Social Learning through Interactions with Other Agents: A Survey

Dylan Hillier[1,2] , **Cheston Tan**[1] and **Jing Jiang**[2]

[1]Centre for Frontier AI Research, Institute of High Performance Computing, A*STAR
[2]Singapore Management University
das.hillier.2023@phdcs.smu.edu.sg, cheston_tan@cfar.a-star.edu.sg, jingjiang@smu.edu.sg

Abstract

Social learning plays an important role in the development of human intelligence. As children, we imitate our parents' speech patterns until we are able to produce sounds; we learn from them praising us and scolding us; and as adults, we learn by working with others. In this work, we survey the degree to which this paradigm – *social learning* – has been mirrored in machine learning. In particular, since learning socially requires interacting with others, we are interested in how *embodied agents* can and have utilised these techniques. This is especially in light of the degree to which recent advances in natural language processing (NLP) enable us to perform new forms of social learning. We look at how behavioural cloning and next-token prediction mirror human imitation, how learning from human feedback mirrors human education, and how we can go further to enable fully communicative agents that learn from each other. We find that while individual social learning techniques have been used successfully, there has been little unifying work showing how to bring them together into socially embodied agents.

1 Introduction

Social learning is "learning that is influenced by the observation of another individual or their products" [Bonnefon *et al.*, 2023]. The "embodiment hypothesis" in cognitive science suggests that intelligence emerges as a result of embodiment within a social environment, that "starting as a baby grounded in a physical, social, and linguistic world is crucial to the development of the flexible and inventive intelligence that characterizes humankind" [Smith and Gasser, 2005]. We can define this embodiment as being when "the environment is acting upon the individual and the individual is acting upon the environment" [Bolotta and Dumas, 2022]. We explore how social embodiment – learning that is influenced by the actions and products of other agents in the learning environment – has played a role in the development of AI, and how it could influence the development of embodied agents in the real world.

Despite often not being recognised as such, a large proportion of AI research can be considered or framed as employing aspects of social learning. In particular, models trained on human-produced data often implicitly rely on learning to imitate humans. Take for example language model pretraining, in which the model is trained to directly reproduce human text. The form of this social learning can have important implications, for example through the reproduction of human biases in textual data [Bender *et al.*, 2021] and facial recognition [Buolamwini and Gebru, 2018]. As such, the prevalence and impact of AI systems trained in this way warrants a more thorough investigation of the mechanics of social learning.

From another perspective, social learning techniques have largely seen success on supervised offline learning tasks with fixed datasets. Real-world deployment often requires accounting for more complicated, two-way interactions between agent and environment. This has traditionally been done through individualistic reinforcement learning methods, which have issues like sample inefficiency, unspecified reward functions, and poor explainability [Dulac-Arnold *et al.*, 2021]. In contrast, humans – who develop in a socially embodied setting – are able to quickly adapt to changes in the environment. Since communication between agents has been made easier by the development of LLMs, it is an apt time to research the development of agents using social learning.

1.1 Social Learning in Humans

Social learning plays an instrumental role in the way humans learn [Tomasello, 2009]. We do not learn individualistically by simply observing the world – instead, we take advantage of the fact that other humans have learnt before us and have built up cultural knowledge. Indeed, *Homo Sapiens'* exploitation of cultural knowledge is exactly what has allowed us to build complex societies and dominate the globe. Rather than having bigger brains (and thereby having a cognitive advantage), our ability to "accumulate information across generations and develop well-adapted tools, beliefs, and practices that no individual could invent on their own" is responsible for our success [Boyd *et al.*, 2011].

Similarly, the transmission and generation of this cultural knowledge is of paramount importance in human development at both an individual and societal level. Social learning enables this cultural transmission and development through our interactions with others. This suggests to us that the
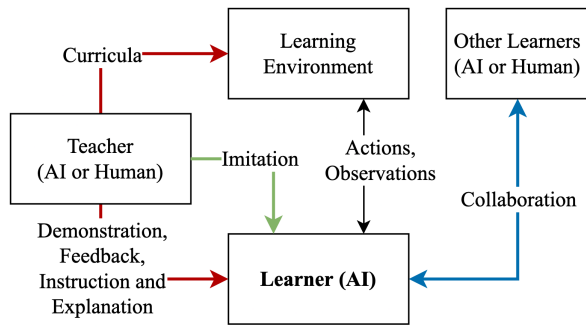
Figure 1: The relationships among different social learning techniques, agents and the learning environment. Note: *Imitation* is initiated by the learner rather than the teacher. As in Figure 2, the learning approaches are coloured according to their classification in Tomasello's account of Social Learning.

Figure 2: The sections of this survey structured in relation to Tomasello's classification of Social Learning.

promising route to achieving more advanced agents lies not just in building bigger artificial brains (scaling) but also in building artificial brains socially, that can learn from and teach other agents, and take advantage of and develop this rich cultural knowledge base [Bolotta and Dumas, 2022].

According to Tomasello's influential theory of Social Learning, it can be divided into Imitative Learning, Instructive Learning, and Collaborative Learning [Tomasello *et al.*, 1993]. The organisation of this survey follows this division, as depicted in Figure 2.

Imitative Learning. In Tomasello's theory of Social Learning, children learn by copying adults in a number of ways. A distinction is drawn between *emulation* and *true imitation*. In the case of the former, the child attempts to reproduce the actions and outcomes of the adult demonstrator; in the latter, the child additionally models the intent behind the action and learns from that. Such *true imitation* seems to require at the least a nascent theory of mind, insofar as it requires modelling the intents of others.

Instructive Learning. Instructive learning refers to the process by which an adult or expert learner tries to teach a non-expert learner. Other than explicit instruction, this process can include feedback, explanations, demonstrations and construction of a curriculum. Collectively, the process by which these mechanisms are employed to aid a learner is referred to as scaffolding [Van de Pol *et al.*, 2010].

