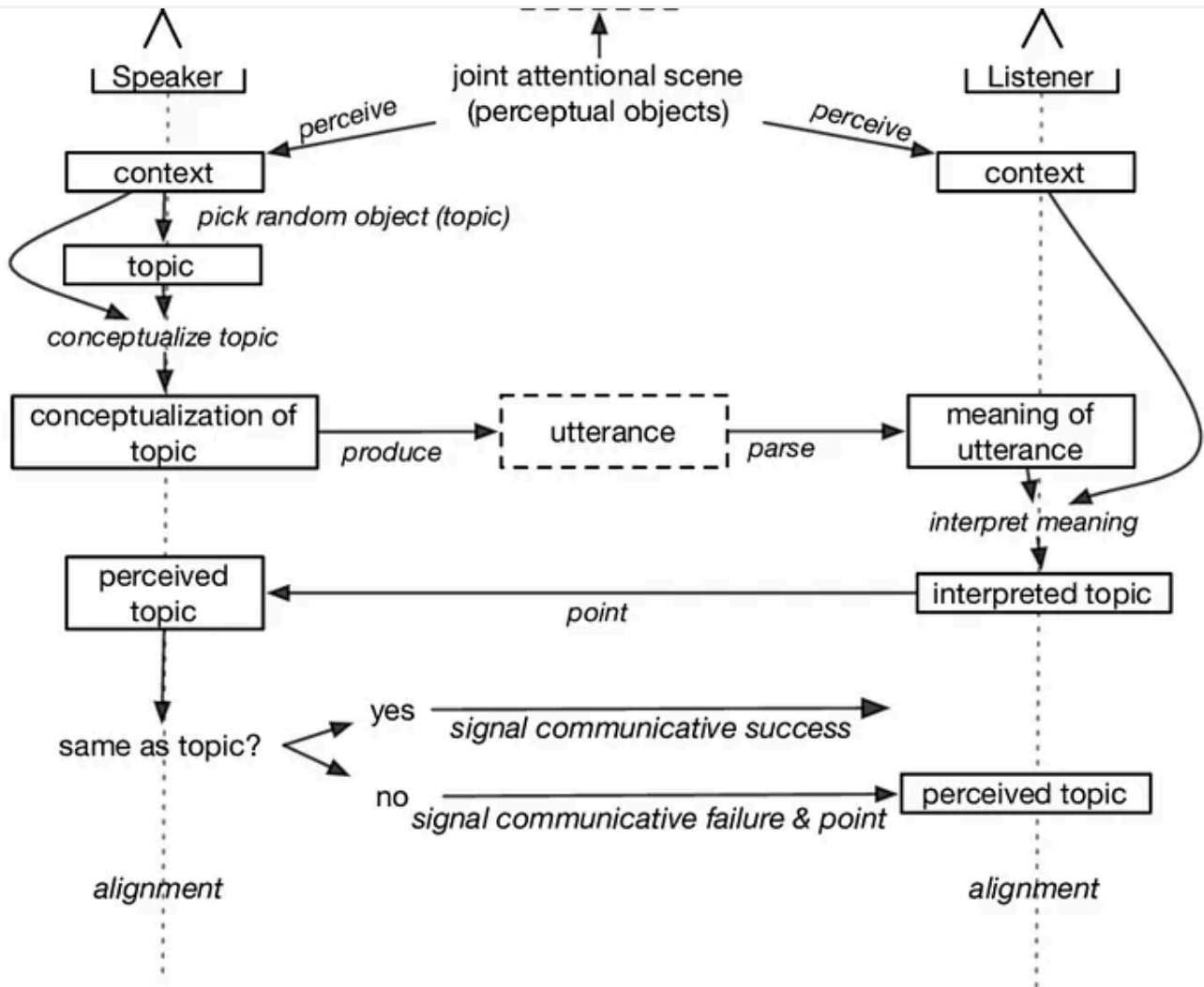Medium   🔍 Search   👤

Typical language game schematic, used for inspiration, not as an example of what UX design might produce.
From Pieter Wellens

Applying communication theory to generative AI design

A look at Dialectical Reconciliation: conversational reasoning with AI

Adrian Chan · Follow
20 min read · Mar 13, 2024

LLMs: Prompts and reasoning

Generative AI, and large language models in particular, continue to become ever more powerful and capable. What began with our fascination with chatbots and crude, lumpy, make-believe digital images has now become an all-encompassing reckoning with Artificial Intelligence.

