	Psychotherapy session logs	Sessions: 341 Utterances/Session: 513.2 Words/Utterance: 9.7	Behavior annotations
SMS Counseling (Althoff et al., 2016)	Therapy & Support	SMS chat logs	Dialogues: 80,885 Messages/Dialogue: 42.6* Words/message: 19.2*	Post-conversation assessment for both the counselor (e.g. suicide risk, main issue etc.) and user (how they feel afterwards)
EmpatheticDialogues (Rashkin et al., 2019)	Therapy & Support	Crowdsource	Dialogues: 24,850 Utterances/Dialogue: 4.3 Words/Utterance: 15.2	—
Hotline Counseling (Demasi et al., 2019)	Therapy & Support	Synthetic Transcripts	Dialogues: 254 Messages/Dialogue: 40-60	Paraphrases by trained counselors
mPED (Liang et al., 2021)	Therapy & Support	Physical activity clinical trials	Sessions: 107 Turns/Session: 423.5 Minutes/Session: 28.8	Demographics; Physical activity related pre and post surveys; Strategy annotations
Congressional Debates (Thomas et al., 2006)	Argumentation	U.S. Congressional transcripts	Debates: 53 Speech segments/Debate: 72.8	Ground-truth label with each speech segment for support/oppose of the proposed bill
Supreme Court (D.-N.-M. et al., 2012)	Argumentation	Oyez.org transcripts	Cases: 7700 Utterances/Case: 220.8	Case-related metadata such as key dates, citation, parties involved, and voting results
DeCour (Fornaciari and Poesio, 2012)	Argumentation	Italian court hearings	Hearings: 35 Utterances/Hearing: 173.4 Tokens/Utterance: 13.9[#]	Metadata for testimonies like place, date, demographics; Hearing related details; Truthfulness annotations
ChangeMyView (Tan et al., 2016)	Argumentation	Reddit	Discussion Trees: 20,626 Nodes/Tree: 61.1	The original post with initial arguments and explicitly recognized successful persuasive arguments from the opposing side
DDO Debates (Durmus and Cardie, 2019)	Argumentation	debate.org logs	Debates: 78,376 Messages/Debate: 7.7	User demographics; Debate metadata like dates and category; Audience votes and comments
Court Debates (Ji et al., 2020)	Argumentation	China Court transcripts	Dialogues: 260,190 Utterances/Dialogue: 13.9	—
Target-Guided (Tang et al., 2019)	Miscellaneous	Crowdsource	Dialogues: 9939 Utterances/Dialogue: 11.4	—

Table 3: Social Influence Dialogue Corpora, grouped by task domains and sorted by publishing year within a domain. All statistics of the form X/Y denote average numbers. MIBT: Multi-Issue Bargaining Task. *Only computed for dialogues with additional survey responses, [+]Only computed for the training data. [#]Only for Speaker utterances in front of the judge (doesn't include other members of the court). Note that not all datasets listed above have been directly used for designing end-to-end dialogue systems, but instead, these have enabled research into various sub-tasks and analyses that can eventually be useful for dialogue systems in this area. Please refer to Section 3 in the main paper for a detailed discussion about these datasets and to Section 4 for information about various methods developed using them.

I. Task Specifications

- **Task objectives**
- **Task Structure**
  - Multi-issue bargaining like DealOrNoDeal and CaSiNo (*Symmetric*)
  - Effective persuader in PersuasionForGood for motivating users to donate to a charity (*Asymmetric*)
- **Context Definition**
  - Therapy interactions like MI and SMS Counseling (*Global*)
  - Argumentation scenarios such as debates and court proceedings (*Local*)

III. Chatbot Capacity

Natural Language Understanding (NLU)

- Implicit: Representation Learning
- Explicit: Eg. Dialog Acts (Intents + Slots like offer details or price amount), Strategy Usage, Emotion Expression

Dialogue Management (DM)

- Persuasive strategy planning
- Task-specific strategies: negotiation tactics, emotional appeal, logical appeals, credibility appeal
- Social Rapport building: Eg. with small talk, empathy
- Knowledge Incorporation

Natural Language Generation (NLG)

- Exhibit desired strategies, social capabilities, and emotional expression

Shared Modules

- Long-term Memory: Relevant for long-term repeated interactions such as therapy and repeated negotiations
- Controls: Providing knobs to control agent behavior

IV. Chatbot Design & Techniques

Model Architecture

- Agenda-based
- Modularized (separate NLU, DM, and NLG components)
- End-to-end

Learning Techniques

- Supervised learning
- Reinforcement Learning - Optimizing on outcome-based rewards
- Transfer Learning - Unifying similar tasks

V. Evaluation Mechanisms

Task-agnostic (Linguistic Performance)

- Automatic: Eg. Perplexity, BERTScore, BLEU
- Human: Eg. Engagement, Coherency, Consistency, Fluency

Task-specific (Outcome and Perception)

- Donation Amount, the final price of the negotiated deal, pedagogical improvements, behavior change, etc.
- User satisfaction, partner perception, etc.

