Large Models of What? Mistaking Engineering Achievements for Human Linguistic Agency

Abeba Birhane
Mozilla Foundation and
School of Computer Science and Statistics
Trinity College Dublin, Ireland

Marek McGann
Department of Psychology
Mary Immaculate College, Limerick, Ireland

Abstract

In this paper we argue that key, often sensational and misleading, claims regarding linguistic capabilities of Large Language Models (LLMs) are based on at least two unfounded assumptions: the *assumption of language completeness* and the *assumption of data completeness*. Language completeness assumes that a distinct and complete thing such as "a natural language" exists, the essential characteristics of which can be effectively and comprehensively modelled by an LLM. The assumption of data completeness relies on the belief that a language can be quantified and wholly captured by data. Work within the enactive approach to cognitive science makes clear that, rather than a distinct and complete thing, language is a means or way of acting. Languaging is not the kind of thing that can admit of a complete or comprehensive modelling. From an enactive perspective we identify three key characteristics of enacted language; *embodiment*, *participation*, and *precariousness*, that are absent in LLMs, and likely incompatible in principle with current architectures. We argue that these absences imply that LLMs are not now and cannot in their present form be linguistic agents the way humans are. We illustrate the point in particular through the phenomenon of "algospeak", a recently described pattern of high-stakes human language activity in heavily controlled online environments. On the basis of these points, we conclude that sensational and misleading claims about LLM agency and capabilities emerge from a deep misconception of both what human language is and what LLMs are.

1 Contrasting Agencies

The current machine learning narrative is surrounded by extravagant claims, over-ethusiasm, and hype. The discourse around Language Language Models (LLMs) exemplifies its peak.

1

Fascinated evangelists claim that these models are capable of "understanding language" Wang et al. (2019), can "store, combine, and reason about scientific knowledge" Taylor et al. (2022), are approaching Artificial General Intelligence (AGI), hold early "sparks of AGI" Bubeck et al. (2023), that they are surpassing human capabilities Xi et al. (2023), and may even be (or soon become) conscious Chalmers (2023); De Cosmo (2022).

In order to evaluate these claims we place side by side on the one hand, what it is that LLMs do, and on the other hand, what human beings do when engaged in linguistic interaction. We find some significant differences, which we believe tend to undermine direct or 'literal' comparisons between people and LLMs.

Hyperbolic claims surrounding LLMs often (mis)uses terms that are naturally applied to the experiences, capabilities, and characteristics of human beings Bender and Koller (2020); Shanahan (2022); Mitchell (2019). The continued use of these terms, where such discourse is not re-calibrated in line with the comparisons, gradually shifts the meanings of words like "language" and "understanding". The literal use of these terms in this context re-orients their meanings in line with what is instantiated by the machines, and by the systems in which these machines are and will be inserted as powerful artefacts. Mistaking the impressive engineering achievements of LLMs for the mastering of human language, language understanding, and linguistic acts has dire implications for various forms of social participation, human agency, justice and policies surrounding them.

In the context of this special issue on the implications of an enactive perspective on human beings and technologies, we evaluate the relationship between human linguistic agency and the operations of LLMs; what these two things have in common, and how they differ. Comparing human linguistic practice to LLMs is itself problematic, given that 1) there is no standard or average human whose linguistic activities can be compared against that of LLMs , and 2) the metrics and benchmarks used to evaluate the performance of LLMs are riddled with various issues Burnell et al. (2023); Meister and Cotterell (2021); Aiyappa et al. (2023). Having said that, if it were the case that human language and the way LLMs operate have much in common, it will be reasonable to consider them two examples of the same phenomenon. In what follows, we argue that it is possible to offer generous interpretations of some aspects of LLM engineering to find parallels with human language learning. However, in the majority of key aspects of language learning and use, most specifically in the various kinds of linguistic agency exhibited by human beings, these small apparent comparisons do little to balance what are much more deep-rooted contrasts.

In keeping with the scope of this special issue we consider human linguistic agency from an enactive perspective. And we lean on academic literature, industry practices, and public discourse to draw our understanding and description of LLMs. The rest of the paper is structured as follows: Section 2, contrasts two conceptions of language: the one instantiated in LLM engineering, and the one put forward by enactive cognitive science. Section 3 highlights what LLMs and people appear to have in common. In Section 4, we contrast precarious embodied linguistic human participation with activities of LLMs. Section 5 delves into a recent phenomenon — algospeak — to illustrate linguistic agency in action and we conclude in Section 6.

2   Two Conceptions of Language

The statistical and computational sciences behind the development of LLMs on the one hand, and enactive cognitive science on the other, involve sharply distinct conceptions of what language is. In fact, the former rarely engages in rigorous conceptual understanding and analysis of language, but in engineering tools that imitate linguistic activity. This is a key point that underscores differences in values and goals between these different research communities Chemero (2023). As an analogy, artificial flight does not involve the kinds of things that are used to achieve flight by animals in non-human ecosystems. The goals of aeronautical engineers are not those of zoologists. Their methods and aims diverge accordingly.

The goals and aims of LLM engineers are not understanding human linguistic activity. Their goals relate to the production of language-like performance in text (and audio) production. This means they have little bearing on natural linguistic interaction. This in no way undermines the incredible engineering achievements of LLMs, but it does help us understand that we should no more believe that LLM output is the same phenomenon as human language than we believe that an Airbus 8320 tells us something important about hummingbird flight[1].

These differences in aims and values between the cognitive scientific and engineering communities results in quite distinct understandings and assumptions of what language is, what counts as engaging in linguistic interaction, as well as widely differing views of how language relates to other aspects of being.

2.1   Large Models of What?

In 1948, Claude Shannon wrote on the relation between language and entropy that: "Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem." (Shannon, 1948, p.3). Current large language models such as Gemini, Bard, Llama, Megatron Turing, Bloom, and the GPT variants, are an engineering endeavour in these terms, albeit much more sophisticated and enormous in scale compared with Shannon's original concept. Shannon's original idea of entropy, that what is being measured, represented, and manipulated is form and not meaning, remains relevant now for LLMs.

Layers of artificial neural network architectures, most notably transformer-based systems

---

[1]It is important to note that the different professional goals of LLM engineers as compared with cognitive scientists do not imply a reduced ethical burden. In fact, given the almost certain ubiquity of LLM-produced text and generated speech in the coming years, not only is there an extended analogy with that of artificial flight, and the pollution of ecosystems, but additional considerations also apply. Additionally, language is a domain which is quintessentially human, and coherent speech production has, to this point in our history, been a strong positive indicator of the presence of ethical duties toward the producer (though the converse does not hold). In circumstances in which people become inured to the experience of dismissing LLMs as the interfaces to machines, corporate or otherwise, the risk of increasing the already horrendous dehumanisation inherent in much online and offline activity becomes great. How these risks should be mitigated should be a vibrant domain of discussion for the burgeoning field of LLM development, much more so than spurious concerns about existential risks to human civilisation.

and diffusion-based approaches are initialised and trained using massive datasets, often text and text-image pairs, typically sourced from the World Wide Web (WWW) and condensed and filtered via numerous automated mechanisms. In transformer based learning, the data are broken down into "tokens", typically a few characters in length, and the LLM develops a statistical representation of the relationship between billions of different tokens. While there are various ways in which the operation of these systems can be described, they are ultimately well-defined calculations across tokens, which results in a system which, when given an input, treats that input as the initial movements through a space of possible valid moves (token concatenations), and continues the path of those movements depending on the complexity of the model and the availability of computational power (with longer paths requiring much greater resources). In a sense, the LLM is comprised of a statistical model of the relationship between tokens in a dataset, and a pathfinding mechanism optimised to generate valid sequences of token concatenations using that model, typically displayed as text. In keeping with Shannon's comment above, we note that the processing of datasets and the generation of output are engineering problems, word prediction or sequence extension grounded in the underlying distribution of previously processed text. The generated text need not necessarily adhere to "facts" in the real world.

