

Some hypotheses on how Chatbots work in problem-solution-driven conversations: Large Language Models as confirmation of the Innovation Illusion

S.F.M. van Vlijmen and H.D. Lethe jr
bas@basvanvlijmen.com

Submitted: 3 May 2026
Revised: 16 August 2026

Abstract

We discuss the nature of chatbots as conversation partners in problem-solving conversations. What can chatbots do and what can't they do? Our analysis draws on insights from Aggregation Dynamics, Cognitive Linguistics, Neuropsychology and Psychology.

We establish that chatbots are multifaceted and composite systems. Our argument focuses on basic chatbots in the hope of thereby making statements about the core functionality of more advanced chatbots. Basic chatbots are assumed to consist of a Large Language Model (LLM) with a simple interface.

The main results of our analysis are: a description of human imagination, understanding and thinking based on so-called metaphorical problem propagations; that the texts in text datasets used for training LLMs have specific characteristics and that these texts only partially imitate human thinking and understanding; that the LLM training process encodes artificial metaphorical problem propagations into an LLM from these text datasets.

Our conclusions are that a basic chatbot cannot be a thinking partner capable of matching the cognitive flexibility of humans, and that further development of LLMs will not lead to this either.

But chatbots exist, they are being used on a massive scale, by both individuals and organisations. It is therefore socially and politically important to understand them. Our article aims to contribute to the discussion on the functioning, benefits and drawbacks of chatbots.

Cognitive Linguistics shows how the use of metaphor is an expression of our thinking. Aggregation Dynamics, is an attempt at a comprehensive systems theory. We believe that the concept of metaphorical problem propagation could provide an interesting addition for both.

Chatbots a solution? For what?

1 Introduction

Chatbots are winning people over! Why, and how? To many users, chatbots come across as fully-fledged, useful and reliable conversation partners. That is undoubtedly an exceptional achievement on the part of their creators. But the success of chatbots also raises many questions.

The seemingly creative and intelligent behaviour of chatbots is, for the time being, difficult to define, verify and explain. What are their characteristics, what exactly can chatbots do, and what can't they do? They are certainly robust. Yet they are also unreliable in various ways. What tasks can be safely entrusted to them, and what absolutely cannot? These issues have attracted the attention of scientists, legislators and philosophers [9, 86, 66, 72, 71, 24, 23, 13].

In our article, we aim to provide an insight into how Large Language Model-based chatbots work and into the degree of freedom in problem-driven conversations with these chatbots, by combining insights from four different fields of research. The four fields are: Aggregation Dynamics, Cognitive Linguistics, Neuropsychology and Psychology. We emphasise it's about problem-driven conversations because, we suspect, it is in this type of human-chatbot conversation that what a chatbot can and cannot do becomes most clearly apparent.

Aggregation Dynamics (AD): Innovation Illusion

AD is an extension of ecology. Ecology studies the distribution and concentration of organisms based on their characteristics and interactions, both amongst themselves and with their environment [7]. AD regards organisms and the environment as impractical abstractions when seeking to understand systems comprising people, products, organisations and cultures. The AD strives for an ecology without exceptions; it strives for totality [82]. The AD believes that this requires a much broader range of interacting players. AD refers to this broader category of players as aggregations. An aggregation is any *thing* that is experienced by a player as a player. In the AD, for example, the following are all treated as equal players: projects, traffic jams, a combination of a cake plate with crumbs from a biscuit and a used coffee cup from a different set, bacteria in their agar-agar dish, the history of political dossiers, the jealousy of the gods and, of course, chatbots. To understand what concerns people and what they do, the AD looks, amongst other things, at what people perceive as opportunities or obstacles, what they identify as problems, and what solutions they devise and implement. The AD argues that people are less innovative in identifying problems and devising solutions than they themselves believe. The AD refers to this as the Innovation Illusion.

Cognitive Linguistics: Metaphor Paradigm

Cognitive Linguistics is fascinated by the relationship between language and thought. Metaphors are said to play a key role. Analyses in Cognitive Linguistics show that metaphors are much more than just exceptional poetic devices. In fact, our thoughts and our texts are teeming with them. Cognitive Linguistics argues that metaphors are the driving force behind human thought.

Neuropsychology: Predictive Processing Theory

A long-accepted theory of perception, thought and action posits that the eyes, and other familiar senses, form the basis of a process that generates a mental and detailed picture of the world. With that picture, and after reflecting on it, one can act. In Predictive Processing Theory, by contrast, the brain already forms hypotheses about the situation in which the thinking body finds itself. Perception, via many more sensory sources than just the familiar senses, serves to subsequently test or adjust those hypotheses. In this theory, the distinction between thinking and acting is fluid. Acting is a form of thinking that serves to refine perception, for example by looking more closely, or to intervene in the world and thereby bring it into line with the hypotheses about it.

Psychology: Thinking Styles

Psychology provides us with definitions of thinking, methods of thinking, and the classification of human thinking into a fast-satisfaction thinking style (System 1) and a systematic analytical thinking style (System 2).

An outline of the argument

The British cognitive scientist Margaret Boden described the field of Artificial Intelligence (AI) as follows: “[it] seeks to make computers do the sorts of things minds can do” [12]. This brief characterisation serves as the starting point for our argument. If we are to say anything about chatbots, we must first have some understanding of those *minds*. The speculative argument consists of four parts, which we summarise briefly below.

1. We formulate a model of human beings’ more or less conscious mental spaces concerning the current situations in which they find themselves, their memories and their plans. These mental spaces are associative and flexible, and the mind can ‘wander’ through them, tinker with them, zoom in and out, move from the concrete to the abstract and back again, and these mental spaces can also be expanded or restricted at will. We call such mental spaces: *metaphorical problem propagations*. Our model arises from the integration of the conceptual system of Cognitive Linguistics (concept, experience) with the problem and solution dynamics of Aggregation Dynamics (motivation, direction, interaction) and the prediction hypothesis of Predictive Processing Theory (perception, thinking, action).

2. Large Language Models are primarily based on written texts available on the internet. These texts are the result of a process of reduction. This is because texts, whether spoken or written, are always the result of a process of selection and combination. Furthermore, written texts lack context and physicality. We therefore argue that most written texts from datasets only partially correlate with human mental capabilities.

3. We propose that a substantial proportion of the texts on which LLMs are trained concern the relationship between problems and solutions. Texts have always been reduced material, and these texts also have a limited repertoire of structural patterns related to the relationship between problems and solutions.

Our hypothesis here is that the training process in a Large Language Model constructs fragments of artificial metaphorical problem propagations. After training, in a conversation, these propagations are activated to a greater or lesser extent and thus, as it were, act as riverbeds, exerting a pull on the course of word generation.

A problem-motivated conversation can now be defined as a path through metaphorical problem propagations, both on the chatbot’s side and on the user’s side. But these propagations differ from one another. The final step of the argument concerns these differences.

4. We argue that a basic chatbot cannot simulate analytical thinking, because it ‘tends’ to seek out the ‘riverbeds’ of the artificial metaphorical problem propagations, and these represent a reduced form of thinking. Since chatbots do not possess a body and have no experience, they are unable to assess and correct their own propagations in terms of effectiveness. That does not mean that a chatbot is incapable of providing certain forms of cognitive support, but this must be guided by an alert and creative prompt writer.

Structure of the article

Sections 2 through to 6 provide introductions tailored to this article on: AI, chatbots and neural networks; the insights drawn from Aggregation Dynamics, Cognitive Linguistics, Neuropsychology and Psychology. Anyone familiar with this material may skip it and start in Section 7.

The main argument, in the four steps described earlier, follows in Section 7 through to Section 10. In the conclusion, Section 11, we summarise everything.

2 AI, chatbots and LLM modelling

Artificial intelligence (AI) is a multidisciplinary field. And, relative to the current hype, it is also an old field. The development of AI went hand in hand with the development of automatic calculators (what we now call computers) from the 1930s onwards [61, 1]. The question of what computers are capable of was already being asked at that time and has occupied generations of researchers and engineers. A prime example of this is the work of Alan Turing. For an overview, see *Mind As Machine: A History Of Cognitive Science* by the British philosopher and cognitive scientist Margaret Boden [11]. The term *artificial intelligence* dates from 1956 [11, p. 331]; previously, the term ‘computer simulation’ was used.

According to Boden, there are five main schools of thought in AI theory [12]. These relate to the way in which information is represented and processed. One of them is inspired by data processing in living organisms: fault-tolerant, parallel and distributed. Artificial neural networks stem from this school of thought, also known as connectionism.

Artificial neural networks are modelled on the neural network structures in the brain and the way in which neurons within them receive and process chemical signals from one another, and then transmit the result to organs or other

neurons. Artificial neural networks work in a similar way. An artificial neuron reads numerical signals from the input ports through which it is connected to other neurons, performs a calculation using these, and outputs the result via its output ports.

How do the neurons in a neural network ‘know’ which calculation to perform? Initially, they know only a very small part of this. It requires a training process to teach them the details. In this process, each neuron and the network as a whole learn, step by step, the desired relationship between input and output. For example, that an image with a cat prominently in the centre corresponds to the word ‘cat’, and an image with a dog corresponds to the word ‘dog’.

To put it very simply, you could say that a neural network implements a lookup table. That sounds cumbersome, and it often is, for example, for a telephone directory. But neural networks also have major advantages. It is a one-size-fits-all computational model: in principle, you can teach the same neural network to function as a telephone directory, but also to answer questions as a helpdesk. Furthermore, neural networks can withstand some damage. Imagine a large neural network distributed across 100 computers. If one of those computers fails, the network is not necessarily rendered unusable straight away. Moreover, neural networks are not fussy about their input. An example is offered by the well-known app Pl@ntNet. This app has been trained to identify a plant based on a photo of it. If you present it with a very blurry photo of a plant, the app will still try to say something meaningful about it. Both of these latter properties are referred to as robustness.

However, neural networks also have disadvantages: they are unreliable. Pl@ntNet could, so to speak, look at a photo of a bicycle and claim with 100% certainty that it is an oak tree. For further details, we refer to the extensive literature on this subject, for example [61].

A Large Language Model is a specific type of neural network. These networks are not trained to link images to the words ‘cat’ or ‘dog’, but are trained on text and text that plausibly connects to it. For example, if you provide an LLM with the word ‘I’, there is a good chance it will return ‘am’. Because ‘I am’ is a very common sentence or the beginning of one. A Large Language Model forms the heart of a chatbot.

2.1 Chatbots and basic chatbots

By the term ‘chatbot’, we refer to systems accessible to a wide audience, such as ChatGPT, Claude, Grok, Lumo and so on. These are not designed to support very specific tasks; they are trained on bulk text from the internet.