II. Chatbot Characteristics and User Backgrounds

Chatbot (Designed)
- Cooperative vs Non-cooperative personality
- Anthropomorphic traits (Eg. Identity, Name, Gender)

Users (Self-reported or inferred)
- Demographics (Eg. Age, Gender, Education, Ethnicity)
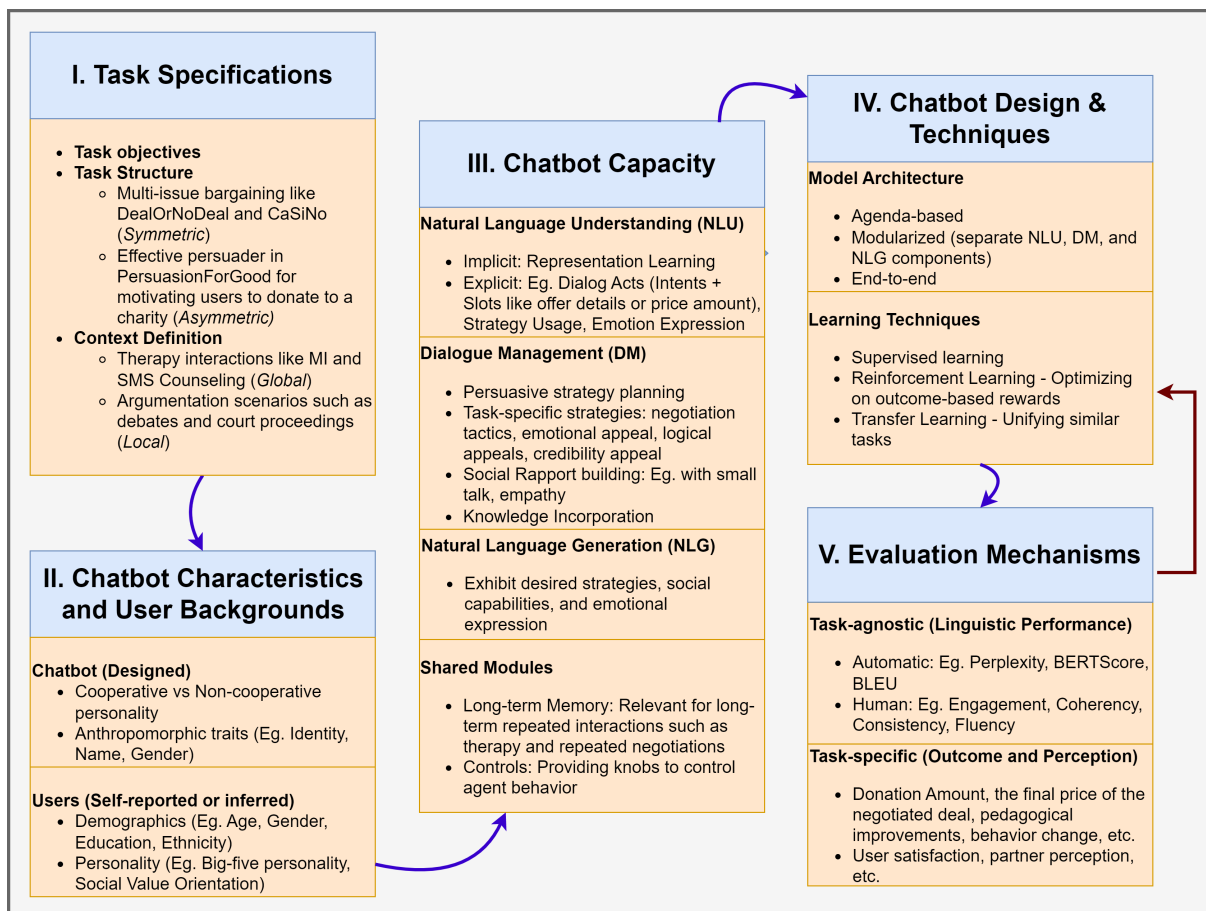- Personality (Eg. Big-five personality, Social Value Orientation)

Figure 2: A theoretical model for the development of dialogue systems for social influence tasks. Curved arrows represent forward relations and the straight arrow represents the feedback. **I. Task Specifications**: Key properties that define the task in consideration and are captured by the collected dataset, **II. Chatbot Characteristics and User Backgrounds**: Attributes for the agent design and target audience, **III. Chatbot Capacity**: The desirable capabilities of the system, **IV. Chatbot Design & Techniques**: The modeling techniques to develop the dialogue system, and **V. Evaluation Mechanisms**: Metrics to evaluate system performance.