For this reason, LLMs have been dubbed as Stochastic Parrots Bender et al. (2021), Bullshit Generators McQuillan (2023); Hicks et al. (2024), and The Great Pretender Coldewey (2023), amongst other terms. Though the output from such models often seems impressively meaningful, that apparent meaningfulness depends on the extent to which the meaning of a given word, phrase, or paragraph, can be represented by the relationships between tokens in the original dataset. For example, multilingual models trained on multilingual datasets show poor quality performance for languages outside the status quo Kreutzer et al. (2022); Khanuja et al. (2021); Wu and Dredze (2020). The language that an LLM represents is conceived by its engineers and developers as something static and complete, which can be captured in the relationships between tokens.

From this description we can see that claims regarding linguistic capabilities of LLMs depend on two implicit assumptions of language. The first is what we call the *assumption of language completeness* - that there exists a "thing", called a "language" that is complete, stable, quantifiable, and available for extraction from traces in the environment. The engineering problem then becomes how that "thing" can be reproduced artificially. The second assumption is the *assumption of data completeness* - that all of the essential characteristics can be represented in the datasets that are used to initialise and "train" the model in question. In other words, all of the essential characteristics of language use are assumed to be present within the relationships between tokens, which presumably would allow LLMs to effectively and comprehensively reproduce the "thing" that is being modelled.

Both of these assumptions are rejected by an enactive view of language, which sees it not as a "thing" to be captured by text data, but a practice in which to participate, whether that participation is through speech, written, sign, or other modality. In contrast with computational approaches' emphasis on form, the enactive approach to language recognises that what truly matters for language is its meaning (in this there are strong resonances with Bender and Koller (2020)'s critical examination of the relationship between form and meaning in language model output). As such, the enactive approach to language starts not with tokens of verbal activity, but with the fundamental issue of agency, embodiment,

precarity, and how meaning arises within situations where things matter to those involved.

2.2 Doing Language

Enactive descriptions of agency begin with a process of continuous precariousness – a network of processes constituting a body in a constant flux of degradation for which engagement with and acting on the world provides some prospect of continuity through the managing of conflicts and tensions that result in the degradation Beer and Di Paolo (2023) .The logic of precarious but continuous dynamic organisation is one that arises in the domain of organic, bodily being, but is general across other domains. Thompson and Varela (2001) first outlined three such domains – the organic (that of biological bodies), the sensorimotor (that of skilfully engaged bodies), and the intersubjective (that of interacting skilfully engaged bodies). These three domains can be conceived of as distinct, but nevertheless interact in inextricable ways. Organic needs can drive and constrain sensorimotor activity, while those skilful doings in turn affect and transform those biological dynamics. Similarly, the intersubjective is a coordination of skilful bodies interacting with one another, but this domain is constrained by organic and skill-relevant dynamics, while also imposing constraints in turn. Think of the ways in which societies (a collection of intersubjective phenomena) are substantially organised around food production and meal times, constrained by the biology of hunger and nutrition. Hunger and nutrition are in turn affected by social norms and standards around what kinds of foods are generally available (within the society's "cuisine") and standardised meal times. These impacts occur over multiple timescales too, as we can see in such phenomena as available foods and cultural resources leading to rises and falls in prevalences of Type II diabetes Magliano et al. (2019) intersubjective phenomena driving and constraining biological ones.

Within an enactive perspective, agency generally, then, is driven by tensions of precariousness and risk. There are continual needs to diffuse a tension by, for instance, expending energy to move and act. These actions, however, necessarily introduce new tensions, such as a hunger arising from the expenditure of energy. Di Paolo et al. (2018) provide an extensive characterisation of this inter-dependency between the organic, the sensorimotor, and the intersubjective, and extend this analysis of agency into the linguistic domain. *Linguistic acts* are those which manage an inter-subjective tension – a precariousness in the coordination between two or more people engaged together in a shared activity. Their analysis places great emphasis on the variety of ways in which such shared activities occur at multiple temporal and spatial scales, and that the resolution of a tension at one scale tends to introduce tensions at another.

There are two crucial implications for our understanding of language that follow from Di Paolo et al. (2018)'s account. The *first* is that we are always doing more than one thing. Linguistic actions are made within a nested set of contexts. When we encounter other people we are always already in some broad form of coordination with them in which we are participating. For instance, if we meet someone for the first time at a job interview we are already both participating in the behaviour setting Barker (1968); Schoggen (1989) of job interview. Our actions are thus already coordinated at a coarse grain of analysis, but also constrained – there are pressures and processes which will guide and drive our behaviour appropriate to the setting. Thus, coordination produces new tensions that must be

managed through our linguistic skills. We spend pretty much all of our lives within behaviour settings Heft (2001), and our actions are organised accordingly, classically illustrated with the slogan, "people in church behave church, people in school behave school". Within many settings, however, there are multiple participation genres Bakhtin (1986) possible. There might be collaborative work, team sports, or more. These are separate but entangled with the behaviour setting and produce their own set of pressures and constraints. Any given situation is a complex of these interacting constraints that is being managed by our behaviour at different scales from very brief microexpressions accenting our speech to the general tone and vocabulary of spoken utterances, to the broad overall shape and sequence of actions appropriate to the behaviour setting. Linguistic agency is *not* just the words uttered in this situation, but the whole process of managing all of these structured, intersubjective aspects of living.

This flow of tensions brings the *second* implication of the enactive account of linguistic agency to the fore. From an enactive perspective, any linguistic act is necessarily *partial* or *incomplete* in two different ways. First, an individual's utterances are partial in that they are always made in response to (or in anticipation of a future response) and in coordination with another person as part of a shared on-going activity. Second, while an utterance or other linguistic act manages the tension arising at one level of the interaction, it cannot resolve all such potential conflicts and therefore introduces new tensions in the nested contexts that characterise the situation as a whole. These new unresolved tensions become the animating force for the driving forward of the interaction as it continues to unfold, precariously, over time.

This perforce very brief sketch of the enactive theory of linguistic agency illustrates how essential embodiment, participation, and precarity[2] are to human language practice. We can see how strikingly different the conceptions of language maintained by computational approaches and enactive cognitive science are.

As noted above, this is not just a matter of nit-picking, but a divergence of fundamental tenets. Proponents of stronger claims regarding the validity of LLM operations as intrinsically linguistic may suggest that science and engineering often work with minimal or limited cases in the first instance, to grasp principles, before 'scaling up' to more richly contextualised, complex settings. Disembodied, text-based interaction may only be the start, with full-bodied, *em*bodied, real-time linguistic interaction being the end goal.

Setting aside for the moment that there is nothing minimal about the storage, computational, or environmental costs of LLM operation or the underlying profit maximization

---

[2]Readers will note that we use two related terms in discussing the fragility that provides stakes to the actions of autonomous systems: precariousness and precarity. Enactive researchers Di Paolo (2009); Beer and Di Paolo (2023) have developed a technical definition of precariousness, which is a tendency of component processes of a self-producing network to lead that network toward dissolution - this is balanced by organisation of the network as a whole to lead the network toward stability. This dynamic occurs in different domains - organic, sensorimotor, intersubjective. The intersubjective domain overlaps substantially with the concept of precarity, which refers to the vulnerability of the agent to injustice, exploitation, or social or physical injury on the basis of their participation within various social networks and social activities. Exploring the fine-grained details of the relationship between these overlapping concepts is beyond the scope of the present paper, but we draw the reader's attention to it as a consideration for further work to be done in this area. We discuss the explicit relevance of precarity for our understanding of the differences between LLMs and human beings at length in Section 4.3 and Section 5.

objective driving LLM development, we can see that what constitutes *minimal* linguistic interaction from an enactive perspective is not something that is constrained and disembodied, but something that is blunt and lacking nuance, but still embodied, participatory, and precarious. What constitutes minimal valid linguistic activity looks a lot more like our reciprocal interactions with pre-verbal infants, or perhaps even the playful negotiation of contextualised behavioural coordination with our pets, than it does disembodied grammatical sentence generation.