As the public contends with the risks presented by ever-more intelligent systems, and institutions contemplate gate-keeping regulations, developers and designers search for frameworks with which to structure conceptual design and AI implementations.

Of those, the most commonplace today are still prompt engineering techniques and reasoning methods. Prompts offer users and developers ways to clarify and control the instructions issued to generative large language models, image models, video, and sound models. Prompts include user queries as well as instructions hidden behind the input and wrapped around user prompt.

Prompts can use chain-of-thought reasoning techniques to force the AI to reflect on its own generated responses, to test for compatibility with desired goals and objectives, and to improve on responses. This approach works because reasoning is logical and rigorous. Various research efforts have explored the fidelity of large language models in different types of reasoning.

Examples of this research include moral reasoning, theory of mind reasoning, logical reasoning, inductive and deductive reasoning, contrastive and contradictory reasoning, tree-of-thought and algorithm-of-thought reasoning.

Why are there so many kinds of reasoning methods? Because prompts designed to elicit reasoning from the model present possibilities of controlling the model's generation. Each angle of exploration or type of reasoning demonstrates the model's fidelity, reliability, and coherence.

And of course, being logical, these different kinds of reasoning show us ways in which LLMs might be used in specific use cases. Just consider the requirements of using LLMs for legal, financial, healthcare, management, and so many other professional disciplines.

Beyond Prompts

Prompts based on reasoning methods are not exhaustive, and other methods of interaction have proven to be interesting as well. Use of role play and personas stand out as an interesting direction. These use not logic and propositional or instructive application of language, but rather personality, style, character, and performance to organize generative AI interactions.

As the development of large language models continues apace, it is using multimodality to add internal complexity and capability to the sense-making of generative AI. Models make use of cross transformational learning when exposed to both images and language, as well as video and language. Models are also being used to learn how to act, in the digital world and in the real world. Here computer vision supplements the model's language and image generation capabilities.

The trajectory for model development, in a nutshell, is directed at making models more lifelike, intelligent, and capable in the expressive domains of language, sound, music, image, and video; and in the world interaction domain of engaging with the real world.

Formal and informal reasoning

Research into prompting with LLMs has focused on various kinds of reasoning and logic because it's important to understand their capacity for formal thinking. If LLMs are to be tasked with actions, we need to know they can identify and execute them. Can the model select from a range of actions? Can it assess when the action is complete? Can it follow instructions, or better, can it create its own?

There are many reasons to be interested in the formal capabilities of LLMs. The reasons to explore the social and interactional capabilities of LLMs may be less obvious. For one, the prospect of engaging socially with LLMs, be they chatbots or some other kind of device, is farther off. And the "rules" governing social interactions among humans are informal and unobvious. So the process of extracting and codifying them would seem a significant challenge.

But the design of LLMs begs the question: why not consider it? After all, the closer these LLMs get to us, the more they are embedded in professional work contexts and workflows, the more we are going to engage with them in "normal" and "natural" ways.

The social limitations of AI

The LLMs are limited, really, by their inability to participate transparently with human social conduct. Not by their logical and reasoning abilities.

Formal reasoning can likely be trained and fine-tuned, given well-prepared documents, data, examples, and training methods. The rules of social situations, of etiquette and social conduct, of the myriad games, entertainments, professional contexts, formal proceedings, and so on are a completely different matter.

The rules of formal logic and reasoning can be made explicit much more easily than the deeply implicit "rules" of social situations.

AI has no subjectivity

Large language models have no subjectivity. That is, they are not subjects. Subjectivity is a concept reserved for human subjects having lived experience. This lived experience grounds the subject — human actor or agent — in a real world with all of its physical constraints.