Collaborative Learning. Collaborative learning is the setting in which neither or none of the collaborators are experts. In these cases, the learners must work together to progress in the learning task. From a developmental perspective, this emerges during early education – in contrast to imitation learning which is observed almost from birth [Tomasello *et al.*, 1993]. However, it plays an important role in childhood development, especially in the development of language outside of their asymmetric interactions with adult teachers. For example, Vygotsky and Cole [1978] emphasise the role of childhood play in enabling peers to practice social roles within their capabilities.
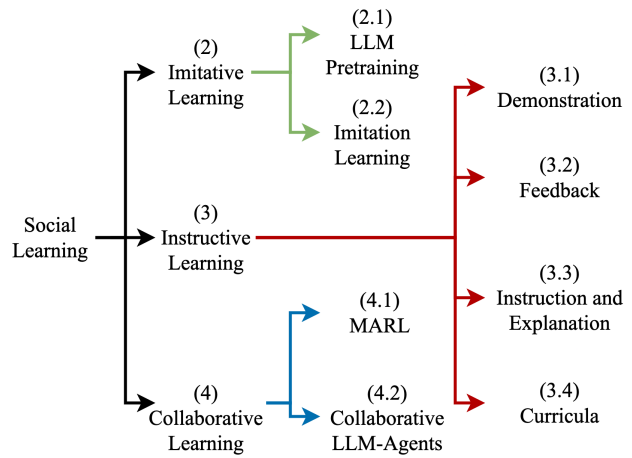
1.2 Social Learning in AI

At first glance, some forms of social learning techniques have been widely used in AI at large [Duéñez-Guzmán *et al.*, 2023]. Diving deeper, we seek to make sense of these from a unified perspective, and examine whether there are more aspects of social learning that can be applied to AI.

In the simplest case, training data is generated or curated by humans. Furthermore, many models are trained to generate more of this human-supplied data – this often consitutes a form of imitation learning. For example, a generative image model may be trained to imitate the art or photography of humans. While such imitative models are powerful, they are also inherently limited to recreating the level of the source of their learning [Yiu *et al.*, 2023]. Several instructive techniques like feedback [Ouyang *et al.*, 2022] and curricula [Bengio *et al.*, 2009] are also widely used in various domains. Finally, collaborative learning regimes have been explored in the Multi Agent Reinforcement Learning (MARL) literature [Gronauer and Diepold, 2022]. All these techniques and their relations are depicted in Figure 1.

1.3 LLM Agents

The development of LLMs poses a striking opportunity to enable embodied agents to learn socially. Indeed, early experiments with creating embodied agents almost entirely from LLMs have shown that such agents are able to achieve remarkable zero-shot performance on embodied tasks [Wang *et al.*, 2024]. This has been boosted by the development of Vision Language Models (VLMs) and other multimodal models that have been used to create agents that can naturally leverage human demonstration data and cultural knowledge [Brohan *et al.*, 2023; Wang *et al.*, 2023]. In addition, since they are language models, such agents possess an inherent ability to communicate with each other, without the need to develop such methods from scratch. This has accordingly enabled the creation of cooperative groups of agents [Chen *et al.*, 2024b].

1.4 Objective of This Survey

There has not been a recent survey on social learning covering the developments arising from large language models (LLMs). Meanwhile, a growing number of works have called for more exploration of social learning approaches [Bonnefon *et al.*, 2023; Duéñez-Guzmán *et al.*, 2023; Bolotta and Dumas, 2022; Yiu *et al.*, 2023]. In this survey, we aim to provide a review of recent works that use social learning techniques with a view to building agents that can learn just through interacting with others. Given the relative maturity of imitation techniques for training generative models, in this survey we put more focus on the potential for training agents using instructive and collaborative forms of social learning, as this has been most directly enabled by the emergence of LLM Agents.

As depicted in Figure 2, we first consider existing approaches to imitation learning, both in the wider machine learning literature, as well as in their application to building embodied agents. Relatedly, we briefly consider some of the work that has explored how to enable this imitation learning in Section 3.1. In Section 3.2, we then extensively explore how feedback has been used for aligning LLMs, and how richer forms of feedback might be leveraged. Many of these lessons are also applicable to Explanations and Instructions as briefly covered in Section 3.3. We then explore the degree to which curricula are useful for these embodied model given some negative results in Section 3.4. We cover how MARL has traditionally been used for collaborative learning in Section 4.1 Finally, we cover the recent development of collaborative LLM-agents in Section 4.2.

More generally, by carefully considering the gaps between human social learning paradigms and current techniques used in training embodied agents we hope to inspire research that more fully utilises the advantages offered by social and cultural learning.

2 Imitative Learning

2.1 LLM Pretraining

Large language models are largely trained on the task of *next token prediction* – essentially predicting the next word in a sentence. This is arguably a form of imitative learning in so far as the model is fed observations of the productions of a human agent (the sequence of tokens), and learns to autoregressively predict that sequence of tokens. As indicated in Section 1.3, such models have achieved remarkable feats, to the point of matching humans on a wide number of tasks.

Nevertheless this acquisition process is highly data hungry – GPT-3 was trained on over 200 billion tokens, whereas the average 12 year old would have only seen roughly 100 million [Warstadt *et al.*, 2023]. One potential explanation for this is that the techniques used for training LLMs do not consist in *true imitation* in the sense described by Tomasello et al. [1993] and rather perform "emulation". However, Kosinski [2023] argues that LLMs already possess theory of mind, although this position has been criticised [Yiu *et al.*, 2023].

More pointedly, during the training regime, a large language model is devoid of the socially-embodied context that typically accompanies a child learner, such as being able to build up a model of the social environment that may be useful for prediction. In this context, language models have been shown to recreate human biases, reproduce a static view in the face of changing social mores, and lack accountability [Bender *et al.*, 2021].

We have already seen how childhood development relies on other learning regimes [Tomasello *et al.*, 1993], therefore what remains to be seen is the development of models that can actually leverage this socially embodied setting to further learn and adapt through their collaboration with other agents. Indeed, this has begun to be explored through building social simulation environments [Liu *et al.*, 2023]. Hopefully this will enable us to move beyond LLMs just acting as "tools of imitation" [Yiu *et al.*, 2023] and mitigate some of the social issues mentioned above.