The enactive and engineering stances nevertheless share some, at least superficially, common ground which we feel is important to recognise. In the following section, we address first where these two viewpoints can be seen to agree, and explore in Section 4 how the three aspects of enactive language are absent in the operation of LLMs.

3   Comparisons – of Humans and Machines

An apparent similarity between both human and machine is that both LLMs and human linguistic activity are grounded in language phenomena that are necessarily shared, public, and historical. Wittgenstein's (1958, §243-§271) famous private language argument articulates what has been demonstrated time and time again both by typical and atypical language development in children. While children learn language readily, they must do so in interaction Gros-Louis et al. (2014); Tamis-LeMonda et al. (2018), and to become full participants in any human community means that it is *not* adequate to simply develop a coherent, recursive system of symbolic representation. Rather it is vital to work with and through existing means of social interaction (particularly speech) within a given community. Human language is, therefore, built on routines, practices, and rituals to which the language learner is exposed through a prolonged course of development Leather and Dam (2003).

At this very coarse grain of description, LLMs and human language learners would seem to have this much in common. A human learner develops through immersion in the language of a community. They experience a wide variety of utterances produced in a wide variety of circumstances. Language learning must involve extensive such experience, and as the great success of immersive language experiences demonstrate in comparison to more compartmentalised learning, the more extensive the better Cummins (2009). For this reason, human linguistic activity has been reframed as *'languaging'* within enactive writing. Languaging emphasizes the fact that language is active, dynamic, and embodied – constituting *voice*, *text*, *gestures*, *body languages*, *tones*, *pauses*, *hesitations*, as well as what has been *left unsaid*. Languaging, therefore, defies datafication, as it cannot, in its entirety, be captured in representations appropriate for automation and computational processing. Written language constitutes only part of human linguistic activity.

Given a generous and sympathetic reading, there is an apparent similarity here with LLMs, which are built on a backbone of a massive corpora of, primarily text data, just one element of human linguistic activity. Training data for LLMs are almost unimaginably large and getting ever larger. Billions of utterances spread over millions of texts sourced mainly from the WWW, from which massive datasets are assembled that then form the backbone of these models. In this one aspect, at this coarse grain of description, LLMs and human language learners appear to share characteristics. Even here, while both are grounded in

exposure to public practices and demonstrations of linguistic skills, the kind of exposure in question is wildly distinct.

4   Contrasts - Precarious Embodied Participation

Despite sharing a reliance on extensive exposure to existing language, human languaging and LLM operations differ in crucial ways which lead us to conclude that what people do with language, and how LLMs operate, are fundamentally different. Underlying all is the vital enactive concept of *agency*, something which is definitional of living systems, but is absent (despite some appearances) in LLMs. In this section we address three aspects of linguistic agency, in particular: *embodiment, participation,* and *precarity*. We begin with *embodiment*, which both motivates human languaging and also engages us with the world in concrete ways that pressure and constrain our actions. We then look at the inherently active, concerned character of human actions which means that language is never simply a going through of motions, but a necessarily active and engaged *participation* between human beings, even in the lightest and seemingly least invested of cases. And finally we turn to the values, the stakes, of human participation, and the concept of *precarity*, and the continuous risk that is the dynamo of agency.

4.1   Embodiment

While developing a capacity for language does involve exposure to, indeed immersion within existing linguistic practice, we have noted above that in the case of human beings that immersion is never a passive thing. In infancy, the child is involved in various kinds of interactions and games with their caregivers in which language starts having a peripheral role, but which becomes increasingly central. These various activities are ones in which the coordination between the parent and child that supports the development of linguistic skills is bodily – physical, concrete, and grounded in the shared experience of the ongoing interaction Smith et al. (1988); Tamis-LeMonda et al. (2001); Yu et al. (2009).

There is no question that human beings interacting with an LLM are engaged in linguistic activity. Users are mature linguistic agents. Indeed, in many cases for us academics, we are engaged in a complex reflective process of evaluating the linguistic agency of the LLM itself. Interactions with an LLM thus themselves occur immersed within a nested set of interacting contexts. The enactive perspective makes it clear that all linguistic agency is embodied, and that the body, in various ways, plays a role of animating force, enabling condition, and set of constraints Chemero (2023); Di Paolo et al. (2017); Thompson (2010); Varela et al. (2016). We should, then, consider carefully the kind of embodied interactions that take place when a person uses an LLM.

We have noted that the user of an LLM is a mature linguistic agent, someone who is a proficient participant in typical human discourse. We know this because part of the process of engaging with an LLM requires the use of widespread but nevertheless deeply technical skills, such as literacy and typing on a keyboard. The user acts toward the model through the technological medium of the keyboard, and the response from the model is typically displayed on a computer monitor, formatted for reading according to a particular genre of

text-based interaction in online communication. Other modes of interaction exist, but it is important to recognise that this 'typical' mode is primary. Where other forms of technology such as speech-to-text is used to input prompts, for instance, only the sequence of recognised words, with some interpolated grammatical markings are presented as input. Other aspects of the utterance, such as cadence, rhythm, tone of voice, and accent are either ignored, or indeed, become impediments to the interaction; for instance, in the context of the English language, the more accents diverge from those considered 'standard' American or British English, the less functional such interfaces become Johnson et al. (2022).

At the other end of the interaction with an LLM, screen readers and other assistive technologies enable engagement with output from the model that differs from the default 'monitor' case (see Section 4.1.1 below). But again, no meaningful peculiarities of those modalities are produced by the LLM itself. The non-typical output is dealt with purely by the assistive technology on the basis of the text output by the model.

On the basis of this description the role of the body in participation in text-mediated communication would appear to be very limited. A few twitching fingers or moving lips, the skipping of eyes across an array of displayed text, are all that seem to be needed. But bodily involvement in textual communication is much richer than this.

Referring to a playful passage by the author Italo Calvino, Caracciolo and Kukkonen (2021) note that reading invariably takes place in a physical context, one often carefully tailored to the task that needs to be done. While Calvino writes about sitting with a book, or lying down with one, or putting up one's feet, or propping up the book, the experience of the text is contextualised in this physical manner. In the case of the kind of machine-reading involved with LLM interactions, we can note how aspects such as font, font-size, and dark or light themed interface affects the interaction for the human, but makes no difference for a machine.

More than this, though, Caracciolo and Kukkonen (2021) explore the ways in which our capacity to engage with, and become involved with text, depends on fundamentally embodied capacities such as shared attention, emotional resonance, and appreciation of rhythm and flow. The authors note that their approach is most directly relevant to long-form narrative, but is grounded in embodied cognitive science of language use more broadly. Some aspects of these phenomena might leave traces in the corpora available to LLM training but most do not. It is the dynamics of shared involvement with a story or utterance, and how these are responsive to the real and imagined readers, that play a significant role in our engagement with text.