Nevertheless, chatbots are difficult to categorise under a single heading. They are, to a large extent, multifaceted and complex or hybrid systems, which are also still very much under development. They are often combinations of different software components based on different approaches and different architectures. The most important part of a chatbot is the Large Language Model (LLM). Other components in a chatbot act as an interface, managing the interaction between the LLM and the user, providing additional functions and

moderating the process. We could call that a software wrapper around the LLM. This ‘wrapper’ ensures that the LLM does not output text deemed undesirable, such as racist language or advice on suicide. Other components control the LLM when summarising text or adjusting the style of a text [6, 2, 80, 36].

It is likely that LLMs will give rise to hybrid forms, because a natural language interface to all manner of tools is, of course, highly appealing. The American mathematician, author and software developer Stephen Wolfram describes the combination of the Wolfram—Alpha computational system and a chatbot [90]. There are many more examples [30, 28]. Exactly how the current public chatbots are constructed is impossible to say for outsiders, who are far removed from the closed-off development process.

The hybrid nature of chatbots as assumed by us, has implications for our argument. For which functionality of which component should we be discussing? Let us define a basic chatbot as consisting of a Large Language Model with a simple interface. When we refer to chatbots and conversations hereafter, we mean chatbots limited to the functionality of these hypothetical basic chatbots. We assume that this abstraction is meaningful. That is to say, statements about basic chatbots are also relevant to their forms expanded with additional functionality. What we discuss in this article essentially concerns what an LLM is capable of and what that means for its current use and future potential.

The Large Language Model is the text-generating engine of a chatbot. Given any input text, the LLM is able to predict, on statistical grounds, what text might follow. The process of generating that text takes place in word fragments, known as tokens, or by words, or in even longer sequences. The LLMs of the aforementioned current chatbots generate text per token. We shall continue with the previous example, in which, for the sake of simplicity, we consider the generation process to be ‘per word’. If ‘I’ is the input, then ‘am’ will score highly in the output, but ‘have’ is also a good candidate. In contrast, ‘traffic sign’ or ‘elephant’ will score low. Because ‘I traffic sign’ is not a typical sentence, except perhaps in absurdist theatre. So there is not just one possible continuation, but rather a long series of possibilities, each with its own probability. A choice is made from that series.

Of course, it may also turn out that there is no plausible continuation, according to the chatbot. In that case, the text generation process stops or the shell surrounding the LLM intervenes. On what grounds? Stephen Wolfram states regarding the selection rules and configuration rules of an LLM [90]:

Particularly over the past decade, there’ve been many advances in the art of training neural nets. And, yes, it is basically an art. Sometimes—especially in retrospect—one can see at least a glimmer of a “scientific explanation” for something that’s being done. But mostly things have been discovered by trial and error, adding ideas and tricks that have progressively built up a significant body of knowledge about how to work with neural networks.

An LLM is a product whose functioning remains largely unexplained; see Section 2.2. Nevertheless, it performs very convincingly in conversations. But

the question remains: how?

LLMs are trained on text fragments and their plausible textual continuations. That works perfectly well. Chatbots effortlessly produce fluent sentences. That was exactly what researchers and developers had hoped for, but they are capable of more.

To many people’s surprise, it turned out that LLMs are also capable of what appears to be creative reasoning. British researcher Geoffrey Hinton, a leading figure in AI, cites as an example the chatbot’s response to his question: ‘What is the similarity between a haystack and an atomic bomb?’ [5]. The chatbot replied that both are self-sustaining combustion processes. That is a striking analogy. A form of analogy-making ability had previously been observed in the semantic networks that encode words, which are components of LLMs [61, p. 248]. Research into the capacity for analogy-finding is a topical subject within AI research [51, 69, 77]. We assume that LLMs possess some ‘sensitivity’ in this area; this observation is sufficient for our argument.

Neural networks also have other desirable properties, such as robustness; and less desirable ones, such as unreliability. These properties play a role at the conclusion of the argument, namely in Section 10. We will explain what we mean by unreliability there.

2.2 *Interpretability*: neural network models

The question of the ‘how’ of neural networks is the subject of the research field *Interpretability* [86]. Within this field, researchers attempt to explain the emergent behaviour of neural networks in terms of the behaviour of individual neurons by creating models of neural networks and by design modifications. Leonard Bereska and Efstratios Gavves describe the state of the art for LLMs in *Mechanistic Interpretability for AI Safety – A Review* [9]. The authors note that developments in interpretability have parallels with those in Psychology.

Psychology began by observing behaviour and developing theories, as the techniques for measuring or simulating brain processes did not yet exist. Step by step, these techniques were developed within new disciplines such as Cognitive Science, Neuropsychology and Neurology. Research into AI interpretability exhibits a similar spectrum of approaches. Bereska and Gavves distinguish four of them, which they describe as follows (our italics): “*Behavioral* analyzes input-output relations; *Attributional* quantifies individual input feature influences; *Concept-based* identifies high-level representations governing behavior; *Mechanistic* uncovers precise causal mechanisms from inputs to outputs.”

Our argument focuses largely on concepts and the relationships between them. See Section 3.4 and the discussion of the conceptual system in Cognitive Linguistics. To us, the results from Concept-based and Mechanistic interpretability seem to align most closely with our argument. By ‘measuring’ within the network, interpretability researchers attempt to isolate network fragments that correspond to specific concepts and activation patterns that can be traced back to behaviour of the artificial neural network [22].

There are clear indications that concepts are not distributed randomly across the network, but can be localised as clusters. Furthermore, related concepts may overlap or lie in close proximity to one another. This does not mean that all concepts are stored in this way. However, we assume that overlap means that if, of related concepts A and B, A is activated, the likelihood increases that B will also be activated. The occurrence of this overlap is understandable: it saves storage space and thereby increases precision, which is precisely what the training of a neural network aims to achieve.

3 Aggregation Dynamics

The philosopher Karl Popper titled a collection of essays *All Life Is Problem Solving* [76, p. 99]. We see it that way too. The interactions between aggregations are about problems. Rather than tackling ‘life’ – that is, all the problems of all aggregations – head-on, Aggregation Dynamics makes a start by investigating problems and solutions as experienced by people.

3.1 Problem and problem position

Imagine you experience something that doesn’t align with your wishes, needs or ambitions. We call that a problem. But the problem only becomes concrete once you articulate or state it: you adopt a problem position. A problem position sets out in detail what we see as the relevant aspects when thinking about and/or communicating on the matter. Typical elements of a problem position are: pressing issues, classifications into main and subproblems, potential avenues for solutions, phasing, background knowledge, facts, rumours, opinions or views of other stakeholders, experiences with certain solutions, priorities, the current status of the implementation of certain subproblems, explanatory histories. Note that this involves weighing up potential solutions, but not the implementation of the solution step itself. A problem situation is a snapshot in time where one pauses for a moment and, for example, asks: ‘Where do we stand now?’

A problem position does not necessarily present a consistent picture. For example, one adviser might say ‘do it’, whilst another says ‘don’t do it’. In fact, the position is probably usually inconsistent, except in the case of extremely trivial problems. It is often precisely this lack of consistency that contributes to the sense that there is a problem.

The AD also claims that it makes sense to distinguish between three types of problem position: the natural, the internal and the external [85]. The natural problem position is one that extends deep into the physical state of a person reflecting on a problem position. This is only partially accessible to consciousness. The internal one is the conscious mental representation. The external one is still a mental representation, but tailored to a recipient. There may be several of these: ‘if someone asks how the project is going, I say *this* to A and *that* to B.’

The AD claims that internal and external problem positions are manageable in terms of scope and complexity [85]. This applies both to the mental representation and to a problem position expressed in words. It is assumed that a simple problem-position jargon, a kind of formulaic language, is sufficient for the description. A problem statement described in that jargon covers the essentials in a concise, matter-of-fact manner. Thus, problem statements concern a limited number of observations, assertions, facts, events from the past deemed relevant, potential next steps, actors and a limited number of relationships between all of these.

There are no rigid boundaries to this mental representation, nor are there any formal requirements. It is what an individual, entirely on a personal level, experiences as a whole and what they consider to be of significance. The words of the American philosopher John Dewey (1859-1952), as quoted by Mark Johnson, seem apt here. The problem position contains that which intuitively aligns with the “pervasive unifying quality” of the moment [37, p. 52].

Finally, the AD assumes that, provided one works carefully, verbalised versions of problem positions can provide a picture of the complexity of a person’s internal problem position [85]. That is to say, through careful interviewing, or by asking someone to work out an internal problem position honestly and fully, or through careful archival research, one can obtain a reasonably clear picture that correlates in terms of complexity with the complexity of an internal problem position. But the natural problem position is assumed to remain largely out of reach.

Although language is a tool for gaining an understanding of the complexity of an internal problem position, this does not mean that just any text already offers the necessary depth. In Section 8, we explore this further and will argue that much of the text produced by humans lacks the complexity of problem positions. Our claim is: most text represents a situation as stable and understood, and then directs the reader towards a goal that takes into account only a limited number of variables.

3.2 Problem propagation

One could describe human behaviour by stating that people move from one problem state to another via solutions or steps in that direction. Leaving aside the question of whether the measures observed are well thought out or not, one could make the following observations. For trivial problems, a single solution step may be all that is needed. For more complex problems, a series of steps often needs to be taken. These are usually not simple linear series of steps. The approach to one problem may clash with the approach to another problem. Or, the approach may seem more complicated than first thought and necessitate splitting the problem into several subproblems. Problem positions are therefore linked through various relationships, through mutual dependencies and through solution steps.

Aggregation Dynamics refers to such a network of problem positions that merge into one another through solution steps as a ‘problem propagation’. Prop-

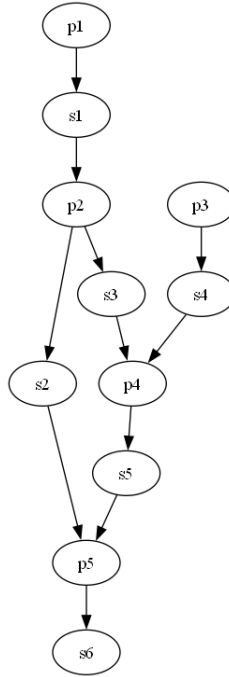


Figure 1: An example of abstract problem propagation. Here, the p’s represent problems and the s’s represent solutions.

agation should be understood as proliferation, expansion or growth: both the approach to a problem and its solution can lead to new problems. Simply because the world does not stand still.