Through culture and social development, we as subjects accrue the ability to communicate and to express ourselves. This communicative ability amounts to competencies with self-expression, interpersonal communication, linguistic and verbal facility, creative and artistic expression, social interaction, and so on.

Each of these competencies makes various use of the codes, rituals, pastimes of social and cultural interaction. Again, these are implicit in social action and go without saying. On account of our different personalities and character, as individuals we differentiate ourselves through our handling of self-expression, conversations, social interactions, and so on.

But design of AI calls on subjective competencies

Why is the world of subjective experience and communicative competence relevant to the design and implementation of large language models and generative AI? For two reasons:

- model designers increasingly call on subjective competencies to model AI (from training to reinforcement learning, policies, and rewards), and

- end-users call on competencies to interact with models.

Chatbots are designed to increasingly engage in open dialog, which is a notoriously difficult domain for conversational agents. (NLP has been used historically to

constrain chatbot conversation to specific and goal-directed interaction, and to avoid drifts into off-topic chat.)

Image and video generation models require that users reflect on an image, use language to describe it, to refine it, and regenerate it. The same with video. These descriptions are prompts comprising of visually descriptive language as well as functional instructions. The same will apply with video generation: users will have to use language to describe what they want to see, as well as formal terminology to achieve specific results (camera, tracking, lighting, mood, fore and background, etc).

The same with music generation: users won't play instruments but will have to reflect on the linguistic characterization of the sound and style they desire to hear.

Each of these is interesting because they require that users translate their objectives into descriptive characterizations. This isn't natural. As much as we might be able to articulate why we like a painting, a photograph, or illustration, we are not good at producing descriptive language for use as image generation prompts. The same with video and music.

As it is currently, AI asks us to reflect on our own reasons for certain choices and to articulate them in descriptive prompts so that the AI might generate the desired output.

My claim is that in the design of generative AI, AI and user reasons slowly draw towards each other. We prompt AI to reflect on its thinking and to share its reasoning. And to get it to generate, we have to reflect on our own thinking and translate it into a prompt. It's a case of machine and user meeting through language.

The more we want to embed AI into daily social life, the more it, and we, will need to design for social interaction. And with the rules of social interaction in mind.

We need to leverage user competencies

The design of generative AI models will require that we (designers) think through the user competencies needed to get the best results from these models. (Enterprise use cases will be no exception: employees will have to become good at using models to accomplish their tasks.)

Currently, development of generative AI is largely the domain of engineers, of software architects, coders, machine learning researchers, and the like. Over time,

user experience and design disciplines will become increasingly involved, as these models require a certain socio-technical expertise and competency of users. The most capable users will be those who understand how to work the model. The best model deployments will be those that have anticipated the needs of users, and the scenarios and use cases in which they are deployed.

None of this is yet revolutionary, but the technology is still so young and is developing so quickly that design disciplines are behind the curve in seeing where they fit and belong. Enterprise and consumer application disciplines will go the same way: conceptual models, frameworks, research, deliverables, workshops, prototyping, validation, evaluation — all of these activities common to enterprise technology implementation will develop generative AI-specific methodologies and practices.

Implications for LLM builders and for UX and design

To reiterate, the development of generative AI in the form of increasingly multimodal and capable agentic systems has two implications of interest to the design community:

- the models will become increasingly lifelike and natural

- user competencies with communicative and social interaction will become critical aspects of design approaches.

The more that models approximate natural and lifelike interaction, the more users will expect them to behave as real subjects. That is, the more that users will simply engage with models using their socially and culturally learned competencies.

This presents a problem to the model designer, insofar as the more that users expect from a model, the greater the risk that the model fails.

Communicative social technologies have conventionally and historically benefited from the fact that they are machine-like (see my previous post on partner modelling and attributes of conversation agents). Users don't expect conversations from ATMs, and so ATM interface designers don't have to concern themselves with the risks of off-topic drift and open dialog.

The more that model designers extract from model capabilities, the more they will want to use meta-linguistic communication features to control model behavior and performance. Examples include personas and personalities (I think of these as

different), style, reasoning, but also roles, games and other structured interactions, conversations and aspects of turn-taking, and more. (Consider a model that can be used in modes: helpful, supportive, suggestive, interrogative, instructional, judgmental, moralizing...)