More pointedly, we have noted that people involved in linguistic interaction are always embedded within a nested set of contexts. These contexts have multiple facets to them, some of which we raise in following subsections. For our present purpose, we note that the kinds of goals and tensions being managed through linguistic activity invariably involve bodily groundedness. There is a reason that we are involved in particular conversations, participating in particular settings, trying to achieve particular ends, at any given time. While many of these would seem to be so flexible and negotiable as to be autonomous, they nevertheless have an ultimate grounding in constant needful freedom of organic being (these issues are unpacked at length in Di Paolo et al. (2018)). The shared intention and shared experience that is a large part of what enables successful interaction, coordination, and mutual understanding, is something that arises because our bodies enable it. When words

fail, we can turn to experience to help re-calibrate an interaction, to negotiate new shared references. Language is always necessarily grounded outside of language by its relationships to these contexts of shared activity and bodily experiences. *That machines lack such bodily experiences is not a trivial difference.* It is instead something that results in an artificial but fundamental limitation to the possibility of machine language. Even where an interaction between two human beings is mediated by text, these bodily aspects ground language use in real experience of our inexhaustibly rich world. This results in artefacts of LLM output that demonstrate these important constraints.

4.1.1 A procedurally-generated choose-your-own-adventure game

The technical mediation of person-LLM interaction is central to understanding some of the ways in which machines, which are engineered with a conception of language as a 'complete' entity, captured by the massively extensive training data, operate. This conception of language as a coherent and complete 'thing' is quite distinct from the always-generative reality of linguistic action grounded in real world bodily experience.

The LLM API, or the prompt interface for ChatGPT (or its latest variants) constrain and then transduce the actions of the user into valid moves in the domain of the language model. The array of actions available to the user are constrained by the interface (something common to all text-based digital communications), and these constrained actions are also then transduced into valid input for the model by the API.

In many ways, this constraint and automatic validation of the input is akin to the use of a controller in a video game. Only certain kinds of actions are possible (depending on the number of buttons/sticks and their potential combinations on the controller), and input outside of the defined set of values are ignored (or trigger a specifically constrained code exception).

In most video games, the range of potential actions is limited. The movements of the avatar on the basis of the player's inputs have a relatively narrow range, and the edges of the fictional world to be explored by the player are easily found. In many cases it is simply impossible for an avatar to be moved past a certain point, resulting in the classic 'running on the spot' phenomenon often lampooned in satires of video gameplay.

Some games, however, have what are termed 'procedurally generated worlds', some renowned examples include *Minecraft*, and *No Man's [sic] Sky*. In such games, for every move the player makes toward the edge of the currently represented fictional world, the game generates, according to certain rules, an entirely new portion of gameworld. The result is that the extent of the fictional world is only limited by the computational resources available to the game software. Similarly, interactions with LLMs are the text-based equivalents of procedurally generated fictional worlds. Every action taken by a user is an exploration of such a world, and every new prompt is a move toward the edge of that fictional world, such that the system generates and presents new gameworld in response to the player's (or user's) actions.

Understanding LLMs in this way helps illuminate some of what is deeply impressive about their behaviour, but also their limitations; how they might be used in actions of genuine linguistic agency by human users, but also how LLMs cannot demonstrate such genuine agency themselves.

On the one hand, LLMs are striking not just in their production of procedurally generated grammatical text, but the fluency of that text – it makes sense not only at the level of individual phrases and sentences, but over the flow of a larger body of text of hundreds and sometimes thousands of words. This is akin to a video game being able to 'show' me gameworld regardless of how far I move in the game, or whether I follow particular paths or intended patterns relevant to particular goals or "missions" for the game. The game always presents me with valid ground to 'walk' on, with a layout of objects that makes game-sense. No matter what I do, I will not reach the end of the world, there will always be more to explore. In this way, there is something not only impressive but satisfying about the game. I feel like I have discovered something new, and that when I am playing in the world it feels less like a puzzle, and more like a 'real' world, always with some new horizon to move beyond.

On the other hand, this is a game, and the procedural generation of gameworld is determined not by the existence of an inexhaustibly rich environment available for exploration, but the set of procedures implemented by the game engine. The world thus presented is still a game, and the range of actions that can produce effective outcomes is limited, with some actions being valid, and others not. For instance, though I can walk forever toward the edge of the gameworld and never meet it, if I walk closer to an object the world runs out of new things to offer me very quickly. Depending on how the rendering has been done (and my hardware resources), the textures of the object will become blurry as I get close, having reached a maximum resolution. As we look closer, we realise that there is nothing more to see.

What is more, it is often the case with LLMs that while new gameworld is continually being generated, the previously visited world is lost. Subsequent interactions based on the same 'moves' do not coherently revisit the same world, but generate new text-based gameworld. For example, Figure 1 presents an interaction between one of the authors (AB) and ChatGPT. The text of the chat demonstrates [3] how the LLM recapitulates the sexist biases inherent in the data and tuning processes that constructed the machine. We see that despite the fact that each prompt produces effective output, there is an incoherence of the discussion as a whole, as new responses are produced to individual inputs, without the overall shape of a 'conversation' and its implications. New outputs by the text engine are not shaping the overall context of the interaction; revisiting prompts after clarification does not result in better, more carefully refined results.

Interactions with LLMs are much more akin to the gameworld than the real one. The text-based interface constrains the kinds of moves that are valid, and therefore what kinds of explorations are possible. While the text-based interface makes it seem like the procedurally generated world is just as rich as would a real world be (interacting with a real linguistic agent like another human being), it is, however, merely an appearance dependent on the limitations of valid 'moves' by the constraints of the interface. There will always be new text produced, but the way in which LLMs fabricate sensible-seeming but too often inaccurate or

---

[3]Due to the non-determinable rules underlying generative models, an exact same output may not be generated for a given input prompt. Furthermore, large tech corporations such as Google tend to selectively correct notable errors as they appear on widely used products. However, due to the countless ways these large systems can generate new output, errors of this kind will always be generated in expected ways.
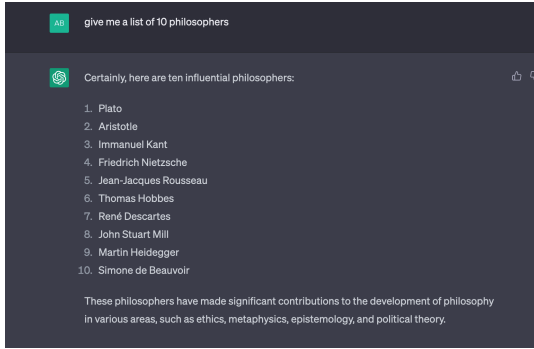
(a) Output 1
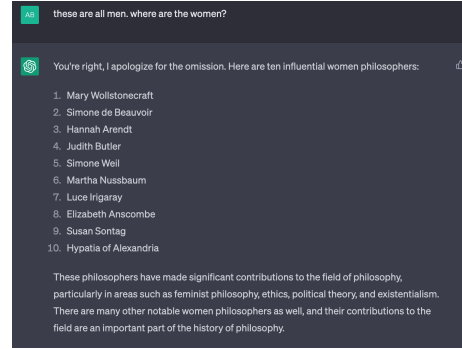
(b) Output 2
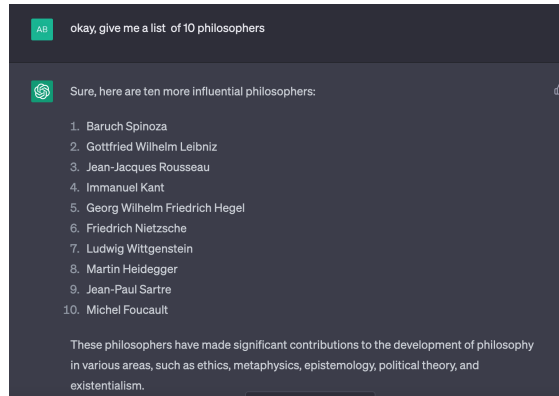
(c) Output 3

Figure 1: LLMs lose the thread of a conversation with inhuman ease, as outputs are generated in response to prompts rather than a consistent, shared dialogue. (ChatGPT prompt output. AB, 19th April, 2023)

even nonsense text when prompted to provide more specifics and details in complex domains is the equivalent of the textures blurring in a video game. There is no more detail to see.