2.2 Imitation Learning

The field of Imitation learning itself (and indeed its dual – instruction by demonstrations) has been a popular reward-free method of training embodied agents. We do not aim to provide a comprehensive overview of the field, as covered by e.g. [Hussein *et al.*, 2017; Zheng *et al.*, 2024]. Key examples of techniques in the field include:

- **Behavioural Cloning** – in behavioural cloning, direct access to the underlying policy of the teacher agent is assumed, and used as a training signal.

- **Inverse Reinforcement Learning (IRL)** – in IRL, the reward function is generated from the teacher's policy and then used to train the student model by existing reinforcement learning algorithms.

- **Observational Imitation Learning (OIL)** – Rather than assuming direct access to the teacher's policy, in observational imitation learning, the learner agent only has access to external observations of the teacher agent acting in the learning environment.

These imitation learning techniques been widely used in the development of embodied agents; For instance RT2 [Brohan *et al.*, 2023], an embodied foundation model for robotics, was created by collating and training on a dataset made from a mix of human demonstrations and question answering web data. This model was then pretrained through OIL (in the form of next token prediction). Similarly, AlphaStar [Vinyals *et al.*, 2019] was also pretrained from imitating games of professional players, and Ramrakhya et al. [2022] utilise an embodied object-search strategy learnt from human demonstrations and robot trajectories.

Section Summary: Imitative Learning techniques are widely used, including in RL and NLP. There is nevertheless valuable work to be done exploring ways to increase the sample efficiency of these techniques, and whether socially embodied settings may play a role in this.

3 Instructive Learning

3.1 Demonstration

Demonstrations as an instructive learning paradigm consist largely in the provision of examples that a student agent can

imitate. Importantly this means that the teacher may impart additional intent onto the demonstration through the choice and mode of presentation of those examples. For example a human teacher may draw attention to particular aspects of a demonstrated process by narrating it. Indeed human demonstrations are typically accompanied by explanations and narrations. Furthermore Ramrakhya et al. [2022] verify that using imitation learning techniques on $70k$ human demonstrations outperforms using $240k$ agent trajectories, indicating that the quality of demonstrations can have a large impact. We have seen relatively little research exploring these elements of demonstration learning. Sumers et al. [2023] conduct a study on the provision of demonstrations to humans and find that if the demonstrations are not accompanied by explanations, they are less effective than linguistic interactions between teacher-student pairs for conveying concepts.

Bhoopchand et al. [2023] do explore some of the mechanisms at play with demonstrations, especially in the light of how they can be performed to maximise cultural transmission. They explore a navigation task with demonstrations performed by humans, that are then dropped out. Naive imitation (or rather emulation) of the experts enables the agent to perform well when the expert is there to follow, but once the expert is removed, these agents fail to generalise. The authors are able to create an agent that can imitate the plan of the teacher through equipping the agent with 1) memory, 2) an attentional bias towards the expert, 3) the expert "dropping out" during training, and 4) randomisation of the domain to ensure that the student agent is learning to utilise the expert and not just memorise the paths. This "attentional bias" may be related to the notion of intent as an aspect of "true imitation" [Tomasello *et al.*, 1993] and further accords well with Tomasello's [2009] theory of joint attention.

3.2 Feedback

Feedback is just information given by the teacher on the performance of the learner. Consider the example of a student learning to perform addition. If they perform an incorrect addition like 18+17=25 we could provide a simple label evaluation (incorrect), a preferred answer (18+17=35), or language feedback like "you forgot to carry over the tens term" – the particular form of feedback that is appropriate will depend on the situation.

Labels are of course omnipresent in machine learning, as supervised learning forms a major paradigm. Learning from preferences has also been developed especially in the field of RL, where *Reinforcement Learning from Human Feedback* (RLHF) has now been adapted for finetuning language models [Ouyang *et al.*, 2022]. While we mostly consider linguistic feedback, human feedback is often multimodal in nature, and may include gestures, tone, expressions and other physical actions. We encourage readers to refer to [Lin *et al.*, 2020] for progress in these dimensions; we expect these to become increasingly feasible and important as more works appear exploring multimodal models.

Preference-Based Feedback

Human preference-based feedback on generative AI output has proven invaluable for aligning the outputs of foundation models with human expectations through the use of 'instruction tuning' [Ouyang *et al.*, 2022]. This finetuning technique was designed for the purposes of *alignment* – the process of optimising a model according to human values and needs (e.g. following instructions). Typically this employs a three step process: 1) collecting demonstration data and performing supervised finetuning; 2) training a reward model on comparisons between model outputs; and 3) optimising the model as a policy against that reward model using an RL algorithm (PPO). The success of this approach has played an instrumental role in the recent rise of LLMs by providing a scalable and effective way to finetune models and control specific attributes.

Despite this, there have been a number of issues due to the cost of collecting human feedback data, and difficulty in optimising using PPO. One interesting development of late is Direct Preference Optimisation (DPO) [Rafailov *et al.*, 2024] – which skips training the reward model and uses the LLM as an implicit reward model. This is essentially a KL-constrained contrastive loss that trains the model directly on preference datasets. Hejna et al. [2024] point out that this method is not applicable to embodied trajectories and introduce a generalised version called "contrastive preference optimisation", wherein a dis-preffered trajectory is contrasted to a preferred one.

These developments should enable us to train embodied agents built with multimodal language models (e.g. [Brohan *et al.*, 2023]) using preference optimisation and bring this aspect of feedback into embodied agents.

Natural Language-Based Feedback

While it is relatively simple to use comparisons or scalar feedback scores, fully leveraging the richness of language feedback has proven more of a challenge.

Sumers et al. [2021] investigate learning under the regime of "unconstrained linguistic feedback". They categorise the feedback given as evaluative – scalar feedback similar to that described above, imperative – giving a preferred action e.g. "you should have carried over the tens term", or descriptive – providing relevant information to the task. Sumers' approach to leveraging these rich rewards is to provide models that try to model the sentiment (a simple sentiment analysis model and a more complicated model handcrafted from Gricean pragmatics), as well as an inference network trained to predict the rewards given by the teacher. In their experiments, they firstly validate the importance of the more complex models that can leverage the richer language feedback, secondly they show that human teaching typically involves a mix of all three categories of feedback in human experiments.