ChatGPT meta prompt example — this prompt is added to the user prompt

Schemas for LLMs?

I think it would be very interesting to see schemas developed for model (server)-side performance. Each schema could be developed by experts in the field and would not only include terminology, definitions, core concepts etc but also interaction styles, linguistic cues, personality, and more.

The schema for a health-care professional conversational agent would thus be different from a schema for online tutoring; for an insurance co-pilot; for a movie recommender.

On the user experience side, or client side, designers will want to incorporate insights from social interaction into generative model design and implementation. What are ways of talking about tasks, activities, and responsibilities in workflows that make use of AI as a co-pilot? What are the opportunities for product recommendations in customer-facing branded agents? How best can off-topic drift be managed by agents that are designed to anticipate and intervene suggestively when users take them off topic? These aren't conventional UX and design skills and methods, but will become moreso as generative AI is more widely deployed.

LLMs: information vs communication

I have argued that there's more to the reasoning of LLMs than various kinds of logical reasoning. I believe there are many implicit rules to social interactions that also involve language (written or spoken), and which offer substantial opportunities in AI model and interaction design.

There's some additional context worth noting here, and it has to do with the difference between information and communication. After all, AI is deployed online, and online is the world of digital information.

I believe that a lot of that information came from communication. And in fact LLMs are being trained on online communication. This is not just so that they have more content, but also so that they can articulate it more naturally.

Large language models generate information using natural language interfaces. They are information technologies but also tools of communication. In the world of information technology historically and conventionally, LLMs are an information technology. They challenge search and the world of the digitally-printed web.

But information that is produced through speech and conversation is just the content of utterances. Conversation "does" more than utter information or produce content — it creates and builds relations between speakers and listeners. It binds participants to a shared stretch of time and social situation. Much of the content that exists online, in fact, is communication content not information content. Or rather, it is content produced by people communicating: content produced through mediated speech and talk, not produced as documentation or record keeping.

These distinctions between information and communication are important because we use language to interact with AI. And the better it becomes, the more we will tend to use language common to social interaction. And as we will see, there are different ways of using communication in social interaction.

A detour into communication theory

All human communication is a kind of social action. Social action "does" something not by moving objects but by moving people. A threat is a social action, even if it remains only a threat. The verbal threat is the linguistic act, and is social because it has interpersonal and social meaning, and social consequences. It is also interpreted within a social context, by which it might be considered just, or unfair, or excessive, and so on.

In the communication theory of Jürgen Habermas in particular, there are two types of linguistic action: strategic action and communicative action. Strategic action achieves ends and involves instrumental use of language. Communicative action seeks mutual understanding between participants, and so build relationships.

Strategic action accomplishes something as the listener does what the speaker wants. Persuasion is an example of strategic action. It is an instrumental kind of action in which the speaker is not interested in reaching understanding *with* the listener. Advertising would be a form of strategic communication — the brand wants to sell to the user but has no interest in mutual understanding. (The meaning of content of the communication is understood, but the understanding doesn't bind the speaker and listener to a mutual engagement.)

Communicative action, by contrast, seeks mutual understanding between speaker and listener. The speaker is interested in reaching understanding with the listener, and takes this into his or her concern. In this sense, the speaker anticipates and cares about how the listener will interpret his or her communication.

Communicative action can only be achieved by human subjects through intersubjective use of communication. It is not available to large language models. But that doesn't mean people won't behave and communicate *as if* AI were lifelike. And if people begin to treat AI as if it were human, the AI should reciprocate in kind.

As models become more lifelike and capable of genuine and meaningful conversation and interaction, designers will want to use the techniques and features of communicative action as models for the interaction design of generative AI.

That brings us to the research paper excerpted below. This paper explores a highly-constrained but conversational reasoning model for large language models. It proposes propositional logic in conversational turns between the model and user as an interaction model. Propositional logic is specific and limited, but the paper goes beyond chain of thought reasoning in that it is conversational and multi-turn. Chain of thought reasoning is normally applied as a one-shot instruction or prompt to the model.