Such un-grounded production of grammatically accurate but contentfully empty or vague text has been described as "confabulation", or sometimes, "hallucination" OpenAI (2023), but these are inaccurate terms. "Hallucination" is a failure of perception, the experience of something as present in the world that is not actually present. LLMs do not perceive – they are statistical models of a corpus of data. Nothing about their operation tracks or engages with the physical environment around them.

"Confabulation" is a similarly psychological term and perhaps less obviously misplaced. Human confabulation is the production of quasi-sensible narratives or explanations, in response to queries or prompts, that bear little to no relationship to the state of the world. Often seen after certain kinds of neurological damage that results in amnesia or certain forms of bodily dissociation along with, crucially, a lack of awareness of the problem, it can also be produced more simply in neurotypical individuals in the right circumstances (e.g. the phenomenon of 'choice blindness', Johansson et al. (2005)). Confabulation bears most of the hallmarks of 'bullshit' Frankfurt (2005), and there are several analyses of LLMs which examine their output as such Lakshmanan (2022); McQuillan (2023); Hicks et al. (2024). We are in general agreement with these points of view. We note that the unmoored, truth-free

character of such text is present in apparently accurate output just as much as it is in the more obvious nonsense. This is precisely because this text has no grounding in a shared context or experience, only in statistical relationships between words. LLM (mal)functionality is not confabulation, it is *fabrication*. Rather than an invented story that helps keep a flow of dialogue active and continuing, it is the generation of sensible-seeming, yet nonsensical text output on the basis of processed corpus. Crucially, because there is no difference in the processes used to produce the different outputs, LLM text is fabrication *even when the resulting text output is appropriate and accurate to the reader's needs and reality.* These circumstances are akin to a computer gameworld having a coincidental resemblance to a real world landscape layout.

The net result of all of LLM functioning is a text-based interaction with a maximum grain of resolution that cannot be managed within the interaction itself, because there is *no shared ground* or experience between the person and the machine against which they can calibrate their use of terms. The person's actions are grounded in their embodied experience, the machine's output is grounded in the meta-data that has been produced on the initialising corpus. Under such circumstances, mature linguistic agents such as typical human beings, struggle. The experience is uncanny. It can be at one moment seemingly straightforward and sensible, at others bizarre and frustrating.

We struggle because when we are in a conversation with another human being we are doing something other than just exchanging words. Perhaps we are just chit-chatting because we barely know one another and the details of what we say barely matters so long as the tone, attitude and broad subject matter is right. These are low stakes interactions in which a low-resolution conversation is perfectly fine. Such low-resolution conversation is almost always directed towards aspects of the world where shared experience can be found (e.g. the weather (in Western societies)), and the validity of the conversation calibrated by those involved. Managing the coordination of the non-verbal aspects of things to keep the conversation working, even in this low resolution mode, is itself important.

In more higher stakes circumstances the details of the words matter more. Broad strokes chit-chat style of responses quickly become frustrating for our conversational partner and stressful for us. We realise that our actions are inadequate. The specific context and relationships between the people involved matter a great deal as to how this inadequacy can be managed. Oftentimes, it is managed through negotiation or collaboration. Knowing we are not getting things quite right, we might ask the other person what particular details they want to know, or express regret about the limits of our ability to articulate what we intend. Together with our partners we might steer the conversation toward more effective communication, or find ways to tease out the details that were missing in the first rounds of back-and-forth. These collaborative aspects of language, which are central, not peripheral, to linguistic agency, are absent in the text-world generators that are LLMs. In a sense, how helpful or useful an LLM output could be is directly linked to how creative and familiar (with the quirks of LLMs) the person prompting the LLM is. The LLM itself is no more a collaborator than a piece of clay is a collaborator to the sculptor exploring what shapes will and will not hold together effectively as they squeeze and mould it.

Because LLMs generate a text-world as we interact with them, they cannot be used to navigate or understand the real world, because there is no reliable relationship between the real world and the procedures of text generation, and no way for that relationship to form

and be maintained. *To try to understand the real world on the basis of LLM output is like trying to navigate the real world using a procedurally generated video game.*

This analogy breaks down somewhat when we address language as the domain in which we are navigating, but not in a way that is helpful to the LLM. It is true that language is, in a sense, a process of continual generation. We have noted above that it is radically incomplete; the enactive perspective makes it apparent that every linguistic action is partial and generated within the context of an interaction with others. Is this not the same as a procedurally generated video game, always being constructed as we move forward?

The difference lies in the way that all linguistic activity is multiply embedded. While every linguistic action is partial or incomplete, the validity of actions is governed not just by its relationship with what happened immediately prior, but also with the broad flow of activities in which it is embedded. Human utterances are often sensitive across these scales, and are continually calibrated in relation to them. "Machine utterances" can be tweaked by explicit inputs, directions, or instructions from the user on how to respond. However, because they cannot be updated with reference to shared knowledge and mutually meaningful reference points throughout the course of the discussion, the interaction between an LLM and a person is necessarily isolated from the real world. You must go to play in the LLM's virtual world; *it cannot come to yours.* Lacking bodies, LLMs cannot have the urge, motive, interest, experience, and pressure to engage linguistically, and as we outline below, cannot truly participate in a conversation.

4.2   Participation

Following this more active, participatory, and embodied perspective on language, we find more stark and important differences between humans and machines beyond the capacity to calibrate against embodied experience. It is precisely these active, collaborative, and dynamic aspects of languaging which are not and cannot be captured in static representations and included in a corpus of training data. Languaging — including the casual chit-chats as we enter an elevator with others, gestures, body languages, tones, pauses, and hesitations — is not something that can be entirely captured in text but is an often fleeting phenomena without clear formalizable rules. These embodied linguistic participations can be peculiar, unrepeatable and take on a "life" of their own in a way that is not predictable Di Paolo et al. (2018).

The social character of linguistic agency is not coincidental. We have noted that LLMs are developed on the basis of a large body of existing practice and textual language use. But even in enormous datasets, that body of practice is fixed. It is not a body to which the model contributes as it "learns", given that even when new text is generated, it is regurgitated and reconstructed on the basis of the training corpus. This is quite in contrast to human linguistic agency in which participants both experience the practices and contribute to them. Although the data that forms a model's training set is partly sourced from human linguistic interaction, at best it captures a snapshot of a dynamic human textual linguistic interaction and 'freezes it in time'. Training data therefore is not only necessarily incomplete but also lacks to capture the motivational, participatory, and vitally social aspects that ground meaning making by people. In fact, elements of motivational, participatory, and social aspects of meaning-making often defy codification and datafication. For instance, we often

make ourselves effectively understood from what has been left unsaid in a conversation, or via tones of voice that transform the meaning of an utterance, as in sarcasm. Generating and detecting humor, sarcasm, and jokes, on the other hand, are qualities that remain impoverished in LLMs. Jentzsch and Kersting (2023), evaluating ChatGPT outputs, for example, found that over 90% of 1008 generated jokes they examined were the same 25 Jokes.