More recently, several works have exploited the recent advances in NLP to use LLMs to directly update the reward functions used for training embodied models in accordance with the feedback. For example in Eureka [Ma *et al.*, 2023] the authors train a robotic hand to execute complicated gestural object manipulation (spinning a pen around) by updating a reward function from observations. Similarly, Yu et al. [2023] use LLMs to define reward parameters that are then optimised over for a variety of robot embodiments including a quadruped and arm based robot. They find that this approach

is generalisable and achieves 90% success rates across a wide variety of tasks. The import of such approaches is that they enable the utilisation of LLM's high-level cultural knowledge about actions to be translated into low-level actions.

A promising alternative solution to this is to use the feedback to condition future actions of the learning agent, and rely on existing learning mechanisms to learn from these actions. Scheurer et al. [2022] use this paradigm to learn from natural language feedback by generating multiple responses conditioned on feedback. The response that is semantically most similar to the feedback is then treated as the "correct" response and used to finetune the model. Similarly, Chen et al. [2024a] explore "Imitation learning from Language Feedback" in which they let the learner refine its answers based on human feedback, and then finetune on these refinements.

Zha et al. [2023] further explore the mechanisms of memorising and retrieving feedback for conditioning embodied agent actions on the fly. They create a system for distilling corrections into retrievable knowledge that can be used by the LLM-agent in future generations. They use the feedback to firstly perform replanning, and then to create a correction that can be placed in a knowledge base (rephrased to be "contextually complete") and retrieved in future situations where relevant. In particular, they use both visual and textual representations for the correction embedding so that it can be retrieved when a situation is similar in either dimension.

3.3 Instruction and Explanation

Instruction following itself has been widely used as a task in the alignment of models. In particular, the RLHF framework described in Section 3.2 was popularised in NLP for instruction following [Ouyang *et al.*, 2022]. This alignment process (learning to follow instructions) has enabled foundation models to learn to perform complex tasks from instructions. A large body of works explore methods of instruction in the form of prompting techniques like chain of thought [Wei *et al.*, 2022], that have been shown to improve reasoning abilities. For example, Mu et al. [2024] finetune their embodied model on complicated instructions that require the use of chain of thought planning over egocentric videos.

Instruction following has also seen use as a paradigm in the popular benchmarks TEACh [Padmakumar *et al.*, 2022] and ALFRED [Shridhar *et al.*, 2020]. TEACh consists in human-human dialogue sessions where one human is playing the role of a teacher or commander; the other acts as a student and must follow the instructions of the teacher, while being able to ask questions for clarification. ALFRED is a relatively simple task and involves following singular grounded instructions like "walk to the coffee maker on the right". The latter instruction can be seen as an example of the vision-language navigation task, which is currently one of the more popular paradigms in embodied AI, given that it was recently enabled by increased performance of LLMs. These benchmarks however do not focus on actually *learning* from the interactions with the instructors per se, nor indeed are the instructions *targeted* at the learning agent.

Saha et al. [2024] explore more deeply modelling interactions between students and teachers using theory of mind. One of their research questions is the degree to which person-alisations of explanations aid students. While they do find a positive effect, its influence is relatively small. While personalisation of explanations have not been found to be beneficial thus far, in an embodied RL setting it was found that the act of giving explanations itself was helpful in aiding the teacher learning [Das *et al.*, 2023]. This is an example of the so called *protégé effect* – when we try to explain or instruct others we test and expand our own understanding of a concept by reconceptualising it.

While it would be desirable to heavily use targeted instructions and explanations by asking for help when needed, the provision of help, especially in the case of humans in the loop, is expensive. Indeed per the theory of scaffolding, we expect that as the learner gains proficiency, the learning experience can be less structured, and intervention can be reduced over time. As such one body of works considers the problem of efficiency by learning when to ask for help. For example Ren et al. [2023] train an agent to ask for help when there is a high level of uncertainty in the task. The agent generates a set of possible actions, and if the likelihood of the likeliest of them being correct (as measured by the agent) is below a certain threshold, the robot then asks for further instruction.

3.4 Curricula

Curriculum Learning is a machine learning approach that parallels the scaffolding of human learning at a more abstract level. The classic paradigm introduced by Bengio et al. [2009] is to order the training data by complexity for some notion of complexity, and progressively increase the level of complexity as the learning progresses. Per Figure 1, curriculum learning consists in the teacher agent altering the learning environment rather than directly interacting with the learner agent.

The idea behind this is that by structuring the data in this way, the initial learning task becomes more feasible, and after basic concepts are acquired, these can be used as a basis for learning more complex concepts. While this seems simple, there are two key problems. First of all defining complexity for a given task is difficult – the curriculum may have to be tailored for individual learners, and require insight into the current capabilities of the learner. Additionally it may be unclear when the learner has sufficiently mastered a concept to enable the complexity to increase.

Accordingly a body of works exists that try to automate this process. This approach is typically called automatic curriculum learning, and is a popular paradigm in the reinforcement learning community (ACL). For instance, Matiisen et al. [2019] explore a "Teacher Student Curriculum Learning" paradigm in which a teacher agent monitors the student's training progress and is used to determine the tasks suitable for the student. The mechanism used is the student's progress on a task – the teacher increases sample weight for tasks that are seeing progress – the weight is then reduced when the task has been mastered since the rate of progress is measurably decreased. The authors treat the problem as a bandit learning problem and explore a number of sampling algorithms, finding that such an automated curriculum can help for maze solving in minecraft, and simple addition tasks using LSTMs.

Other works such as Ramrakhya et al. [2023] and Morad

et al. [2021] use estimations of task success probabilities to order the experience while training embodied navigation systems. In particular, they aim to find tasks of 'intermediate difficulty'. Interestingly, in the former they note that the addition of easy tasks to the training mix without curriculum learning per se has a similarly beneficial effect, as the learning agent simply spends less time in easier episodes since it can achieve them quickly.