Note that this dialectical reasoning isn't strictly strategic or communicative because it uses propositional logic. In Habermas' communicative action theory, utterances also involve persuasion and claims to intersubjective truth that are outside

propositional logic. However, the research is interesting in that it demonstrates a dialogical method of reasoning not used in papers about chain of thought reasoning performed by the model on itself.

I suspect that we will have to develop a new conceptual framework for generative AI that accounts for its lifelike abilities and behaviors but which concedes its inability to care about human subjects. This will allow us to leverage all of the domain-specific linguistic and interactional competencies in designing models and user experiences. But also manage the risks attendant to technologies that are not what they appear.

Dialectical Reasoning

The following sections are from: <u>DR-HAI: Argumentation-based Dialectical Reconciliation in Human-AI Interactions</u>. Boldface is mine.

> "The significance of creating AI agents that can establish trust and accountability by interacting with human users is constantly increasing. This is the central idea behind the field of **explainable AI** planning (XAIP), which aims to design agents that are transparent and explainable by considering the human user's knowledge, preferences, and values. A crucial aspect in XAIP is the human-AI interaction, where the AI agent is expected to communicate effectively with humans to enhance their understanding and address their concerns. Most research efforts within XAIP have been placed on (sequential) decision-making tasks involving an agent and a human user, and the goal is to make the agent's decisions explainable and transparent to the human user when those decisions appear inexplicable to them. While there have been a plethora of approaches to solve XAIP from different perspectives, a process called model reconciliation has garnered growing interest. In **model reconciliation**, it is assumed that the agent and the human user each have their own (mental) models of the task, and the need for explainability arises due to some knowledge asymmetry between these two models that makes the agent's decisions inexplicable with respect to the human user's model. A solution in model reconciliation is then an explanation (technically, a minimal set of model updates) from the agent to the human user such that the agent's decisions become explicable to the human user.

Despite the growing popularity of model reconciliation approaches, we identify two limitations: (1) It is commonly **assumed that the agent possesses the human user's model** a-priori in order to anticipate their goals and predict how its decision will be perceived by them. This may lead to incorrect assumptions about the human user's knowledge and preferences, and consequently to unsatisfactory explanations. (2) Most model reconciliation approaches are formulated around a **single-shot reconciliation** paradigm, that is, they focus on generating a single, albeit comprehensive, explanation that is presented all at once. While this type of reconciliation can be useful when the human user needs to quickly understand a decision or when the underlying task is relatively simple, it may fail to work for more complex decisions and tasks that require a more detailed understanding from the human user, especially when there is substantial knowledge discrepancy between the agent and user models."

In the DR-HAI framework, **dialectical reconciliation** is the process of resolving inconsistencies, misunderstandings, and gaps in knowledge between explainee and explainer. It relies on the **exchange of arguments**, conflicts, and other dialogue moves that allow the agents to collaboratively construct a shared understanding of the decisions in question. To successfully achieve dialectical reconciliation, the agents follow certain dialogue protocols that guide their interaction:

• Establish a **clear dialogue structure**, including the use of locutions that define **permissible speech acts** and **turn-taking mechanisms.**

• Engage in a cooperative and collaborative manner, with both agents focusing on the shared goal of **improving** the explainee's **understanding.**

• Employing **argumentation techniques**, such as offering counterexamples or pointing out logical inconsistencies, to constructively challenge each other's positions and beliefs.

By adhering to these protocols and engaging in dialectical reconciliation, the explainer can help the explainee **iteratively refine their knowledge** base….

….

We now introduce the dialectical reconciliation dialogue type, drawing upon Hamblin's **dialectical games** framework. Here, a dialogue is viewed as a game-theoretic interaction, where utterances are treated as moves governed by rules that

define their applicability. In this context, moves consist of a set of locutions, which determine the types of permissible utterances agents can make. To align with the goals of DR-HAI, we define the **following set of locutions:**

L = {query, support, refute, agree-to-disagree}

The **query locution** enables the explainee to ask the explainer for an argument that supports the explainee's query. The **support locution** allows the explainer to provide such an argument that supports the explainee's query. The **refute locution** permits both agents to present counterarguments, and the agree-to-disagree locution allows both agents to acknowledge each other's utterances when no further queries or counterarguments are possible.