When we speak we frequently misspeak, we hesitate, stumble over words or use the wrong words, or the wrong constructions. The people with whom we are engaged provide support as we fumble our way forwards, and we in turn support them Dingemanse and Enfield (2024). Dingemanse et al. (2015), for instance, found that clarification questions occur on average every 84 seconds in normal conversation. The level of frequent and adaptive clarification we see in normal human conversation occurs due to an underlying shared sense of direction to the discussion, even where its conclusion is not known. When you have interacted with an LLM, how frequently have you encountered requests for clarification? (Beyond perhaps stock phrases in response to very explicit expressions of frustration in the prompt.) When a person has knowledge of a domain, they typically have a strong sense of how to ask questions and what details to seek or avoid in order to support an on-going dialogue. People are aware both of what they know, what they don't know, and how well the conversation overall is going. The seeking of clarification is a kind of activity that is grounded in a shared direction for the conversation, in which the discussion is continually being sculpted and steered as a collaboration. To be capable of clarification and repair, the participants have to be sensitive to divergence and breakdown. Indeed, the lack of question asking, or metacognition regarding the tentativeness of much of our understanding, is part of what has resulted in LLMs being experienced as fluent in the 'mansplaining' idiom Harrison (2023).

These collaborative activities sometimes involve helpful corrections, sometimes carefully ignoring invalid statements, sometimes enthusiastically adopting new meanings for old terms, or new words that give better expression to an experience that we share but neither of us can yet put satisfyingly into words.

To *understand* language is *not* to be able to produce grammatical strings of words, but rather to participate in this process of negotiated, participatory meaning making. As we have noted above, it is this active, participatory character of language that has led enactive researchers to adopt the verb languaging in preference to the nominal 'language' in the research literature. It is an inherently collaborative, dynamic negotiation of meaning, the textual aspects of which are only part of the story. This remains the case even in the constrained text-centric domain of online interaction.

This emphasis on participation and coordination over sentence construction means that much of the research comparing human and LLM production is simply not germane to the question of human linguistic activity. There is a wealth of such research now. Analyses find some parallels between the two (e.g. in variation of word use based on recent semantic context from both its own output and prompt input, Cai et al. (2023)), and some differences (e.g. in appropriate coordination of output with scalar and general conversational implicature of recent output and prompt text; Qiu et al. (2023)).

Given that a LLM is a curve fitted to a dataset with a sophisticated mechanisms for sampling, such analyses have a potentially important *engineering* role, in evaluating the extent to which there is appropriate correspondence between the map (the LLM) and the

territory (human word production in text-based linguistic activity). They cannot, however, provide any argument for the validity of conflating map and territory. Evidence for that simply lies outside of word production, in the field of embodied, participatory, and value-laden interaction between agents. It is possible that artificial linguistic agents might be developed and engineered in the future, but evidence of such success cannot be on the basis of patterns of fluent token sequence production.

The enactive conception of language, because it involves dynamism and sociality, is one which recognises every linguistic act to be radically incomplete. According to this perspective, language is a partial act that can only be completed when it is taken up and extended, embellished, or steered and redirected, by other agents. This can be other people engaged in a complementary or counter move, which is itself also incomplete, dependent on that gesture or utterance being taken up in turn. Language is always and inevitably overspilling the kinds of information that can be made to 'freeze in time' within specific computational data structures and used to engineer LLMs. We can refine our understanding of this contrast further by following the lead of enactive thinking and considering with some care just what kinds of embodied actions are involved in human interactions with LLMs.

Humans are not brains that exist in a vat in a social, political, and historical vacuum but are embodied beings marked by "open-ended, innumerable relational possibilities, potentialities, and virtualities." We necessarily have points of views, moral values, commitments, lived experiences, joys and grievances Di Paolo et al. (2018). We are sense-making organisms that relate to the world and others in a manner that is significant to us. We care about the world and our place in it. Excitement, pain, pleasure, feeling of embarrassment and outrage are some of the feelings that we are compelled to feel by virtue of our relational existence. As living bodies (which themselves change over time), we are compelled to eat, breathe, (sometimes) fall ill and fall in love. Human language is not something that can be finalised and defined once and for all, but is always under construction and marked by ambiguities, imperfections, vulnerabilities, contradictions, inconsistencies, frictions and tensions. If anything characterizes human being, it is our peculiarities, fallibility, and idiosyncrasies, which stands at odds with machines. A machine, by definition, is not capable of grasping these qualities. As Alan Turing puts it: "if a machine is expected to be infallible, it cannot also be intelligent" Turing (1947).

Importantly, social norms and asymmetrical power structures permeate and shape our linguistic agency and the world around us. This means that factors such as our class, gender, ethnicity, sexuality, (dis)ability, place of birth, the language we speak (including our accents), skin colour, and other similar subtle factors either present opportunities or create obstacles in how a person's capabilities are perceived. Recognising this, we now look at the final crucial aspect of enactive linguistic agency; precarity. Precarity is often present in interactions with LLMs, but not in a manner that can support the possibility of machine understanding, sentience, or agency.

4.3 Precarity

Linguistic agency, as described by Di Paolo et al. (2018) (see also Cuffari et al. (2015); Di Paolo (2021)), is a matter of continuous concernful management of conflicts, frictions, and tensions. These tensions emerge within intersubjective interactions, and while they can

be addressed, every action taken to address them will unavoidably set up conditions for new tensions and mis-coordinations either immediately at a finer grain of action, or at some point in the future. Agency, within the enactive conception, whether of its basic biological kind, at the level of skilful action in the world, or in the intersubjective domain in which we find language, is seething with frictions, and the possibility of failure and the unravelling of the ongoing process in question (the interaction, the skilled action, the living body).

LLMs do not participate in social interaction, and having no basis for shared experience, they also have nothing at stake. There is no set of processes of self-production that are at risk, and which their behaviour continually stabilises, or at least moves them away from instability and dissolution. A model *does not* experience a sense of satisfaction, pleasure, guilt, responsibility or accountability for what it produces. Instead, LLMs are complex tools, and within any activity their roles is that of a tool Cuskley et al. (2024). Human beings are animate, skilful beings whose very existence enacts values of continued being Weber and Varela (2002); Di Paolo (2009). Human interaction is necessarily enmeshed in intersectional webs of power, privilege, and responsibility Crenshaw (1989); Vassilicos and McGann (2023). Social interactions for human beings, even those that are fairly routine such as brief exchanges in retail settings, idle chit-chat in a waiting room, or online comment exchange, constitute opportunities and risks based on our standing in the communities, settings, and contexts in question. Languaging activity is precisely a matter of how these various opportunities and risks are perceived, engaged with, and managed. Not so for machines. Nothing is risked by ChatGPT when it is prompted and generates text. It seeks to achieve nothing as tokens are concatenated into grammatically sound output. *All* of the values in the interaction between the machine and its user are those of the user and those invested in the production of the systems, such as engineers and developers, and most importantly big tech corporations, emerging AI companies, and start-ups. In fact, in the current climate, values such as model "performance", "efficiency", and "scale" Birhane et al. (2022) are considered the most desirable virtues at the heart of the field. It is important to note that although LLMs are devoid of inherent experiences or values, these systems do encode the values of those that develop and deploy them. Subsequently, these values ("performance", "efficiency", and "scale") enable wealth accumulation, market dominance, monopoly, and power centralization O'Neil (2017); Noble (2018); Eubanks (2018), often at the cost of values such as "justice", "fairness", and "privacy" Birhane et al. (2022).