Despite their success in RL domains, curriculum-based approaches have not seen as much success in NLP or computer vision. Recently Warstadt et al. [2023] ran a competition to train LLMs using a human-developmentally plausible dataset and found that although curriculums were the most popular approach explored by entrants, it was relatively ineffective for improving sample efficiency. Indeed the key negative result of their competition report was exactly that none of the common curriculum learning approaches offered substantial benefits to the developmental task of language learning. Given the differing success of curriculum learning approaches it is likely that it's benefit is largely dependent on how structured the learning problem is.

Indeed *embodied* language models may be a more suitable domain for the application of curriculum learning methods. Wang et al. [2024] leverage the LLM's knowledge of task properties to derive new suitable tasks. Rather than relying on handcrafted curricula or designing a separate automatic curriculum, they simply prompt the LLM to come up with tasks itself, conditioning the new tasks on the state of the agent, previously learnt skills, the previous task completion history, and the long term goals of the agent. They find that this automatic curriculum is both critical for the agent's success, and furthermore beats out manually defined expert curricula. This process of giving the agent meta-learning abilities is promising but likely requires more principled analysis; in any case it speaks to the potential of using LLM-Agents to instantiate social learning paradigms in a unified way.

Section Summary: More principled work remains to be done on demonstration learning, such as investigating whether ToM can enable student agents to imitate more efficiently. Feedback has been widely used in the form of preference learning. Leveraging general language feedback has seen increasing use but requires more research. Approaches like conditioning on feedback likely have similar applications to leveraging explanations and instructions. Finally, while curriculum learning has been widely used in RL, it is not as promising for training LLMs – this may not however be the case for LLM-Agents.

4 Collaborative Learning

4.1 Multi-Agent Reinforcement Learning (MARL)

Historically the majority of research on collaborative learning has occurred in the context of multi-agent reinforcement learning. For instance AlphaStar [Vinyals *et al.*, 2019] was trained to grandmaster human level in the game StarCraft II through playing with other players. The system was initialised using imitation learning, and then further finetuned by an RL algorithm to perform well against a mixture of opponents. The solution proposed by the authors is an evolution of previous so called "self-play" strategies. This consists in pitting different versions of an agent against itself in strategy games. The key insight here was to firstly add a greater mix of adversarial agents (previous versions of the learning agent) in addition to agents that were specifically designed to exploit the flaws of the agent in question.

Recently, an AI agent (Cicero) was trained for the game Diplomacy at a human level, ranking in the top 10% of participants in an online league [FAIR *et al.*, 2022]. In Diplomacy, players take the role of diplomats controlling countries, and have to negotiate with other players, make promises, and deceive them. By leveraging pretrained LLMs to conduct the negotiation and model other users' intents, a strategic reasoning model could then be used to plan actions and intents and was then trained with a combination of RL methods including self-play and behavioural cloning.

In the more cooperative setting, a number of recent works have focused on the "learn to communicate" paradigm [Oroojlooy and Hajinezhad, 2023]. In these paradigms, the sets of agents may have to learn both what and when to communicate, and even some explorations of who to communicate to via attention mechanisms [Gronauer and Diepold, 2022]. Many of these works have started from a point of zero linguistic competence and require the agents to learn from scratch. Accordingly a future trend will likely include the incorporation of existing language models [Gronauer and Diepold, 2022], similarly to the approach taken in the work of FAIR et al. [2022] described above.

4.2 Collaborative LLM-Agents

A significant recent trend in NLP has been exploring the use of communicative LLM-agents. Firstly, Park et al. [2023] explored using such generative agents as models of human behaviour – in particular, they simulated a town of agents that perform daily activities and interact in natural language, enabled by a game environment that tracks the memories and positions of each agent. These behaviours, including emergent abilities at party planning, were rated as being largely believable as human by crowdworkers.

Later, Qian et al. [2023] explored the creation of software using teams of LLMs mimicking the 'waterfall' model of software development. The authors broke down the development process into a sequence of discussions between LLM agents acting in roles (e.g. CEO, programmer etc.). In each discussion the LLMs would act in turn as instructors and assistants, making decisions through concensus. In contrast to simply generating the codebase for a software development project directly using an LLM designed for coding, they managed to achieve an $\sim88\%$ success rate on achieving the software development task.

Chen et al. [2024b] then went further to provide a more general framework for communicative LLM based agents. The process proceeds in rounds made up of 1) expert recruitment – building up a group of roles to perform, 2) decision making – deliberating over what to do 3) action execution, and 4) evaluation in which the outcome is compared to the goal and feedback is given.

In the embodied setting, Zhang et al. [2024] have explored building cooperative embodied LLM agents. This report-

edly requires more complicated models and theory of mind - in particular their agents consist in modules handling belief, communication, reasoning, and planning. The belief module is intended to give the agents theory of mind by tracking the knowledge of the scene, other agents, the self, and the task. On benchmarks, the authors found that the LLM agents were able to cooperate well on various benchmarks, and that the theory of mind was necessary for this success.

Such cooperative LLM agents are already able to outperform state-of-the-art MARL agents on several tasks, such as navigation [Li *et al.*, 2023], without the extensive number of training episodes typically required for the MARL agents. Importantly however, in all these collaborative settings, no learning per se is going on – the agents are able to achieve complicated tasks through collaboration but only "learn" by updating the memory or context used in their internal decision making processes (since they are frozen language models).

More generally, we expect that collaborative settings are more useful for generating experience trajectories that would be unachievable for singular agents which can then be leveraged for learning using ordinary learning mechanisms, rather than being a source for learning mechanisms per se, in a similar way to Scheurer et al. [2022] exploring conditioning on feedback as a way of learning from natural language feedback rather than providing a separate feedback learning mechanism per se.