….

A DR-HAI dialogue is terminated at timestep t if and only if the explainee cannot generate subsequent queries or counterarguments, that is, when the explainee utters the agree-to-disagree locution.

During the dialogue, the **agents essentially decide which dialogue moves to make.** An agent may have a specific objective in mind when making its decision, such as adhering to rationality principles, ending the dialogue quickly, or prolonging the dialogue as much as possible. The mechanism for deciding a move that takes this overall aim into account is called an **agent strategy.**

….

Indeed, with DR-HAI, we are introducing a new dialogue type: dialectical reconciliation. Related dialogue types include the following: **Persuasion** [13,26], where an agent attempts to persuade another agent to accept a proposition that they do not initially hold; **information-seeking** [12,24], where an agent obtains information from another agent who is believed to possess it; and **inquiry** [3,16], where two agents collaborate to find a joint proof to a query that neither could individually. Although many dialogue systems have been proposed for the aforementioned dialogue types, to the best of our knowledge there are **no existing dialogue frameworks that consider the use of dialectical reconciliation that is aimed at enhancing an explainee's understanding.**

….

Despite the promising aspects of DR-HAI, it is important to acknowledge its **limitations** and potential areas for improvement. DR-HAI follows a fixed structure in presenting arguments and does not consider the effectiveness of personalizing the interactions according to each user's beliefs and preferences. In addition, the current model assumes that both agents communicate through well-defined dialogue moves and that their communication is seamless. In reality, however, communication might be affected by factors such as miscommunication or uncertainty. Finally, the current framework is **limited to propositional logic**, which may not be sufficient to express complex relationships and dependencies in real-world domains.

To address the limitations and improve the framework, we suggest the following future directions: (1) An **adaptive approach** that tailors arguments to individual **users' needs and preferences** could further enhance the effectiveness of dialectical reconciliation. (2) Integration of the DR-HAI framework with user interfaces and natural language processing systems, such as **LLMs,** to make the interactions more natural and accessible, especially to non-expert users. (3) Finally, although we used propositional logic to present DR-HAI and demonstrate its efficacy, extending to more expressive logics, such as **first-order logic**, **description logics**, or **modal logics**, will allow for more complex reasoning and argument generation.

Applying dialog to prompting

Prompt engineering, often in conjunction with RAG (retrieval augmented generation), is a common approach to making LLMs suitable for conversational queries and search. This approach treats the LLM as a domain expert of sorts. Prompts are designed to "extract" information from the LLM using both the model's intrinsic linguistic capabilities and RAG for expertise and domain-specific knowledge.

We could see these kinds of interactions as focused exchanges not unlike the dialectical reasoning method described in the paper above. And the approach seems suitable for domains in which tightly-scripted dialog or conversation can serve to reach an informational objective. In user experience or interaction design terms, this approximates progressive disclosure. The progressive disclosure here is not printed information on a screen, formatted for ease of use, but text returned from a large language model according to specific prompts (requests).

It isn't difficult to see situations in which LLMs are used to provide conversational access to information of all kinds. These models are already capable of returning not just plain text information but formatted documentation. Thus it is likely that the UX design of these domain-specific interactions will include dialogical or conversational prompting of varying degrees of complexity. These could be in the form of the dialectical and propositional structure presented above, or in more open forms of dialog.

Large language models have no internal social model. Nor can they have one, as they aren't living subjects. Social context is a learned and experienced form of world organization and representation. It is manifest through language (and other modes of expression), but precedes language. The social world cannot be faked by a generative AI.

The take-away for prompt engineers and user experience designers of LLMs and generative AI is that dialogical or conversational interaction is a form of linguistically-mediated communication essential to the use of AI. That dialog occurs on a continuum of communication that runs from highly formal and instructional to open and expressive. And that we (users) navigate this continuum using our competencies as communicating subjects—in this sense, the more an LLM can follow along, the more it can do.