When one maps out a problem propagation, one can look back (history) and look ahead (plans, scenarios). The ‘present’ in a network is referred to by AD as the problem front. At the front, problem positions can collide and choices are made.

Problem propagations can be represented graphically and/or in text. Figure 1 illustrates an abstract problem propagation, with the p’s as problems and the s’s as solutions.

It doesn’t have to be so abstract. The representation of a problem propagation is not subject to any formal requirements [84, p. 7]. A problem propagation can be elaborated to a greater or lesser extent with the details of a specific solution process. In Figure 2, the earlier propagation is supplemented with concrete information drawn from the Los Angeles Smog case [85, 35, 18]. To help the reader understand the annotations in Figure 2, we will briefly explain what the Los Angeles Smog case is about.

Around 1943, air pollution in the industrial city of Los Angeles and the surrounding towns reached alarming levels. It became an issue that would dominate

local political debate for decades, and which had a major influence on technical developments and environmental thinking both nationally and internationally. Initially, it was thought that the smog was caused by the use of coal for heating, power generation and industrial processes. Such smog was familiar from places such as St. Louis, Pittsburgh and London. But after years of research, the main cause in Los Angeles turned out to be of an entirely different nature: the incomplete combustion of petroleum products in car engines. In summer, the volatile hydrocarbons did not blow away but lingered over the cities due to meteorological conditions and the mountains surrounding Los Angeles County. The abundant sunshine now had free rein to cause the hydrocarbons to react into all sorts of nasty substances: photochemical smog. It took many years to bring this problem under control, for a great many reasons. For example, the car-addicted and vociferously complaining public refused to believe that they themselves were partly to blame. Reluctantly, the oil and automotive industries set about adapting their products to achieve better combustion. One example of this is the use of a catalytic converter in the exhaust (the precursor to our current three-way catalytic converter). However, this turned out to require unleaded petrol. But that lead was in petrol to prevent premature ignition in the cylinders, known as knocking. Without lead, knocking had to be prevented in some other way, and so on. The Los Angeles Smog case is rife with problems multiplying.

Just as with problem positions, problem propagations can also be described with varying degrees of accuracy or in varying degrees of detail. That said, according to the AD, it is never complete. Language always falls short. For example, because internal problem positions flow in an inimitable way into unattainable, deeper natural problem positions. That does not detract from the fact that problem propagations can be worked out in great detail by incorporating complete problem positions and unravelling the solution steps much further.

3.3 The Innovation Illusion

People seek stability, value predictability and like to work according to established patterns. Standardisation and working to procedures are the result of this. The Innovation Illusion argues that people regularly tend towards predictability when solving problems, whilst believing they are actually being innovative. This tendency towards predictability and repetition only becomes apparent when one looks more closely at the pattern of interaction and abstracts from details that have no causal significance, from the domain, the players, and from scales of time and space.

The Innovation Illusion posits problem types as a finite set of player, scale-, time- and domain-independent abstract undesirable situations and abstract solutions for them. The set of solutions is also assumed to be finite. The set is called the repertoire of the problem type; we sometimes also refer to solution types.

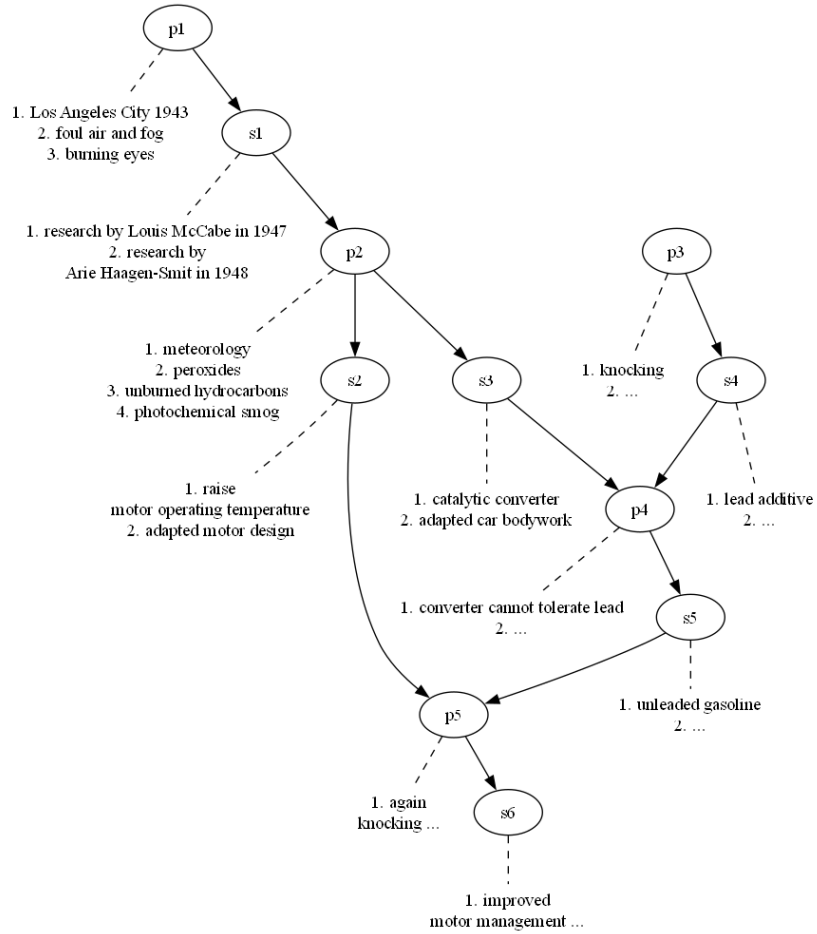


Figure 2: An example of problem propagation and a suggestion of some context from the Los Angeles Smog case, as seen from the perspective of a fictional observer.

An example is the problem type congestion. In the Netherlands, there is both a problem with traffic congestion on motorways and an overload on the national electricity grid. Both problems are instances of the congestion problem type. The solutions to these problems show many parallels. Two examples: additional lanes in one domain correspond to thicker cables in the other; avoiding rush hour is comparable to ‘a request for limited electricity consumption by households between 9 am and 4 pm’. These are analogous situations and interventions. In terms of the type of problem, we are dealing with an aggregation (road/cable) with a function (facilitating transport) that, compared to other aggregations (road users/electricity users), is unable to fulfil its role sufficiently.

The Innovation Illusion is likely exaggerating, but where the line between repetition and innovation lies remains unclear for the time being. For our argument, it is important to assume that in most problem formulations and the subsequent solution processes, the tried and tested path is followed.

3.4 Problem propagation and its patterns

An important question is: are there patterns in problem propagations? The hypothesis, from a previous section, that descriptions of problem positions consist of a finite number of statements seems to suggest as much. The Innovation Illusion, explained above, significantly reinforces that suspicion.

The Innovation Illusion asserts that the number of problem types is finite and that the corresponding repertoires are also of finite size. Evidently, a problem propagation abstracted down to the problem-type level is then full of repetition and patterns. For, suppose there are n problem types; then the next problem in a problem propagation that goes n problem steps deep must be of a type that has already occurred. For example, the maintenance of a complex device leads to the deployment of a particular expert, and that expert must also maintain her own equipment.

It follows from the above that, when considering problem positions and problem propagations relating to concrete problems, one can also envisage more abstract variants. Step by step, one can work towards representations at the type level, or even more abstractly, types of types. And from there, one can of course also go back, namely by detailing more concrete cases.

4 Cognitive Linguistics

Here we outline the key concepts from Cognitive Linguistics that are relevant to us, as developed in particular by the linguist George Lakoff and the philosopher Mark Johnson.

Everyone is familiar with metaphors and analogies as figures of speech. Both refer to the way in which we understand and experience a thing, behaviour or situation in terms of another thing, behaviour or situation. Cognitive Linguistics favours the term ‘metaphor’. The metaphor perhaps has a more poetic connotation, in which the choice of words carries somewhat more weight than

in the analogy. The latter is perhaps more descriptive, and the chosen words are instrumental; in other words, it can be put differently. Be that as it may, the metaphor or analogy sometimes comes in very handy for making something clear.

Sometimes? No, the opposite turns out to be true. Virtually every sentence contains a metaphor because we cannot understand or express ourselves without making comparisons. This insight gained widespread recognition with the publication of Lakoff and Johnson's *Metaphors We Live By* in 1979. This theory is of great importance to our argument. Texts are rich in layered meaning and give an insight into how we think. The field of research that Lakoff and Johnson helped shape is now referred to as Cognitive Linguistics [25, 26, 14]. We will discuss the central elements, which are the most important to us.

Cognitive Linguistics holds that human mental activity – thinking, assigning meaning, and motivated action – is profoundly shaped and constituted by bodily experiences, movements, sensory stimulation, emotions, feelings and actions. There is a constant interaction between the environment and the body-brain.

The concepts, their properties and relationships, with which we are so familiar in our daily lives, are not objective facts, but are learned step by step. This begins as early as infancy with our physical experiences, known as perceived correlations: if this happens, then that happens too. Direct physical experience leads to what are known as primary concepts or metaphors. Some examples are: *Knowing Is Seeing*, *More Is Up* and *Speech Act is Physical Force*. We shall explain them.

Knowing is understood in terms of *Seeing*. As a child, you learn that looking more closely helps you understand better. You learn this even before you can put it into words. Once you have mastered language, you understand expressions such as: 'I see what you mean', 'Could you elucidate this aspect further' or 'I don't find the conclusion clear yet'. *More Is Up*, for example, is something you learn when playing with water. You see that the level in your cup rises when you add more water. You learn the fundamental concept that *Speech Act Is Physical Force* when your parents grab hold of your body and move it, for example your hand, whilst giving a verbal instruction: 'Move your hand! It's hot!'

The metaphor is the mental tool for giving meaning to new experiences and situations based on concepts, possibly primary concepts, that you already know. Anyone who conceives a new concept for the first time or learns it from another person will initially be aware of the metaphor, but as soon as the concept becomes more familiar, it takes on a life of its own and you are no longer aware of the metaphor [47]. Higher-order concepts, i.e. those above the primary level, are called conceptual metaphors. Metaphors elevated to the status of concepts. The more concepts one has at one's disposal, the more situations can acquire meaning and the greater the scope for mental manoeuvre.

Concept formation and conceptual metaphors do not provide a complete explanation of abstract thinking, but are an important mechanism and key elements within it [37, 34]. The metaphor *Categories Are Containers* illustrates the relationship between primary and more abstract concepts. Hundreds of

times a day we experience the concept *container*; it is a recurring property of many things we deal with daily, from an early age: bags, bottles of apple juice, shoes, houses, trouser pockets. You can put something inside something. But sometimes it fits and sometimes it doesn't. This is how we learn the spatial logic of cases within the category of containers. If I have my keys in my hand, and I have my hand in my trouser pocket, then my keys are in my trouser pocket. If one now elevates the container metaphorically to a category or concept, then abstract concepts can also contain one another, and this becomes set theory. If a blackbird is a bird and birds are animals, then a blackbird is also an animal [37, p. 52-53].