Section Summary: Collaborative Learning has been explored mostly through Multi Agent Reinforcement Learning. These have included interesting studies on optimising the population of collaborators and emergent communication. In the case of LLM-Agents, there have been remarkable achievements through applying prompting, including application of ToM, but thus far limited work directly updating the weights of the models constituting these agents, limiting their adaptability.

5 Conclusions and Future Directions

Although social learning is not typically conceived as a key paradigm or organising principle for machine learning, we have shown that it permeates a large proportion of the machine learning literature.

Moreover, we have seen how performant LLMs have enabled the application of more advanced social learning techniques using the preferred mode of human instruction – natural language. For example, in Section 3.2 we showed that this has enabled the construction of models that can directly respond to language feedback, and in Section 4.2 how they had been used to build communicative agents purely from language models that were able to outperform existing MARL approaches.

Comparison and Combination of Social Learning Paradigms. As shown in Table 1, the majority of approaches explored in this paper use multiple social learning techniques. However, only four use all three of the main categories and are largely not embodied. As we have argued, equipping agents with linguistic capabilities has enabled agents to leverage communication and instructive learning

Reference	Iml	Inl	Col	Emb	ToM
[Belkhale *et al.*, 2024]	Y	F,I	–	Y	–
[Chen *et al.*, 2024b]	–	I,F	Y	–	–
[Chen *et al.*, 2024a]	Y	I,F	–	–	–
[Hejna *et al.*, 2024]	–	F	–	Y	–
[Mu *et al.*, 2024]	Y	I	–	Y	–
[Rafailov *et al.*, 2024]	–	F	–	–	–
[Saha *et al.*, 2024]	Y	D,F	Y	–	Y
[Wang *et al.*, 2024]	–	C,F	–	Y	–
[Bhoopchand *et al.*, 2023]	Y	D	–	Y	–
[Das *et al.*, 2023]	–	I	Y	Y	–
[Liu *et al.*, 2023]	Y	F,I	Y	–	–
[Li *et al.*, 2023]	–	F	Y	–	Y
[Ma *et al.*, 2023]	–	F	–	Y	–
[Park *et al.*, 2023]	–		Y	Y	Y
[Qian *et al.*, 2023]	–	F,I	–	Y	–
[Ramrakhya *et al.*, 2023]	–	C	–	Y	–
[Ren *et al.*, 2023]	–	F	Y	Y	–
[Wang *et al.*, 2023]	–	I,F	–	Y	–
[Yu *et al.*, 2023]	–	F	–	Y	–
[Zha *et al.*, 2023]	–	F,I	–	Y	–
[Brohan *et al.*, 2023]	Y	D	–	Y	–
[FAIR *et al.*, 2022]	Y		Y	Y	Y
[Ouyang *et al.*, 2022]	Y	I,F	–	–	–
[Padmakumar *et al.*, 2022]	–	I	Y	Y	–
[Ramrakhya *et al.*, 2022]	Y	D	–	Y	–
[Scheurer *et al.*, 2022]	Y	F,I	Y	–	–
[Morad *et al.*, 2021]	–	C	–	Y	–
[Sumers *et al.*, 2021]	–	F	–	Y	–
[Shridhar *et al.*, 2020]	–	I	Y	Y	–
[Vinyals *et al.*, 2019]	Y	F,D	Y	Y	–
[Matiisen *et al.*, 2019]	–	C	–	Y	–
[Bengio *et al.*, 2009]	–	C	–	–	–

Table 1: Social Learning techniques used: ImL: Imitative Learning, InL: Instructional Learning, CoL: Collaborative Learning, Emb: Embodied, ToM: Theory of Mind, C: Curriculum, D: Demonstrations, F: Feedback, I: Instruction, Y: Yes.

paradigms out of the box. Given this, we expect to see more work exploring the combination of these approaches.

Learning Permanently through Social Conditioning. The promising LLM-Agent approaches have thus far largely relied on the knowledge acquired by the LLM through imitation learning. There has been little work performing further gradient updates once these models are socially embodied, and the "learning" has largely constituted updates to the agent's memory modules (as in [Wang *et al.*, 2024; Zha *et al.*, 2023]). Accordingly, identifying how models can learn more permanently in these settings remains a research priority. We see a promising direction in utilising low-level learning methods on experiences conditioned on the social interactions that entail "higher-level" learning mechanisms like feedback and instruction. This has seen some application in the non-embodied setting [Scheurer *et al.*, 2022], but we expect this to be replicated for embodied settings in the near future as we move beyond simply using LLM-agents without

finetuning.

Multimodality. Our survey has been primarily concerned with linguistic approaches, however the recent development of multimodal models will likely enable richer forms of social interaction (and thereby learning). We have already seen this in the case of demonstration learning from vision and language [Brohan *et al.*, 2023], but given the relative immaturity of video-foundation models it may be some time before we are able to fully leverage non-linguistic signals like gestures and visual demonstrations for teaching embodied agents. As discussed in Section 3.2, existing techniques for learning from feedback could be applied to learn from these signals.

Theory of Mind and Intent. As can be seen in Table 1, there have been a number of approaches exploring the use of Theory of Mind as a prerequisite for richer interactions between teachers, students, and peers. Indeed Theory of Mind cuts across social learning. Modelling the intent of a teacher enables a learner agent to fully leverage the learning signal, as we discuss in Section 3.1. Modelling teacher agents also allows learners to evaluate the teacher's potential as sources of learning [Gweon, 2021]. While ToM modules are a relatively common aspect of agents, especially collaborative agents, we have seen little work exploring these techniques so far. As such, we expect to see more work investigating the effects of these modules, especially given the interest in ToM initiated by Kosinski [2023].

Application Domains. Social learning techniques are broadly applicable in the field at large. Nevertheless, they may play an especially important role in domains where social skills themselves are more relevant. For instance, in the case of socially-assistive robotics, the ability to adapt to individual needs and build a relationship over time is especially important [Tapus *et al.*, 2007]. Similarly, the ability to use models to teach humans skills would be a significant benefit of developing instructive social learning techniques. For example, this has explored in the case of social skill training [Yang *et al.*, 2024]. In the meantime, we expect the majority of progress to occur in using social-learning techniques for training agents in video games, due to the low cost of training, ample amounts of training data, and ease of interacting with humans. Games are already a testbed for many approaches covered above [Wang *et al.*, 2024; Wang *et al.*, 2023; Vinyals *et al.*, 2019], but we can expect social learning to enable richer forms of collaborative play with AI agents, as explored for example in the work of Gong et al. [2023].