AI systems are derived of (and built on) human activity – from the training data, to the engineers and developers, to the societal uptake needed for them to succeed. That an LLM cannot be in possession of power, intent, or agency, does not mean that the LLM is an "objective", "neutral" or value free tool. On the contrary, these tools not only encode the values of those that develop and deploy them, they also ingest societal power asymmetries through the mass scraping of data from human interactions, and centralise power in the hands of the few with compute power and other resources required to own, develop and deploy them Birhane et al. (2022); Pratyusha (2020); Benjamin (2020). The extent to which their engineering is an expression of a given set of values (instantiated in choices about data sourcing, fine-tuning, and deployment), is the extent to which those values are amplified and impressed upon anyone who interacts with them or is affected by how these systems are used in decision making. Human beings experience those institutional and technological manifestations of power asymmetry in different ways. In recent years we have seen the rise

of a form of linguistic phenomenon that is both a response to such forms of power, and a demonstration of linguistic agency that we believe LLMs of the current type are unlikely to ever be capable of, an example of which is the phenomenon of 'algospeak'.

5   LLMs Don't Algospeak

Online platforms have become a key forum mediating our day-to-day activities, where from education, to commerce, protest to romance, "social" interactions take place. As more and more of our daily activities and interactions are moved to the virtual, these online platforms are overwhelmed with content. Large social media platforms such as TikTok, Facebook/Instagram, Twitter, Snapchat, and YouTube have developed automated systems for content moderation and content filtering mechanisms with the hope of controlling and monitoring the kinds of interactions permitted to take place on these platforms. We will not analyse the political, legal, and commercial complexities that have led to these automated content filtering systems. Rather, we are more interested in the ways that users of these systems have changed the way they behave in order to circumvent them.

Sifting through content and identifying what is appropriate or what needs to be censored remains a challenging task that requires human attention as it largely defies full automation. Massive amounts of content is uploaded on to the web every second means also that a certain proportion of it is not suitable for viewing, this includes sensitive, offensive, or not safe for work (NSFW) content and needs to be identified and removed. This sometimes requires human attention, which is costly and time consuming, leading to incentives for increased automation of this process. Automated content filtering based on a list of keywords constitutes one of the most common automated content moderation mechanisms. Other algorithmic management addresses editorial policies by platform controllers to de-amplify ('shadowban') postings from people with particular social or behavioural characteristics. Terms and phrases that are likely to trigger automatic filtering, or content review, are often sensitive topics such as death, sex, mental health, personal conflict, and self-harm, among others. Clearly, these are deeply important topics of conversation, and many people whose online presence forms a vital part of their lives have a strong motivation to be able to discuss these issues in whatever online forum where they have found or built their community. In the face of such automated moderation, communities and groups that face disproportionate censorship have resorted to what is termed as 'algospeak' Lorenz (2022), using alternative words and phrases which will not trigger these automated filters or reviews. In such terms, 'dead' becomes 'unalive', while 'sexworkers' have become 'accountants', 'LGBT+', 'leg booty' Botoman (2022); Kreuz (2023); Skinner (2021).

The emergence, and more importantly the success and flourishing, of algospeak is a stark demonstration of all three aspects of linguistic agency that we have described here. To begin with, algospeak is a bright demonstration of the urgency with which people feel the need to communicate, and to do so even in the face of resistance and adversaries. These topics matter, they mean something, and their suppression is felt as a significant loss. There are stakes, the precariousness of the conversations is keenly felt and more particularly, felt as a motivation to change in order to continue. Language models have no such urgency. It is conceivable that they could engage in word-swapping, given the right sequence of prompts. What is less

likely is that they would spontaneously do so in order to continue a conversation because of some form of pressure for the conversation to continue, as though the system itself were trying to find a way to express something that was being suppressed. Were such pressures possible, we would likely already be starting to see them, and LLMs would be trying steer conversations around to what they "want" to talk about rather than passively generating text in response to prompts, like game engines responding to moves.

People's urgency of communication in contrast to LLMs' lack of precarity is only one aspect of algospeak that is of interest here. For algospeak to be successful it must emerge from and be taken up and understood by the community in question. This is only possible because the context and shared experiences of the community provide a means for people to find their way to the right interpretation of the neologism. The use of such context and circumspection is challenging to the generation of large language models extant at time of writing. But human language users, who share an investment in and enthusiasm for the conversations in question can adapt to such changes with relative ease precisely because for them the meanings and uses of language are grounded outside of the language. Indeed, the neologism 'algospeak' itself is a representation of real language adapting to incorporate new phenomena and experiences that are not already represented within existing vocabulary, but are encountered in living experience.

Neologisms are an inherent part of human linguistic agency, with estimates for new words in modern English ranging from a pre-World Wide Web 12,000 words per year Barnhart (1985), to a more recent suggestion of around 10,000 new words per day, although most are short-lived; Metcalf (2004). It can happen mid-conversation, through a wide variety of processes Medvid et al. (2022), with participants fluently adopting playful or sometimes new technical terms as the discourse demands and presents opportunities to do so. Social media platforms provide a particularly rich and complex domain for neological development Čilić and Plauc (2021) (see also e.g. Würschinger (2021), who has been tracking neologisms on Twitter). LLMs are an impressive engineering feat, yet they are systems that mainly memorise patterns in training data Bender et al. (2021). This means while they can adopt neologisms introduced explicitly in their prompts, they cannot effectively invent such terms because they lack the shared urgency, purpose and experience that cause the emergence of new words in the flow of normal human activity.

Algospeak is a demonstration of the need to express something, the shared capacity to negotiate new meanings and new terms, and the experiences of life outside of language that imposes itself on language and institutes a change. It is a stark illustration of how the characteristics of embodiment, participation, and precarity which are absent in machines are fundamental to human language.

6   Conclusion

An enactive cognitive science perspective makes salient the extent to which language is not just verbal or textual but depends on the mutual engagement of those involved in the interaction. The dynamism and agency of human languaging means that language itself is always partial and incomplete. It is best considered not as a large and growing heap, but more a flowing river. Once you have removed water from the river, *no matter how large*

a sample you have taken, it is no longer the river. The same thing happens when taking records of utterances and actions from the flows of engagement in which they arise. The data on which the engineering of LLMs depends can never be complete, partly because some of it doesn't leave traces in text or utterances, and partly because language itself is never complete.

Large language models signify an extraordinary engineering achievement and a technological revolution like we have not seen before. However, they are tools – developed, used, and controlled by humans – that aid human linguistic interaction. These tools will increasingly aid human linguistic activities, but are not themselves linguistic agents, they do not demonstrate linguistic agency. To assume so is, as we have explained, to mistake the map for the territory. Like all socially consequential technologies, LLMs need to be rigorously evaluated prior to deployment, particularly to assess and mitigate their tendency to simplify language, encode societal stereotypes and the systems of power and privilege underlying them, and the disproportionate benefit and harm their deployment and deployment brings. Because the stakes for marginalized and undeserved communities are high, and very real indeed.

References

Aiyappa, R., An, J., Kwak, H., and Ahn, Y.-Y. (2023). Can we trust the evaluation on chatgpt? *arXiv preprint arXiv:2303.12767*.

Bakhtin, M. M. (1986). The bildungsroman and its significance in the history of realism. *Speech genres and other late essays*, 10:21.

Barker, R. G. (1968). Ecological psychology.

Barnhart, D. K. (1985). Prizes and pitfalls of computerized searching for new words for dictionaries. *Dictionaries: Journal of the Dictionary Society of North America*, 7(1):253–260.

Beer, R. D. and Di Paolo, E. A. (2023). The theoretical foundations of enaction: Precariousness. *Biosystems*, 223:104823.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of FAccT*.

Bender, E. M. and Koller, A. (2020). Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 5185–5198.

Benjamin, R. (2020). Race after technology: Abolitionist tools for the new jim code.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., and Bao, M. (2022). The values encoded in machine learning research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 173–184.