Cognitive Linguistics suggests that all of our concepts take the form of metaphors. Even the most abstract conceptualisations in the exact sciences acquire meaning through the engagement of very ordinary “sensory-motor parts of the brain”. The meaning of concepts is therefore grounded in our ‘sensory-motor and affective experience’ [37, 33, p. 52].

5 Neuropsychology

The work of Lakoff and Johnson and others, dating from the late 1970s, predates by several decades a major leap forward in empirical research in the neurosciences. The results of this research confirm the picture outlined by Cognitive Linguistics and further flesh it out [26]. To illustrate this, Johnson refers, amongst other things, to the work of the American psychologist D.M. Tucker, who stated: “there are no brain parts for disembodied cognition” [37, p. 51]. Here, we draw in particular on the work of the British researchers Andy Clark and Anil Seth and their perspective on Predictive Processing Theory [16, 79].

Predictive Processing Theory posits that what we experience as perception is the brain's best *prediction* of the situation we find ourselves in. This theory assumes that our thought processes constantly generate hypotheses about the situations in which the body and brain find themselves, and then test these against sensory data. This includes data about the external world, but also about the body's internal state and posture. So it concerns sensory data from familiar senses such as sight, hearing, smell, taste and touch, but also body temperature, heart rate and the orientation of body parts.

Integrating all this hypothesising yields a complete and coherent picture. At least, that is how we experience it. We think we are truly perceiving the world. But it is merely a projection. Following in the footsteps of others, Anil Seth refers to the perception as we consciously experience it as a *controlled hallucination*.

A long-accepted, and intuitively understandable, theory of our perception, thought and action posits that the eyes, and other familiar senses, form the basis of a process that generates and updates our mentally experienced, clear picture of the world. Only then can thinking take place and action be taken. Predictive Processing Theory reverses the order of the process of perception, thinking and action. Here, perception begins with prediction. Conflicts be-

tween the prediction and sensory information lead to an adjustment of the prediction, thinking about it, and/or a physical action to adapt the situation to the prediction, and/or to investigate the situation further. Cognitive Linguistics provides insight into the structure of our prediction process by emphasising the metaphorical aspect of our thinking.

Both theories can be seen as part of a naturalistic strand of cognitive research. The reason for briefly exploring this side path is that work on artificial neural networks such as LLMs is typically part of it. The precursors of Predictive Processing Theory have a long history, which we shall not discuss further here [67, 58, 11, 14, 68].

6 Thinking styles and conversation partners

The American Psychological Association defines thinking as follows [3]:

Thinking. Cognitive behavior in which ideas, images, mental representations, or other hypothetical elements of thought are experienced or manipulated. In this sense, thinking includes imagining, remembering, problem solving, daydreaming, free association, concept formation, and many other processes. Thinking may be said to have two defining characteristics: (a) It is covert – that is, it is not directly observable but must be inferred from actions or self-reports; and (b) it is symbolic – that is, it seems to involve operations on mental symbols or representations, the nature of which remains obscure and controversial.

This is a very general statement. Psychological research into the supposed rational behaviour of *Homo economicus*, with which the names of the American-Israeli psychologists Daniel Kahneman and Amos Tversky are closely associated, has characterised thinking styles in greater detail [39, 78]. They argue that thinking activity can be broadly divided into two styles. These styles are referred to as System 1 and System 2. System 1 “operates automatically and quickly, with little or no effort and no sense of voluntary control.” System 2 “allocates attention to the effortful mental activities that demand it, including complex computation. The operations of System 2 are often associated with the subjective experience of agency, choice and concentration.” System 2 takes much more time and effort than System 1. Kahneman et al. note that System 1 is by far the most commonly used thinking style in our daily actions. The division into System 1 and System 2 is useful for modelling purposes; in practice, people often use both styles interchangeably and they overlap.

Although the American Psychological Association (APA) does not refer to System 1 and System 2, we believe that the APA’s supplementary definitions provide a useful further elaboration of System 2.

Critical thinking. A form of directed, problem-focused thinking in which the individual tests ideas or possible solutions for errors or

drawbacks. It is essential to such activities as examining the validity of a hypothesis or interpreting the meaning of research results.

Creative thinking. The mental processes leading to a new invention, solution, or synthesis in any area. A creative solution may use pre-existing elements (e.g., objects, ideas) but creates a new relationship between them. Products of creative thinking include, for example, new machines, social ideas, scientific theories, and artistic works.

Furthermore, the APA also distinguishes between *divergent thinking*, *lateral thinking* and *convergent thinking* as variants of critical or creative thinking, in which the emphasis lies respectively on new thinking strategies, the systematic reconsideration of assumptions, and the systematic verification of the logic of thought processes, that is, reflecting on one’s own thinking. We take the liberty of interpreting all these modes of thinking, from *critical* to *convergent*, as characteristic of System 2 thinking. We shall refer to this as analytical thinking hereafter.

In the literature on LLMs, we came across the distinction between *reasoning behaviour* and *reasoning performance* [62]. This is worth noting. Our classification clearly focuses on behaviour and not on performance, *accuracy* or *efficiency*. Those aspects are relevant in data centres seeking to process queries quickly, not for a discussion of principles. Another aspect we wish to mention here is the evolutionary basis of human thought [32]. We believe that humans represent merely a form of cognition determined by the terrestrial situation [57, p. 12].

7 Metaphorical problem propagations

The preceding Sections 2 through to 6 form the basis for the main argument that begins below. As indicated in the introduction, the argument consists of four parts. The first part follows in this section and the other three in Section 8, 9 and 10.

In brief, the four parts deal with: 1. a model of the human thought space; 2. the verbalisation of fragments from that thought space into text, and the properties of that text; 3. the training of LLMs on text and the artificial thought space that this produces; 4. what a chatbot can and cannot do within its artificial cognitive framework, and what implications this might have.

In this argument, the central question is: what kind of conversation partner is a chatbot? Following on from the previous section on thinking styles, we can now refine that question to: is the chatbot more of a System 1 or a System 2 conversation partner? Let us regard conversations as problem-motivated, as stated earlier. The conversation partners want to solve a problem or develop a perspective on it.

7.1 Problem types and repertoires are metaphors

Consider the conceptual system from Cognitive Linguistics. Lakoff and Johnson state [48, p. 3]:

In most of the little things we do every day, we simply think and act more or less automatically along certain lines. Just what these lines are is by no means obvious. One way to find out is by looking at language. Since communication is based on the same conceptual system that we use in thinking and acting, language is an important source of evidence for what that system is like.

Here, we understand language as language-in-action, as concrete linguistic expressions about concrete matters. What is the purpose of that “thinking and acting”?

Mark Johnson elaborates on this by, among other things, referring to the philosophy of the American pragmatists William James and John Dewey. Johnson summarises this in five key points; points 3 and 4 are the cherries we would like to pick here [37, p. 69]:

- Embodied cognition is problem-centred, and it operates relative to the needs, interests, and values of organisms.
- Embodied cognition is not concerned with finding some allegedly perfect solution to a problem, but one that works well enough relative to the current situation.

The American philosopher Donald A. Schön uses a few key terms relevant to our argument when he writes about social policy and generative metaphors in a collection *Metaphor and Thought* [70, p. 138]:

When we examine the problem-setting stories told by the analysts and practitioners of social policy, it becomes apparent that the framing of problems often depends upon metaphors underlying the stories which generate problem setting and set the directions of problem solving.

We interpret *framing of problems* and *metaphors underlying the stories* in this quotation as problem types and problem positions, respectively. We therefore believe that the process described by Schön also applies outside the domain of social policy. There is a further indication of this correspondence with Cognitive Linguistics.

In the AD, problem types and solutions are understood as patterns of roles and interactions between parties. In our view, this conception aligns with Cognitive Linguistics, which does not seek the nature of an object’s concept in terms of objective aspects and properties. Having a concept of an object is “to be able to simulate the kinds of perceptual, motor, and affective interactions you typically have with that kind of object” [37, p. 24]. And these interactions yield primary and conceptual metaphors via experiential correlation (see Section 4).

So the similarity lies in interactive experiences. Think back for a moment to the metaphor *Knowing Is Seeing* (example from Section 4). Thinking is not seeing. But looking closely at something often goes hand in hand with better understanding. The AD discusses such relationships in terms of roles and role patterns.

7.2 The notion metaphorical problem propagation

The notion *Metaphorical problem propagation* is a model of the human space of thought and experience that consists of a combination of the problem concepts from Aggregation Dynamics, the conceptual system of Cognitive Linguistics, and the predictive dynamics from Predictive Processing Theory.

This human space of thought and experience is not new; the reader experiences it every moment. What is new, perhaps, is that the combination of these three theoretical perspectives gives us the opportunity to reflect on problem-solution-driven conversations with chatbots. We believe this combination is best explained, to begin with, using an example.

Problem position 1: ‘A driver is driving on the motorway. The cars in front suddenly brake hard. Problem: how to react. One solution is to brake.’ Problem position 2: ‘The driver is braking. The sub-problem is now: braking adequately. The solution: to brake in such a way that the driver does not collide with the vehicle in front, and the cars behind also have the opportunity to come to a halt.’ Problem position 3: ‘The driver has come to a standstill. The braking manoeuvre was successful. Problem: the driver’s journey and daily schedule are no longer on track.’

The above is a simple example of problem propagation. Let’s make the events a bit more personal.

Problem position 1: ‘I’m driving on the A9 highway at 8 a.m. when I suddenly see the brake lights of the car in front of me flashing on and off. I want to stay alive. One solution is to brake.’ Problem position 2: ‘I brake. I still want to stay alive, but the subproblem now is: braking in such a way that I don’t crash into the car in front of me, and the car behind me doesn’t crash into me. The solution is to continue braking appropriately.’ Problem position 3: ‘I’m at a standstill. I realize I’m still alive. But I’m definitely not going to make it to my 10 a.m. appointment, with the possible consequence that the others in the meeting will decide something I oppose!’