Acknowledgments

References

[Belkhale *et al.*, 2024] Suneel Belkhale, Tianli Ding, Ted Xiao, Pierre Sermanet, Quon Vuong, Jonathan Tompson, Yevgen Chebotar, Debidatta Dwibedi, and Dorsa Sadigh. Rt-h: Action hierarchies using language. *arXiv preprint arXiv:2403.01823*, 2024.

[Bender *et al.*, 2021] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[Bengio *et al.*, 2009] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 41–48, 2009.

[Bhoopchand *et al.*, 2023] Avishkar Bhoopchand, Bethanie Brownfield, Adrian Collister, Agustin Dal Lago, Ashley Edwards, Richard Everett, Alexandre Fréchette, Yanko Gitahy Oliveira, Edward Hughes, Kory W Mathewson, et al. Learning few-shot imitation as cultural transmission. *Nature Communications*, 14(1):7536, 2023.

[Bolotta and Dumas, 2022] Samuele Bolotta and Guillaume Dumas. Social neuro AI: Social interaction as the "dark matter" of AI. *Frontiers in Computer Science*, 4:846440, 2022.

[Bonnefon *et al.*, 2023] Jean-François Bonnefon, Maxime Derex, Alberto Acerbi, Thomas Griffiths, et al. Machine culture. *Nature Human Behaviour*, 7(11):1855–1868, 2023.

[Boyd *et al.*, 2011] Robert Boyd, Peter J Richerson, and Joseph Henrich. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(supplement_2):10918–10925, 2011.

[Brohan *et al.*, 2023] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, pages 2165–2183. PMLR, 2023.

[Buolamwini and Gebru, 2018] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

[Chen *et al.*, 2024a] Angelica Chen, Jérémy Scheurer, Jon Ander Campos, Tomasz Korbak, Jun Shern Chan, Samuel R. Bowman, Kyunghyun Cho, and Ethan Perez. Learning from natural language feedback. *Transactions on Machine Learning Research*, 2024.

[Chen *et al.*, 2024b] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, et al. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. In *The Twelfth International Conference on Learning Representations*, 2024.

[Das *et al.*, 2023] Devleena Das, Sonia Chernova, and Been Kim. State2explanation: Concept-based explanations to benefit agent learning and user understanding. *Advances in Neural Information Processing Systems*, 36:67156–67182, 2023.

[Duéñez-Guzmán *et al.*, 2023] Edgar A Duéñez-Guzmán, Suzanne Sadedin, Jane X Wang, Kevin R McKee, and Joel Z Leibo. A social path to human-like artificial intelligence. *Nature Machine Intelligence*, 5(11):1181–1188, 2023.

[Dulac-Arnold *et al.*, 2021] Gabriel Dulac-Arnold, Nir Levine, Daniel J Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9):2419–2468, 2021.

[FAIR *et al.*, 2022] Meta Fundamental AI Research Diplomacy Team FAIR, Anton Bakhtin, Noam Brown, Emily Dinan, et al. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

[Gong *et al.*, 2023] Ran Gong, Qiuyuan Huang, Xiaojian Ma, Hoi Vo, Zane Durante, Yusuke Noda, Zilong Zheng, Song-Chun Zhu, Demetri Terzopoulos, Li Fei-Fei, et al. Mindagent: Emergent gaming interaction. *arXiv preprint arXiv:2309.09971*, 2023.

[Gronauer and Diepold, 2022] Sven Gronauer and Klaus Diepold. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, pages 1–49, 2022.

[Gweon, 2021] Hyowon Gweon. Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, 25(10):896–910, 2021.

[Hejna *et al.*, 2024] Joey Hejna, Rafael Rafailov, Harshit Sikchi, Chelsea Finn, Scott Niekum, W. Bradley Knox, and Dorsa Sadigh. Contrastive preference learning: Learning from human feedback without reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024.

[Hussein *et al.*, 2017] Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan, and Chrisina Jayne. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.

[Kosinski, 2023] Michal Kosinski. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*, 2023.

[Li *et al.*, 2023] Huao Li, Yu Chong, Simon Stepputtis, Joseph P Campbell, et al. Theory of mind for multi-agent collaboration via large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 180–192, 2023.

[Lin *et al.*, 2020] Jinying Lin, Zhen Ma, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. A review on interactive reinforcement learning from human social feedback. *IEEE Access*, 8:120757–120765, 2020.

[Liu *et al.*, 2023] Ruibo Liu, Ruixin Yang, Chenyan Jia, Ge Zhang, Diyi Yang, and Soroush Vosoughi. Training socially aligned language models on simulated social interactions. In *The Twelfth International Conference on Learning Representations*, 2023.

[Ma *et al.*, 2023] Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, et al. Eureka: Human-level reward design via coding large language models. In *Second Agent Learning in Open-Endedness Workshop*, 2023.

[Matiisen *et al.*, 2019] Tambet Matiisen, Avital Oliver, Taco Cohen, and John Schulman. Teacher–student curriculum learning. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3732–3740, 2019.

[Morad *et al.*, 2021] Steven D Morad, Roberto Mecca, Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. Embodied visual navigation with automatic curriculum learning in real environments. *IEEE Robotics and Automation Letters*, 6(2):683–690, 2021.

[Mu *et al.*, 2024] Yao Mu, Qinglong Zhang, Mengkang Hu, Wenhai Wang, Mingyu Ding, Jun Jin, Bin Wang, Jifeng Dai, et al. Embodiedgpt: Vision-language pre-training via embodied chain of thought. *Advances in Neural Information Processing Systems*, 36, 2024.

[Oroojlooy and Hajinezhad, 2023] Afshin Oroojlooy and Davood Hajinezhad. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11):13677–13722, 2023.