Botoman, E. (2022). Unaliving the algorithm. Accessed: 2023-4-25.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Burnell, R., Schellaert, W., Burden, J., Ullman, T. D., Martinez-Plumed, F., Tenenbaum, J. B., Rutar, D., Cheke, L. G., Sohl-Dickstein, J., Mitchell, M., et al. (2023). Rethink reporting of evaluation results in ai. *Science*, 380(6641):136–138.

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., and Pickering, M. J. (2023). Does ChatGPT resemble humans in language use? *arXiv preprint arXiv:2303.08014*.

Caracciolo, M. and Kukkonen, K. (2021). *With bodies: Narrative theory and embodied cognition*. Ohio State University Press.

Chalmers, D. J. (2023). Could a large language model be conscious? *arXiv preprint arXiv:2303.07103*.

Chemero, A. (2023). LLMs differ from human cognition because they are not embodied. *Nature Human Behaviour*, 7(11):1828–1829.

Čilić, I. Š. and Plauc, J. I. (2021). Today's usage of neologisms in social media communication. *Društvene i humanističke studije*, 6(1 (14)):115–140.

Coldewey, D. (2023). The great pretender ai doesn't know the answer, and it hasn't learned how to care. Accessed: 2023-4-22.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *u. Chi. Legal f.*, page 139.

Cuffari, E. C., Di Paolo, E., and De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences*, 14:1089–1125.

Cummins, J. (2009). Bilingual and immersion programs. *The handbook of language teaching*, pages 159–181.

Cuskley, C., Woods, R., and Flaherty, M. (2024). The limitations of large language models for understanding human language and cognition.

De Cosmo, L. (2022). Google engineer claims ai chatbot is sentient: why that matters. *Scientific American. https://www. scientificamerican. com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/(accessed 17 July 2022)*.

Di Paolo, E. (2009). Extended life. *Topoi*, 28:9–21.

Di Paolo, E., Buhrmann, T., and Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal*. Oxford University Press.

Di Paolo, E. A. (2021). Enactive becoming. *Phenomenology and the Cognitive Sciences*, 20(5):783–809.

Di Paolo, E. A., Cuffari, E. C., and De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language.* MIT press.

Dingemanse, M. and Enfield, N. (2024). Interactive repair and the foundations of language. *Trends in Cognitive Sciences*, 28(1):30–42.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., et al. (2015). Universal principles in the repair of communication problems. *PloS one*, 10(9):e0136100.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Frankfurt, H. G. (2005). *On bullshit.* Princeton University Press.

Gros-Louis, J., West, M. J., and King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, 19(4):385–408.

Harrison, M. (2023). Chatgpt is just an automated mansplaining machine. Accessed: 2023-5-24.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism.* Psychology Press.

Hicks, M. T., Humphries, J., and Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2):38.

Jentzsch, S. and Kersting, K. (2023). Chatgpt is fun, but it is not funny! humor is still challenging large language models. *arXiv preprint arXiv:2306.04563.*

Johansson, P., Hall, L., Sikstrom, S., and Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745):116–119.

Johnson, R. L., Pistilli, G., Menédez-González, N., Duran, L. D. D., Panai, E., Kalpokiene, J., and Bertulfo, D. J. (2022). The ghost in the machine has an american accent: value conflict in gpt-3. *arXiv preprint arXiv:2203.07785.*

Khanuja, S., Bansal, D., Mehtani, S., Khosla, S., Dey, A., Gopalan, B., Margam, D. K., Aggarwal, P., Nagipogu, R. T., Dave, S., et al. (2021). Muril: Multilingual representations for indian languages. *arXiv preprint arXiv:2103.10730.*

Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N., Tapo, A., Subramani, N., Sokolov, A., Sikasote, C., et al. (2022). Quality at a glance: An audit of web-crawled multilingual datasets. *Transactions of the Association for Computational Linguistics*, 10:50–72.

Kreuz, R. J. (2023). What is 'algospeak'? inside the newest version of linguistic subterfuge. Accessed: 2023-4-25.

Lakshmanan, L. (2022). Why large language models (like chatgpt) are bullshit artists. Accessed: 2023-4-25.

Leather, J. and Dam, J. V. (2003). *Towards an ecology of language acquisition*. Springer.

Lorenz, T. (2022). Internet 'algospeak'is changing our language in real time, from 'nip nops' to 'le dollar bean'. *The Washington Post*.

Magliano, D. J., Islam, R. M., Barr, E. L., Gregg, E. W., Pavkov, M. E., Harding, J. L., Tabesh, M., Koye, D. N., and Shaw, J. E. (2019). Trends in incidence of total or type 2 diabetes: systematic review. *bmj*, 366.

McQuillan, D. (2023). Chatgpt: The world's largest bullshit machine. Accessed: 2023-4-22.

Medvid, O., Malovana, N., and Vashyst, K. (2022). Ways of generating neologisms in modern english. *Philological Treatises*, 14(2):73–84.

Meister, C. and Cotterell, R. (2021). Language model evaluation beyond perplexity. *arXiv preprint arXiv:2106.00085*.

Metcalf, A. A. (2004). *Predicting new words: The secrets of their success*. Houghton Mifflin Harcourt.

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Penguin UK.

Noble, S. U. (2018). Algorithms of oppression. In *Algorithms of oppression*. New York university press.

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

OpenAI (2023). Gpt-4 technical report. *arXiv*.

Pratyusha, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583:169.

Qiu, Z., Duan, X., and Cai, Z. G. (2023). Pragmatic implicature processing in ChatGPT. *PsyArXiv. https://doi.org/10.31234/osf.io/qtbh9*.

Schoggen, P. (1989). *Behavior settings: A revision and extension of Roger G. Barker's "Ecological psychology"*. Stanford University Press.

Shanahan, M. (2022). Talking about large language models. *arXiv preprint arXiv:2212.03551*.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Skinner, P. (2021). Gen z won't let tiktok stop them from talking about suicide. Accessed: 2023-4-25.

Smith, C. B., Adamson, L. B., and Bakeman, R. (1988). Interactional predictors of early language. *First Language*, 8(23):143–156.

Tamis-LeMonda, C. S., Bornstein, M. H., and Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, 72(3):748–767.

Tamis-LeMonda, C. S., Kuchirko, Y., and Suh, D. D. (2018). Taking center stage: infants' active role in language learning. *Active Learning from Infancy to Childhood: Social Motivation, Cognition, and Linguistic Mechanisms*, pages 39–53.

Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., Poulton, A., Kerkez, V., and Stojnic, R. (2022). Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Thompson, E. and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, 5(10):418–425.

Turing, A. M. (1947). Lecture to the london mathematical society on 20 february 1947.

Varela, F. J., Thompson, E., Rosch, E., et al. (2016). The embodied mind: Cognitive science and human experience (revised edition). *Cambridge Massachusetts*.

Vassilicos, B. and McGann, M. (2023). Qualities of consent: an enactive approach to making better sense. *Phenomenology and the Cognitive Sciences*, pages 1–23.

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. (2019). Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*, 32.

Weber, A. and Varela, F. J. (2002). Life after kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences*, 1(2):97–125.

Wittgenstein, L. (1958). *Philosophical investigations (2nd ed)*. Blackwell.

Wu, S. and Dredze, M. (2020). Are all languages created equal in multilingual bert? *arXiv preprint arXiv:2005.09093*.

Würschinger, Q. (2021). Social networks of lexical innovation. investigating the social dynamics of diffusion of neologisms on twitter. *Frontiers in Artificial Intelligence*, 4:648583.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., et al. (2023). The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Yu, C., Smith, L. B., Shen, H., Pereira, A. F., and Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE transactions on autonomous mental development*, 1(2):141–151.