The above is a simple example of metaphorical problem propagation. In this example, we see problem-solution relationships, predictions, and metaphorical relationships. Before I had to brake, I was under the assumption: ‘Everyone drives carefully and keeps their distance; business as usual; I’ll be at my appointment in 15 minutes, right on time.’ The brake lights of the car in front of me, which flash on, disrupt that image. Several predictions arise. We’ll mention two: ‘Someone accidentally taps the brake pedal, if the lights go out immediately and the car keeps moving, I can ignore this red light signal’; ‘this is actual braking, but how hard are they braking?’ The last prediction calls for closer observation to better understand what is happening. These predictions are part

of our conceptual system, which we usually take for granted. In this situation, our conceptual system says *Pumping braking Is Immediate danger*, and that calls for immediate action: braking. This example illustrates how the various elements of metaphorical problem propagation interact, generate meaning, and prepare and activate actions. If we take all these emotions, predictions and metaphorical relationships into account – which we are only doing to a limited extent here – the result is a holistic experience: an example of metaphorical problem propagation.

Metaphorical problem propagation is undoubtedly a complex concept. We find it helpful to visualize the mental architecture of metaphorical problem propagation as follows. The conceptual layering, from primary concepts to conceptual metaphors, constitutes a ‘vertical’ dimension here. In a ‘horizontal’ dimension, the propagations unfold. It is not appropriate to speak of a time axis. That suggests a precision that does not exist, and it would also limit mental freedom, for example – for other, possibly contradictory – orderings of events, or for projections into the future or the past. Finally, Predictive Processing Theory adds a third dimension, a prediction dimension. In this dimension, various predictions operate in parallel, at different levels of abstraction, with different emphases on the role of the conceptual system, and using different words.

Reflecting on current situations, formulating and testing hypotheses, and reflecting on and remembering experiences can thus be understood as the processing of malleable and flexible metaphorical problem propagations. A metaphorical problem propagation is not objective; it is partly unconscious, and chance also plays a role. Some observations stand out, others are missed. Some are perceived as problems, others are not.

Johnson describes a “representation theory”, referring to the work of J. Fodor, as “the idea that human thought consists of a series of functional computational operations on language-like symbols “in the mind” that can be used to represent external states.” Although we speak of it in words and words are often a consequence, our notion of metaphorical problem propagation is not a symbolic representation in this sense. Metaphorical problem propagation is our interpretation of what Johnson, with reference to Lakoff and Gallese, describes as an “interactionist, multimodal, simulation theory of meaning and thought” [37, p. 145]. In this, we emphasise problem-solution relationships.

We will explore the intersections between the three theoretical perspectives in greater depth. As mentioned earlier, one can envision the conceptual system as a hierarchy with primary metaphors at the bottom and, moving upward, more abstract concepts and sometimes even complex and inconsistent combinations [37, p. 53]. The concepts derived from AD simply fit into this hierarchy, with, for example, practical problems corresponding to the primary concepts at the bottom and abstract problems higher up, all the way to the most abstract category, namely that of the ‘unspecified problem’; “Houston, we have a problem” [88].

The AD also posits a similar hierarchy, from concrete to abstract, for problem situations, for problems, and for solutions (see the last paragraph of Section

3.4). The relationships that one mentally brings to life when contemplating a problem position, are experienced through metaphors. We therefore imagine that it is possible to initiate and continue the thought process at different levels of abstraction. And also that it can then proceed in parallel at other levels. So thinking about ‘the neighbours John and Jack who are arguing over the hedge between their gardens’ proceeds, so to speak, also at a more abstract level as P_1 , P_2 , and h in relation r , and perhaps also at more concrete or even more abstract levels. Or, for example, it proceeds in parallel in another remembered propagation, concerning a different argument, via associations with certain aspects of the context or Jack’s character traits. The freedom to switch between these levels and to introduce new levels are part of our conceptual system. It is also easy to imagine that it is possible to think backward and forward in the metaphorical problem propagation. Forward and backward exist by virtue of memory and experience; there is no fixed timeline.

The problem concepts from the AD and the conceptual system not only align well, but their combination also expands the range of possible applications. The conceptual system of Cognitive Linguistics already provides a dimension for describing action elements. *Image schemas* are “recurring sensory-motor patterns” that “afford us possibilities for meaningful interaction with our surroundings, both physical and social” [37, p. 21 & 100]. An extension of these are *X-schemas* [4, 47], which can be understood as composite processes of sensory-motor patterns. Sensory-motor patterns remain a current topic [64].

But these links between perception and action involve relatively basic interactions and short chains. Aggregation Dynamics operate on a different scale: memories, histories, plans, scenarios, interests, and hypotheses about what opponents are thinking (Theory of Mind). On that scale, it involves long, interconnected sequences of events or projected actions. This vastly expands the action dimension of the conceptual system and makes it relevant in many more situations, which, we believe, is entirely conceivable through metaphors as tools of thought. The sensory-motor patterns are, so to speak, metaphorically stretched to encompass problem situations between, for example, superpowers. This larger scale is not entirely foreign to Cognitive Linguistics, as is clearly evident from the earlier quote from Schön concerning “social policy.” There is quite a lot involved there. But it does not seem to be the scale at which one prefers to analyse.

But how do these mental images come to life? In his 2017 book, Johnson focuses specifically on the predictive brain [37, p. 142], but Predictive Processing Theory is all about that [16, 79]. This theory describes how the brain constantly generates predictions about the situation we find ourselves in and how it will unfold. But these predictions can just as easily pertain to the past, for example, regarding the question, ‘How could this have gone wrong?’ Predictive Processing Theory also explains how, from these competing and often mutually inconsistent predictions, a comprehensive and emotionally coherent picture, the Controlled Hallucination, is nevertheless generated.

We have the impression that these predictions are not fleshed out in very concrete terms by Predictive Processing Theory. This seems understandable

given the highly technical, empirical setting in which this theory is developed, where the focus is on correlations between sensory stimulation, neural effects, perceptions, and actions. But we take the liberty here of stretching the concept of prediction and postulating the metaphorical problem propagations as its structure.

Predictive Processing Theory deepens the notions of problem and solution. It is obvious that we identify problems with the mismatch between the predicted and the perceived world. A problem is then a *prediction error*. We naturally identify a solution with adjustments to the prediction or an action. A solution or step towards a solution brings the world into line with the prediction, or brings the prediction and the world closer together.

We believe that the concept of metaphorical problem propagation could be an interesting addition to Cognitive Linguistics and Predictive Processing Theory. For Aggregation Dynamics, as an attempt at a total system theory, metaphorical problem propagation represents a significant expansion of its potential applications. Furthermore, the connection with Cognitive Linguistics and Predictive Processing Theory strengthens the theoretical foundation.

Just like the problem propagation from Section 3.2, metaphorical problem propagations form a category of mental structures. A metaphorical problem propagation can be on a very large scale, encompassing everything a person knows and has experienced, but it can also concern a fleeting thought fragment that is quickly forgotten, or a fragment centred on a specific problem. In our argument, the context makes it clear what scope we mean, whether it concerns a total mental space, a cluster of related issues, a fragment, or a single thought.

8 Metaphorical problem propagations and texts

In Sections 3.1 and 3.2, we anticipated the conversion of the mental world of problem positions and problem propagations into textual or graphical form. Below, we propose a process that describes this transition and ask ourselves what the properties of the resulting text are.

Now suppose someone has a thought and wishes to share it with another person, or tell a story to another. What happens when that thought or story is told? Based on what we postulated above, we imagine that thought or story as a memorised or spontaneously unfolding metaphorical problem propagation. When we convert the metaphorical problem propagation from Figure 2, which represents memorized events, into text; this could, for example, yield something as illustrated in Figure 3.

But this is rather clunky language still. One can imagine that, with a little more editing, a smooth-flowing text could be produced from the list of sequentially arranged statements in Figure 2. We envisage the entire process of verbalisation as comprising three transformations. Firstly, the problem positions and the solution steps must be elaborated into text, for instance ‘Los Angeles City 1943; polluted air and fog; stinging eyes’. The problem positions break down into sequences that already suggest sentences or paragraphs. Secondly,

The situation and problem was
'Los Angeles City 1943; polluted air and fog; stinging eyes'
The first steps towards a solution were
'Louis McCabe's research in 1947; research by Arie Haagen-Smit in 1948'
This identified the causes
'meteorology; peroxides; unburned hydrocarbons; photochemical smog'
The solutions for these problem were sought in two ways
'increasing engine temperature; modified engine'
and
'catalytic converter; modified car body'
But it turned out that
'catalytic converters cannot withstand lead; ...'
which called for
'unleaded petrol; ...'
This, however, brought back an old problem
'knocking; ...'
The
'lead additive'
prevented that, now it was back
'knocking again; ...'
A solution was found in
'improved engine management; ...'
this solved knocking in a different way

Figure 3: A linearisation of the problem propagation shown in Figure 2, in which a few words have been added in **bold** to link the statements together.

the logic of the network structure must be represented in supplementary text with introductory and concluding sentences, references, headings, layout, summaries and examples. Here these are texts as ‘**The situation and problem was**’, and ‘**The first steps towards a solution were.**’ Thirdly, an editorial process turns this into a smoothly flowing text. That final step is not shown here. We call the result of this three-step process a *verbalised metaphorical problem propagation*. In the same way, we can of course also speak of a *verbalised metaphorical problem position*.

Verbalisation is the result of choices: the text ‘1. increased engine temperature 2. modified engine’ could also have been placed after the catalytic converter complications. The ‘mental space’ becomes smaller, but stability increases. Using a physics metaphor, one could say that the metaphorical problem propagation is in a state of superposition; in the text, a choice has been made.

That effect is stronger in written text than in spoken texts, where interaction is still possible. Then someone can still ask a question and the speaker can respond with a nuance and then continue the story. Since, our “body cannot not communicate” [19]. But our argument concerns written text, as that is the bulk raw material for training LLMs.

8.1 Structure and contents of text: static text

If “all life is problem-solving”, then problem types are important concepts in our thinking [76, p. 99]. Anyone can observe that the word ‘problem’ crops up frequently in any conversation. We assume this is also true of texts. That is already one characteristic.

Communication requires effort and must yield a benefit for the sender. This is the view put forward by evolutionary biologist Arik Kerschenbaum in *Why Animals Talk. The New Science of Animal Communication* [41]. The message must therefore be meaningful for the sender’s survival. This in turn suggests that the message must contribute to solving the sender’s current problems. This is most successful when the receiver can understand the message. It is therefore unlikely that anyone, when asked about her problem position, would really bring up absolutely everything under the sun or structure her message as a multidimensional “stream of consciousness” as James Joyce did in *Finnegans Wake* [38, 54]. That won’t work for a manual, a report or a leaflet. If the reader loses interest, the text misses its mark. The text must be to the point.