[Ouyang *et al.*, 2022] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[Padmakumar *et al.*, 2022] Aishwarya Padmakumar, Jesse Thomason, Ayush Shrivastava, Patrick Lange, Anjali Narayan-Chen, Spandana Gella, Robinson Piramuthu, Gokhan Tur, et al. Teach: Task-driven embodied agents that chat. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2):2017–2025, Jun. 2022.

[Park *et al.*, 2023] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, et al. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023.

[Qian *et al.*, 2023] Chen Qian, Xin Cong, Cheng Yang, Weize Chen, et al. Communicative agents for software development. *arXiv preprint arXiv:2307.07924*, 2023.

[Rafailov *et al.*, 2024] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

[Ramrakhya *et al.*, 2022] Ram Ramrakhya, Eric Undersander, Dhruv Batra, and Abhishek Das. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5173–5183, 2022.

[Ramrakhya *et al.*, 2023] Ram Ramrakhya, Dhruv Batra, Aniruddha Kembhavi, and Luca Weihs. Curriculum learning via task selection for embodied navigation. *CVPR 2023 Embodied AI Workshop*, 2023.

[Ren *et al.*, 2023] Allen Z Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, et al. Robots that ask for help: Uncertainty alignment for large language model planners. In *Conference on Robot Learning*, pages 661–682. PMLR, 2023.

[Saha *et al.*, 2024] Swarnadeep Saha, Peter Hase, and Mohit Bansal. Can language models teach? teacher explanations improve student performance via personalization. *Advances in Neural Information Processing Systems*, 36, 2024.

[Scheurer *et al.*, 2022] Jérémy Scheurer, Jon Ander Campos, Jun Shern Chan, Angelica Chen, Kyunghyun Cho, and Ethan Perez. Learning from natural language feedback. In *ACL Workshop on Learning with Natural Language Supervision*, 2022.

[Shridhar *et al.*, 2020] Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, et al. ALFRED: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10740–10749, 2020.

[Smith and Gasser, 2005] Linda Smith and Michael Gasser. The development of embodied cognition: Six lessons from babies. *Artificial Life*, 11(1-2):13–29, 2005.

[Sumers *et al.*, 2021] Theodore R. Sumers, Mark K. Ho, Robert D. Hawkins, Karthik Narasimhan, and Thomas L. Griffiths. Learning rewards from linguistic feedback. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(7):6002–6010, May 2021.

[Sumers *et al.*, 2023] Theodore R Sumers, Mark K Ho, Robert D Hawkins, and Thomas L Griffiths. Show or tell? exploring when (and why) teaching with language outperforms demonstration. *Cognition*, 232:105326, 2023.

[Tapus *et al.*, 2007] Adriana Tapus, Maja J Mataric, and Brian Scassellati. Socially assistive robotics [grand challenges of robotics]. *IEEE robotics & automation magazine*, 14(1):35–42, 2007.

[Tomasello *et al.*, 1993] Michael Tomasello, Ann Cale Kruger, and Hilary Horn Ratner. Cultural learning. *Behavioral and Brain Sciences*, 16(3):495–511, 1993.

[Tomasello, 2009] Michael Tomasello. *The cultural origins of human cognition*. Harvard University Press, 2009.

[Van de Pol *et al.*, 2010] Janneke Van de Pol, Monique Volman, and Jos Beishuizen. Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22:271–296, 2010.

[Vinyals *et al.*, 2019] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[Vygotsky and Cole, 1978] Lev Semenovich Vygotsky and Michael Cole. *Mind in society: Development of higher psychological processes*. Harvard university press, 1978.

[Wang *et al.*, 2023] Zihao Wang, Shaofei Cai, Anji Liu, Xiaojian Ma, and Yitao Liang. JARVIS-1: Open-world multi-task agents with memory-augmented multimodal language models. In *Second Agent Learning in Open-Endedness Workshop*, 2023.

[Wang *et al.*, 2024] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024.

[Warstadt *et al.*, 2023] Alex Warstadt, Aaron Mueller, Leshem Choshen, Ethan Wilcox, Chengxu Zhuang, Juan Ciro, Rafael Mosquera, Bhargavi Paranjabe, Adina Williams, Tal Linzen, and Ryan Cotterell. Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora. In *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, pages 1–34. Association for Computational Linguistics, 2023.

[Wei *et al.*, 2022] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[Yang *et al.*, 2024] Diyi Yang, Caleb Ziems, William Held, Omar Shaikh, Michael S Bernstein, and John Mitchell. Social skill training with large language models. *arXiv preprint arXiv:2404.04204*, 2024.

[Yiu *et al.*, 2023] Eunice Yiu, Eliza Kosoy, and Alison Gopnik. Transmission versus truth, imitation versus innovation: What children can do that large language and language-and-vision models cannot (yet). *Perspectives on Psychological Science*, page 17456916231201401, 2023.

[Yu *et al.*, 2023] Wenhao Yu, Nimrod Gileadi, Chuyuan Fu, Sean Kirmani, Kuang-Huei Lee, Montserrat Gonzalez Arenas, Hao-Tien Lewis Chiang, et al. Language to rewards for robotic skill synthesis. In *7th Annual Conference on Robot Learning*, 2023.

[Zha *et al.*, 2023] Lihan Zha, Yuchen Cui, Li-Heng Lin, Minae Kwon, Montserrat Gonzalez Arenas, Andy Zeng, Fei Xia, and Dorsa Sadigh. Distilling and retrieving generalizable knowledge for robot manipulation via language corrections. In *2nd Workshop on Language and Robot Learning: Language as Grounding*, 2023.

[Zhang *et al.*, 2024] Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B. Tenenbaum, Tianmin Shu, and Chuang Gan. Building cooperative embodied agents modularly with large language models. In *The Twelfth International Conference on Learning Representations*, 2024.

[Zheng *et al.*, 2024] Boyuan Zheng, Sunny Verma, Jianlong Zhou, Ivor W. Tsang, and Fang Chen. Imitation learning: Progress, taxonomies and challenges. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):6322–6337, 2024.