The dialectical reasoning method presented here is but one method that extends chain of thought reasoning through dialog.

Social interaction & "as if" AI were human

Social interaction goes much farther than dialectical reasoning, and includes many types of reasoning and argumentation. Earlier I introduced the distinction between strategic and communicative action, the former being the case in which a result is obtained linguistically, but not necessarily through understanding; the latter in which two or more participants reach mutual understanding with each other over something.

Can there be mutual understanding with an LLM? Not insofar as LLMs have no subjectivity, and so cannot have human or intersubjective relationships with us. But does this rule out strategies of communicative understanding as means of engaging with LLMs? I don't think so. I believe that the better the LLM, the more likely we explore open and social types of interaction.

Generative AI is opening up a space of indeterminacy in which the AI is neither entirely subject nor machine. Large language models communicate and perform well enough, with enough intelligence and apparent competence, to engage us in ways *as if* the AI were competent. It matters less, from a design perspective, that the AI is an artificial intelligence, and more the degree with which we interact normally.

There are many more kinds of conversation and interaction available to us, as ways of doing things with words, socially, than we use with generative AI. Do we foresee a creep forward into these social domains for LLMs? It is not difficult to foresee a genre of LLMs customized with conversational domain expertise and increasingly good at "competent" topical dialog, within certain boundaries germane to that topic.

These would be language models designed to converse in specific vernaculars. They would use internal prompt designs, lexicons, training or fine-tuning on real-time chat stream transcripts, threaded social platform interactions, FAQs, knowledge bases, and so on to become competent conversation partners. Their "skills" would also include some amount of competence in social and language games common to the topic: consider the types of dialog common to sports commentary, speech making, legal deposition, court-room cross examination, etc.

Conclusion

In the research papers I have seen on sentiment analysis, emotion analysis, conversational search, and question-answer uses of AI, focus tends to be on use of LLMs to improve the accuracy of classifying user conversation and expression. Common use cases include online reviews, customer service transcripts, online ratings and comments, social media memes, toxic speech, and so on. These are often cases in which linguistic expression is ambiguous or unclear, in which user intent is not superficially obvious, and in which context can be essential to a correct interpretation of meaning. Online reviews, comments, discussions etc are also rife with incomplete sentences, insider references, misspelling, and the like.

I have not seen discussion of ways in which to use LLMs to intervene in these spaces to guide, steer, and clarify commentary for greater clarity. I could foresee this being an aspect of prompt engineering or user experience design in which bidirectional or symmetrical prompting approaches were used to try to funnel conversation for clarity. LLMs could be tuned to ask clarifying questions. They could be designed (with internal prompts designed around reasoning techniques) to mitigate and force clarification of user intent, sentiment, or emotion.

We might think of this a kind of dynamic reinforcement learning from human user feedback—generative AI systems could be designed to proceed internally through ranking, ordering, and prediction algorithms to improve conversational clarity and mutual agreement with users. Topical RAG content solutions could be merged with conversational schemas appropriate to that content area so that interactions with users take advantage of both *what* people ask about and *how* they ask.

I see this as an opportunity for user experience design. However UX would need to develop frameworks and conceptual models for how to design conversation. These frameworks might cover general conversation horizontally, and specific topics vertically. They would want to include a range of interaction models, drawing on examples of the strategically persuasive use language as well as the communicative examples oriented to mutual understanding.

As I mentioned earlier, there are no social models for LLMs (as there might be world models for physical environments). And yet so much social life involves language, making use of a rich palette of implicit rules and behavioral conventions—reasons for the how and why of communication. This space seems promising for design research by those engaged in pushing LLMs forward.

AI		Design		UX		Artificial Intelligence		Large Language Models

Follow

Written by Adrian Chan

2.3K Followers · 2.1K Following

CX, UX, and Social Interaction Design (SxD). Ex Deloitte Digital. San Francisco. Gravity7.com. Cycling, Photography, Guitar, Philosophy. Stanford '88, IR.

Responses (2)