As mentioned, the person producing the text wants to achieve something with it. It is therefore to be expected that there is a certain degree of direction in it and that it is engaging for the user. This is consistent with the distinction made in Section 3.1 between internal and external problem positions. The external problem position coincides with a sender’s text strategy.

We can further clarify these ‘text strategies’ based on the work of Kahneman et al. They argue that System 1 dominates our daily actions. The direct, fast and uncomplicated System 1 suffices in most situations. This seems entirely in line with “we simply think and act more or less automatically along certain lines” from an earlier quotation of Lakoff and Johnson (Section 7.1). We suspect

that a significant proportion of the text that people produce also has a System 1 character.

In summary, we believe we can hypothesise that most texts are written from the perspective of a strategically oriented sender, deal with relevant issues that this sender wishes to address, are coherent and consistent, point towards solutions, are to the point, are non-experimental, are not difficult to understand, and are tailored to an audience seeking reassurance. Furthermore, these texts, like all written texts, are fixed, static and closed. They cannot be compared to the fluid and associative process of human thought or to people discussing a particular topic. The latter is also a form of text. Yet it is precisely this ‘limited’ type of text that is used for training LLMs.

We argue that examples of these texts can be found in a wide range of categories: newspaper articles; brochures; manuals; website posts; tweets; instruction guides; reports; Wikipedia articles; historical narratives; novels; (popular science) articles. In short, very everyday categories. Of course, within these categories there will be examples that do not quite meet our criteria, such as in-depth background articles or debate reports that record a great deal of interaction. But these remain static nonetheless.

The set of characteristics outlined above is difficult to summarise in a single word. Somewhat useful terms include: professional text, functional text, purposeful text, goal-oriented text, or solution-oriented text. But these are not quite right either and may already have other meanings. As we believe that the lack of flexibility is the most important factor, we refer to the whole as *static* text. This is a nice short term. But keep in mind this notion covers a range of text characteristics.

If we combine the findings of the previous section with those of this one, we believe we can conclude that the text people produce does not represent the intellectual richness from which it stems. Firstly, the verbalisation of any metaphorical problem-propagation already results in a fixation and simplification. Secondly, people do not simply produce text for the sake of it. In most cases, it is simple, purposeful text. This text lists a few points, offers some considerations, draws a conclusion or notes a line of thought or an action, and then moves on to the next point that needs to be made.

We have the impression that researchers analysing text datasets for LLM training focus primarily on the content-related characteristics of texts, such as unacceptable statements [8, 45, 65]. The thought processes and storylines in text datasets receive less attention. The authors of *Narrative Theory-Driven LLM Methods for Automatic Story Generation and Understanding: A Survey* state [55]: “Overall, [narrative] researchers rarely conduct unsupervised learning and pre-train LLMs on large volumes of texts themselves” and “(...) it cannot be ignored that narrative texts make up large portions of LLM pre-training.” This assertion supports our hypothesis, formulated above, regarding the structure and content of the texts in the text datasets, namely that they are of a specific type.

They then raise two highly relevant questions: “What effect does pre-training on more or less narrative texts have on an LLM?” and “(...) what kind of narrative artefact is an LLM?” These questions are also on our minds, and we offer some preliminary answers to them.

These questions are the subject of Section 9 and Section 10. We believe the second question is comparable to our own, namely, ‘what sort of conversation partner is a chatbot?’ We will argue that the static nature of the text dataset significantly determines the nature of the LLM as a ‘narrative artefact’. Assuming, therefore, that we may equate ‘narrative artefact’ with ‘conversation partner’.

The application of the various narrative theories discussed in the survey compels us to refine our arguments. For example, the quantification of narrativity, which “denotes the extent to which a text or discourse exhibits the qualities of narrative, including temporal progression, causal linkage, and the presence of agents and events.” We find this quantification interesting in its own right in relation to text datasets [74]. This metric could perhaps also be extended to include a measure of reflexivity or other characteristics we have identified.

9 The thinking space of chatbots

The raw data used to train LLMs consists of texts that are available online in one form or another [17, 56]: websites, brochures, wikis, novels, history books, e-mails, reports, manuals, accounts, and laws. This material is not immediately suitable for use. There are various reasons for making a selection – for example, ethical, content-related or technical grounds – and methods – for example, automated or partially manual – to make the texts suitable for training purposes [81, 91]. We now assume that most of the texts in the text dataset for training publicly accessible systems such as ChatGPT, Claude and Grok are of the, in the previous section, postulated static type.

These static texts share structural similarities because verbalisation is a process of strategic selection and fixation. However, we believe there are further similarities. There are three further reasons why texts within a dataset, however large, are likely to resemble one another: 1. Verbalisation creates connections between problems, problem positions and solutions. The Innovation Illusion posits that the majority of combinations of these are familiar patterns and that a limited set of jargon suffices to describe them; 2. Problems, problem positions and solutions correlate with conceptual metaphors from the conceptual system. This system is rich in patterns because conceptual metaphors build upon one another; 3. Even statements that are not directly related to problems, problem positions and solutions are based on metaphors, because we verbalise through comparisons. These, too, are patterns.

It is well known that neural networks are good at recognising patterns. It is well known that LLMs can store human concepts and features. And it is also known that LLMs can encode and utilise analogies. Suppose that this charac-

teristic is not merely incidental but a relatively systematic effect of the training process.

Our hypothesis on this is now that the LLM training process reconstructs the metaphorical problem propagations inherent in the texts of a text dataset. The resulting artificial metaphorical problem propagation, with varying degrees of coherence, together with word prediction, forms the basis for a chatbot’s ability to conduct a conversation that appears convincing to humans.

Is there any evidence to support this hypothesis? We see a connection between our hypothesis and the empirical results described in *Evidence of a predictive coding hierarchy in the human brain listening to speech* [15]. In this study, participants listen to stories. Whilst they hear a story, brain measurements track their predictions regarding the story’s continuation. The stories were also fed to LLMs. It appears that: i) the activation patterns in the LLMs “linearly map onto the brain responses to speech”; ii) the LLM operates on multiple timescales; iii) predictions in humans “are organised hierarchically: frontoparietal cortices predict higher-level, longer-range and more contextual representations than temporal cortices.” These three findings appear to align with the parallel we propose between metaphorical problem propagation and its artificial reconstruction and expression in chatbots; metaphorical problem propagation involves projection across time frames of varying degrees of detail; and, the parallel conceptualisation across multiple levels of abstraction.

We find another indication in *Implicit Representations of Meaning in Neural Language Models* [53]. The authors “identify contextual word representations that function as models of entities and situations as they evolve throughout a discourse.” In the authors’ words, the key difference from other studies that empirically attempt to uncover what an LLM encodes, is that they “aim to recover a representation of the situation described by a discourse rather than representations of the sentences that make up the discourse.” Based on examples, they succeed in demonstrating that these situations carry meaning because the actions described by the discourses have consequences for the statistical prominence of subsequent situations. We see a clear connection between our problem positions and these *situations*.

Leaving aside the question of whether even larger datasets are a desirable development [8]. With the current text datasets and LLMs, which are enormous compared to the past, the training process is quite successful in reconstructing metaphorical problem propagations – assuming our hypothesis is correct – as evidenced by the public’s positive reception of chatbots. But the reconstruction is not perfect; for example, because a chatbot can make things up and often has no answer to simple questions. Are these minor glitches that just need to be ironed out, or are there fundamental limitations? The question we wish to answer in Section 10 is: what kind of metaphorical problem propagation does an LLM simulate, and what can the chatbot do with it?

9.1 Conversations

Our hypothesis regarding conversations is as follows. A prompt activates a spectrum of artificial problem positions within the LLM. These are loosely connected groups of loosely connected clusters of neurons that encode concepts.

The activation of an artificial problem position is more or less strong. The degree of activation is the statistical degree of recognition of the prompt as a conceptual pattern. The prompt also activates a spectrum of solution steps (likewise groups of clusters of concepts) and problem positions that result from them, and so on. These are all activated to a lesser degree.

The tokens generated by the LLM lie within the ‘contexts’ of the statistically most appropriate problem positions and the statistically most appropriate solution steps. The text generation process has immediate effect: the spectrum of activated groups shifts as a result of the text generation, their composition changes, and the statistical relevance of the groups changes. Some problem positions become more prominent and others less relevant. When text generation stops, the neural network has reached a new state. There are ‘new’ problem positions that now take centre stage in processing the user’s next prompt.

We suspect that the ‘software shell’ surrounding the LLM does not simply follow the LLM blindly, but that it attempts to determine which text generation options yield the ‘best’ conversation. The shell will make choices based on certain parameters, but exactly how is irrelevant here. It may be that, according to the shell, there is no significant activation of problem positions. In that case, the chatbot responds with some seemingly appropriate phrase, such as a request for further explanation.

Depending on the outcome of the text generation process, with or without adjustment or selection by the shell, the chatbot responds with a counter-question, a suggestion, an answer or a solution to the problem posed. The conversation unfolds like a path through the chatbot’s simulated metaphorical problem propagation and, simultaneously yet fundamentally differently, like a path through the user’s metaphorical problem propagation.

In our view, the distinction between a conversation and a prompt is not clear-cut. On both the chatbot’s side and the user’s side, elements from the conversation continue to have an impact, or rather, influence the effect of the most recent prompt. This is essential, because otherwise a user’s response to a chatbot’s question would have no effect: the range of problem positions activated must change. The interaction between conversation and prompt can also be explicitly controlled using *in-context learning* as described in the literature; this should elevate the chatbot to a higher cognitive level [75].

It is tempting to imagine the successive artificial problem positions through solution steps as steps through the neural network, from one group of neurons to a completely different group of neurons. We do not believe that is what usually happens. If a prompt activates a spectrum of artificial problem positions, these will overlap, and they will also overlap with solution steps and the associated problem positions. This is not surprising because choosing a solution

step generally only partially changes the problem position. This is certainly the case with complex problems. If someone formulates a prompt ABC , and the chatbot suggests D as a solution step, then $ABCD$ is the new problem position. The training process is a form of data compression, and this means that these problem positions become encoded in an overlapping manner.

10 What kind of conversational partner is the chatbot?

Suppose an LLM encodes metaphorical problem propagation, as argued above. What kind of cognitive aid is a chatbot, and what kind of cognitive tasks is a chatbot capable of performing?

10.1 What kind of thing is a chatbot?

First of all, we need to free ourselves from the thought that chatbots are some kind of magical entity that defies human understanding.

Cognitive Linguistics posits that meaning arises from a sentient and experiential body that seeks to navigate a physical environment and a socio-cultural situation. We agree with this. A chatbot does, in a sense, have a ‘body’, but it is not human. The data centre is not a human socio-cultural environment. The chatbot therefore lacks the foundation of human primary subconscious concepts. This is a well-known argument [61]. So whatever form of aggregation a chatbot may be, at best it can simulate human interactions that it finds in text. The words it produces do not have the meaning they have for us.

That does not, however, mean that an intelligent system requires a biological body or that it cannot possess its own form of understanding [59]. But that is a different discussion [32]. Our argument concerns the basic chatbots to be analysed and, by extension, chatbots in general. These therefore exist within a limited human world of words, and how that relates to the world we experience lies beyond the capacity of the chatbots.

Is it still conceivable, then, that a chatbot could fully simulate our cognitive world, even if the chatbot does not understand what we are talking about? We find this idea reflected in a hypothesis by blogger janus: “a model [as an LLM] whose objective is text prediction will simulate the causal processes underlying text creation if optimised sufficiently strongly” [9]. This statement suggests that if you continue to optimise the model – for example, by using larger datasets, more intricate and extensive neural networks, and a more intensive training regime – you will end up with a model capable of simulating the causal functioning of our body and brain. In our view, this is a vain hope. As concluded at the end of Section 8.1, dimensions and associations are lost in the transition from the mental world of human thought to text. That is a fundamental problem. A practical problem associated with this is that people do not express everything that goes on inside them in text. The available text we have to work

with for LLM training is predominantly of the static type, as we argued. This text certainly provides an incomplete picture of human cognitive capabilities.

The chatbot is therefore limited in two ways. It is a simulation because the chatbot lacks a human, experiencing body; the simulation is flawed because it is trained on fundamentally incomplete data. It remains an artefact which, in our view, is far removed from all claims regarding Artificial General Intelligence or even stronger claims [46]. In Section 10.4, we argue that it is not plausible that ‘more complete’ data would lead to significantly ‘better’ results.

10.2 Static-thinking, no reflection

Our hypothesis is that the training text for public chatbots has an static character (Section 8.1). This has an effect on the artificial metaphorical problem propagation distilled from it. It seems highly plausible that this, too, therefore has an static character. By this we mean that a chatbot simulates a form of thinking that is more akin to System 1 thinking than to analytical thinking, and is therefore not very critical or analytical [39]. This thinking moves swiftly from problems and circumstances to conclusions, or solution steps, or actions, without much reflection. Consequently, whilst the circumstances – an important part of problem situations – may well be extensive, the assertions are not treated critically; they are simply accepted. The chatbot generally does not stray from the problem front.

It is difficult to determine precisely to what extent this ‘thinking’ mirrors human thinking, as System 1 thinking is not precisely defined, and there is, of course, a ‘grey area’ where System 1 thinking transitions into more analytical thinking. However, given the success of chatbots, it appears to align well with users’ everyday needs. It seems plausible, in line with Kahneman et al., that if many people rely on System 1-like thinking in many activities, those everyday needs are also characteristic of conversations with chatbots. Let us assume that. Users are then looking for quick answers and accept them if they seem plausible. We suspect that the average user would quickly tire of a chatbot that nitpicks over every little detail or starts nitpicking about supposed assumptions that might not actually be true. No, the chatbot fits neatly into the dominant System 1 part of the thinking spectrum. The chatbot provides what many people are looking for in their daily conversations: confirmation of their world view. They avoid difficult questions if they believe they can achieve their goals with a quick fix.

Our conclusion comes as no surprise to those who critically study the behaviour of chatbots:

[62]: “It is likely that the apparent success of LLMs in reasoning tasks predominantly reflects their ability to memorise the extensive data they have been trained on.”

[87]: “In terms of the Dual Process Theory of human cognition, LLMs are analogous, at best, to System 1 cognitive processing, and

it remains unknown how anything approaching System 2 cognition could be incorporated into a machine.”

We, of course, agree with these statements. The difference between our argument and that of these authors is that our model explicitly structures ‘LLM reasoning’ and ‘cognitive processing’ respectively by asking about the motivational basis of knowledge and concepts. Our perspective focuses on the characteristics of the texts on which LLMs are trained and emphasises problem-solving relationships. These two aspects appear to receive little attention in the literature.

10.3 A thinking aid after all

Is the chatbot limited to problem scenarios present in the training data, or is it capable of more? Geoffrey Hinton’s example from Section 2.1 shows that chatbots can find apt analogies. According to Hinton, this was an analogy that was not present in the training data. We suspect that this is a fundamental characteristic. So, we suspect that a chatbot could serve as a thinking aid, but how does it go about this, and how far does this thinking aid reach?

Various studies show that people can come up with ideas when working with chatbots. People faced with a creative task perform better when assisted by a chatbot than when left to their own devices or when they only have web searches at their disposal [50].

However, follow-up research prompted by these results shows that when the creative performance of the group being studied is aggregated, creativity decreases. In short, the chatbot steers the group towards the familiar path, the statistically obvious path [89, 60, 63, 21]. In the words of the title of one of these articles: ‘Generative AI enhances individual creativity but reduces the collective diversity of novel content.’

We therefore note that Hinton puts the word ‘similarity’ into the chatbot’s mouth. The careful design and optimisation of prompts is a topical subject, see, for example, the Google page on prompt engineering [29]. Work is also underway to gain a better understanding of the ‘thought processes’ a chatbot follows, thereby enabling greater control over them; for example, *chain of thought prompting* [43, 20, 75]. However, ‘chain of thought prompting’ does not appear to be without its problems [40]. We refer to all these methods and tools for guiding the chatbot as ‘targeted prompting’. Targeted prompting allows the chatbot to search for analogies, metaphors or other relationships between concepts. These concepts or relationships may be new or useful to the user. In the literature, this is referred to as *dark knowledge* [86]. Such a metaphor suddenly frames a phenomenon that is normally described in certain terms using different terminology, and this can lead to new, useful insights or breakthroughs. Will the chatbot successfully get off the mark as a thinking partner through carefully targeted prompts?

You can ask a chatbot what it thinks of a particular assumption. However, you would expect a critical, analytical thinking partner to introduce new elements of its own, to recap – ‘what do we have so far’ or ‘where do we stand now’

– or to breathe new life into a stalled analysis. In the latter case, this might involve uncovering a hidden and incorrect assumption, adjusting a definition, proposing a new hypothesis, going back in the line of reasoning and starting afresh, or re-examining all the steps in a dubious sub-argument. These are all mental strategies which the American Psychological Association categorises as critical, creative, lateral, divergent and convergent thinking. We believe that, with only the conversation transcript to rely on, the chatbot lacks the mental tools to carry out these kinds of mental operations itself. Therefore, the chatbot needs to be guided in order to be of any use in the analytical part of the spectrum. The researchers Taylor Webb, Keith J. Holyoak and Hongjing Lu confirm this view when they state [34]: “But there is no reason to suppose that the same system, absent human-generated inputs, would spontaneously develop a disposition to think analogically, as apparently happened at some point in human evolution. Thus, to the extent that LLMs capture the analogical abilities of adult human reasoners, their capacity to do so is fundamentally parasitic on natural human intelligence.” The chatbot is therefore not a fully-fledged thinking partner.

That is not to say that a chatbot cannot play a role in a thought process, as already indicated above [50]. However, in our view, this does require that the results produced by the chatbot in its role as a respondent are reliable and demonstrate systematic execution. Unfortunately, this is also where it falls short.

In the introductory sections, we have already described chatbots as unreliable. It is time to explain what we mean by ‘unreliability’. We use this term to describe the following characteristics: 1. unreliable in the sense of misleading the user; 2. unreliable in the sense that the chatbot harbours biases; 3. unreliable in the sense of being inconsistent, i.e. first saying ‘A’ and later, without giving any reason, saying ‘not A’; 4. unreliable in the sense of not exhibiting the same behaviour in situations that are identical or comparable for the user; 5. unreliable in the sense of fabricating and hallucinating, making things up out of thin air.

In our view, the lack of a physical body largely explains aspects of unreliability 3–5, and contributes to 1–2. We believe that problems of type 4 can be observed in work on analogy puzzles and mathematical problems [52, 69, 27]. However, points 1–2 are an effect of the training data; people are not always equally reliable, and much text is more or less strategic, and in some cases even manipulative. The concepts employed by the chatbot lack a foundation; the chatbot cannot assess whether a sentence makes sense, or whether it goes too far in manipulating the user; the chatbot lacks *common sense*.

Aspects 3–5 may offer advantages within the analytical spectrum of thought, namely in thinking outside the box. However, the user must be very alert. We believe, however, that these characteristics pose a risk to users who operate in the System 1 part of the spectrum of thought. We will return to this in Section 10.5.

10.4 Enhancing reasoning skills?

Could reflective ability be strengthened with an ‘analytical training set’? What would constitute a typical analytical text? We envisage in-depth philosophical inquiry into a subject that is far from mainstream science, which systematically unpicks and weighs up all manner of arguments and which, ultimately, draws no definitive conclusions, only conditional ones. An example might be Jan Verplaetse’s overview and critique of the free will debate [83]. We suspect it will not be easy to compile this analytical training set, but that is beside the point.

Nevertheless, even if an analytical training set could be compiled, we do not believe that a chatbot could thereby elevate itself to the status of an analytical thinking partner. We suspect that if the simulated metaphorical problem propagation partially loses aspects of its static nature by replacing certainties in the statements with reflective feedback, chatbots might start to engage in more elaborate speculation. You might then be able to use them more effectively as thinking partners, but the need for vigilant supervision of the chatbot also increases.

Analytical text does not solve the problem of the chatbot lacking an experience-based conceptual system. Even between people, text often falls short when it comes to conveying meaning. But we do possess an experience-based conceptual system. This helps us to reconstruct what another person means. Sometimes this is only possible through repeated reading and rereading (*close reading*) in order to construct a mental structure that is consistent with what you are reading. Even then, you may still be mistaken; you do not really understand the text. Lakoff and Johnson state, as previously quoted, [48, p. 3]: “language is an important source of evidence for what [the conceptual system] is like.” Text is, in this respect, a widely available medium. But it is a problematic medium.

Our conclusion, therefore, is that a chatbot, as a conversational partner, is not an analytical thinking partner, nor can it become one with its current architecture and through text-only training. Larger and larger models will not make chatbots more analytical as thinking partners; moreover, there are many other objections to moving in that direction [8]. Yann LeCun, a leading AI researcher and Meta’s former Chief AI Scientist, states: “Animals and humans exhibit learning abilities and understandings of the world that are far beyond the capabilities of current AI and machine learning (ML) systems.” And he does not believe that current AI technology is capable of bridging this gap [49]. We agree with him on this point.

10.5 Concept-based interpretability

Metaphorical problem propagation links human thought, human text and the functioning of LLMs. Research in the field of interpretability shows that it is possible to identify concepts and properties. If our model is of any use, we suspect it applies to concept-based interpretability (Section 2.2). After all, our

model suggests a meta-structure over concepts. We should point out, however, that we are too far removed from this highly technical field of research to assess the value of our model.

In any case, if concept-based interpretability succeeds in unravelling the conceptual structures of LLMs, this might open up the possibility of controlling certain properties, such as the various aspects of reliability. That sounds appealing because the combination of unreliability and robustness – the chatbot always has a smooth-running response – has proven dangerous for users who use chatbots too frivolously or are too trusting [44]. Does concept-based interpretability offer hope that the risk of accidents can be mitigated?

As we have already argued above, the reasoning ability cannot be improved and robustness is an inherent characteristic. This leaves only the option of increasing reliability to reduce the risk of accidents. We can identify a number of approaches to achieve this: safety barriers, proofs of correctness and redesign.

By ‘safety barriers’, we mean software surrounding an LLM that monitors the conversation and that intervenes when things go wrong. It is comparable to an automatic fire extinguisher in a kitchen with a gas hob. Correctness proofs are mathematically precise methods for demonstrating the properties of software. Redesign involves configuring an LLM in such a way that it always behaves correctly.

We believe that all these approaches face a fundamental problem: it is rather difficult to specify precisely which behaviour is acceptable and which is not. After all, we only understand ourselves to a very limited extent. Prejudices, tendencies, and bias present a separate problem within this context [71, 8]. Suppose one were to succeed in specifying an ‘acceptable’ set of desired properties; in that case, we expect that demonstrating those properties in an implementation would be very difficult [31]. So, there are two major obstacles, then.

If, however, both obstacles are overcome convincingly enough – that is, if one succeeds reasonably well in specifying the desired behaviour and demonstrating that a system behaves accordingly – then, in accordance with the precautionary principle, we suspect there will be a strong tendency to maximise reliability [73]. If that choice is made, we suspect it will be at the expense of the playfulness and associative ability of LLMs. And, we believe, that the usability of chatbots based on these LLMs would then decline significantly. We disagree with a claim in the literature that *hallucination* and *knowledge conflict* are challenges that “mainly derive from improper learning data” [86].

It seems that, if chatbots are to be of any use as creative thinking aids, one must be prepared to settle for a loosely defined and domain-specific “comfort zone” [42]. And users must themselves take care to remain within that zone and not be carried away by the illusion that the chatbot is an all-knowing innovation. It offers some hope that researchers Eric Bigelow and Tomer Ullman, in *People Evaluate Agents Based on the Algorithms That Drive Their Behaviour*, found that, given the chance, people take into account how a system works before assessing its results [10].

11 Conclusion

Chatbots are computer programs that serve as conversation partners or sources of information for humans in written or spoken conversations. Examples include ChatGPT, Grok, Claude, Deepseek, Gemini, and Lumo. The purpose of this article is to offer a perspective on the nature of chatbots as genuine conversation partners regarding problems in relation to their solutions. What can chatbots do and what can't they do in that context, and how can this be explained?

Our argument focuses on basic chatbots in the hope of making statements about the core functionality of chatbots in a general sense. Basic chatbots are assumed to consist of a Large Language Model (LLM) with a simple interface.

To better understand the behaviour of these chatbots in problem-solving-driven conversations, we examine the relationship between: human thinking, human text production, the text dataset used for LLM training, and the functioning of LLMs. We introduce the concept of *metaphorical problem propagation* to reflect on this relationship.

A metaphorical problem propagation is a model of the human thinking space, or a part of it related to a specific issue, consisting of the combination of the problem concepts from Aggregation Dynamics, the conceptual system of Cognitive Linguistics, and the prediction dynamics from Predictive Processing Theory.

People produce texts. We wonder what the characteristics of those texts are. We argue that the texts used to train chatbots only partially represent human thinking. We argue that this is because these texts are purpose-driven, intended for specific users and closed. We call this *static* text. Furthermore, these static texts share similarities because people tend to repeat themselves when identifying and solving problems; they are less innovative than they think (the Innovation Illusion); and because of the central role played by metaphors – which are closely intertwined – in human conceptualisation, the articulation of thoughts and the composition of texts.

It is well known that neural networks can identify patterns. Our hypothesis is that, during the learning process, an LLM reconstructs the implicit, reduced and static metaphorical problem propagations contained within the texts it is presented with. A prompt makes multiple problem positions and solution steps statistically more or less plausible. Whether the LLM works out one of the solution steps (which may also be a request for more information) or whether the chatbot intervenes with a response is not particularly important. The response alters the landscape of activated problem positions and the subsequent development. In that state, the user's next prompt alters the landscape once more. A conversation can now be defined as a path through the problem positions in the LLM's artificial metaphorical problem propagations and a path through the user's metaphorical problem propagations.

A chatbot can think along familiar patterns, look up facts, and generate text based on known lines of reasoning. That is a far cry from analytical thinking. Nevertheless, a chatbot can be coaxed into taking creative 'steps in its thinking.' But to do so, the chatbot must be guided. If properly directed, it can find

analogies in the enormous database at its disposal, and in its unrestrained nature, it can thus establish relationships between systems, situations, or domains that are new and potentially meaningful. We argue that, for various reasons, chatbot technology cannot be developed into a reliable, independent, analytical thinking partner.

Our conclusions align with Yann LeCun’s scepticism and are at odds with the optimism of Big Tech. That does not change the fact that chatbots exist, that they are being used on a massive scale, by both individuals and organizations, and that it is therefore socially and politically important to understand them. A key question for the discussion in this regard is: for which problem is a chatbot a solution? Our article aims to help answer that question.

Acknowledgements

We would like to thank Atze Bosma and Jos van Wamel for their comments on earlier versions of this article.

References

- [1] Gerard Alberts and Bas van Vlijmen. *Computerpioniers. Het begin van het computertijdperk in Nederland*. AUP, Amsterdam, 2016. In Dutch.
- [2] Tomas Humberto Montiel Alcantara, David Krütli, Revathi Ravada, and Thomas Hanne. Multilingual Text Summarization for German Texts Using Transformer Models. *Information*, 14(6), 2023.
- [3] Apa.org. *Dictionary of the American Psychological Association*. [URL](#), consulted on March 14, 2026.
- [4] David R. Bailey. *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. PhD thesis, EECS Department, University of California, Berkeley, 1997.
- [5] Steven Bartlett. Godfather of AI: They Keep Silencing Me But I’m Trying to Warn Them! Interview with Geoffrey Hinton by Steven Bartlett, DOAC series, Youtube, published on June 16, 2025, [URL](#), consulted on September 24, 2025.
- [6] Lochan Basyal and Mihir Sanghvi. Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models. *ArXiv*, 2023. [DOI](#).
- [7] Michael Begon and Colin R. Townsend. *Ecology: From individuals to ecosystems*. Wiley Blackwell, 5th edition, 2021.

- [8] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Conference on Fairness, Accountability, and Transparency*, ACM FAccT '21, pages 610–623, 2021. [DOI](#).
- [9] Leonard Bereska and Stratis Gavves. Mechanistic Interpretability for AI Safety – A Review. *Transactions on Machine Learning Research*, August 2024. [URL](#).
- [10] Eric Bigelow and Tomer Ullman. People Evaluate Agents Based on the Algorithms That Drive Their Behavior. *Open Mind*, 9:pp. 1411–1430, August 2025. [DOI](#).
- [11] Margaret Boden. *Mind As Machine: A History of Cognitive Science*, volume 1 & 2. Clarendon, Oxford University Press, 2006.
- [12] Margaret Boden. *Artificial Intelligence*. Oxford University Press, 2018.
- [13] John Brockman, editor. *What to Think about Machines that Think*. Harper Perennial, 2015.
- [14] Alexander Cameron. What is 4E cognitive science? *Phenomenology and the Cognitive Sciences*, 2025. [DOI](#).
- [15] Charlotte Caucheteux, Alexandre Gramfort, and Jean-Remi King. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, 7(3):430–441, 2023. [DOI](#).
- [16] Andy Clark. *The Experience Machine. How Our Minds Predict and Shape Reality*. Penguin, 2024 (2023).
- [17] Commoncrawl.org. *Home page*. [URL](#), consulted on April 13, 2026.
- [18] S.H. Dewey. *Don't Breathe the Air: Air Pollution and U.S. Environmental Politics, 1945-1970*. Number 16 in Environmental History Series. Texas A&M University Press, 2000.
- [19] De.wikipedia.org. *Paul Watzlawick*. [URL](#), consulted on May 29, 2026.
- [20] Aniket Didolkar, Anirudh Goyal, Nan Rosemary Ke, Siyuan Guo, Michal Valko, Timothy Lillicrap, Danilo Rezende, Yoshua Bengio, Michael Mozer, and Sanjeev Arora. Metacognitive capabilities of LLMs: an exploration in mathematical problem solving. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. [URL](#).
- [21] Anil R. Doshi and Oliver P. Hauser. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290, 2024.

- [22] Maximilian Dreyer, Jim Berend, Tobias Labarta, Johanna Vielhaben, Thomas Wiegand, Sebastian Lapuschkin, and Wojciech Samek. Mechanistic understanding and validation of large AI models with SemanticLens. *Nature Machine Intelligence*, 7(September):1572–1585, 2025. DOI.
- [23] Giuliano da Empoli. *Het uur van de wolven*. Atlas Contact, 2025. Original title: *L’Heure des prédateurs*. In Dutch, translated from French by Hans E. van Riemsdijk.
- [24] European Union. *Rules for trustworthy artificial intelligence in the EU*. EUR-Lex, URL, consulted on February 9, 2026.
- [25] Gilles Fauconnier and Mark Turner. *The way we think. Conceptual blending and the mind’s hidden complexities*. Basic Books, 2002.
- [26] Jerome A. Feldman. *From Molecule to Metaphor. A Neural Theory of Language*. MIT Press, 2006.
- [27] Elliot Glazer, Ege Erdil, Tamay Besiroglu, Diego Chicharro, Evan Chen, Alex Gunning, Caroline Falkman Olsson, Jean-Stanislas Denain, Anson Ho, Emily de Oliveira Santos, Olli Järvinemi, Matthew Barnett, Robert Sandler, Matej Vrzala, Jaime Sevilla, Qiuyu Ren, Elizabeth Pratt, Lionel Levine, Grant Barkley, Natalie Stewart, Bogdan Grechuk, Tetiana Grechuk, Shreepranav Varma Enugandla, and Mark Wildon. FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI, 2025. arXiv, DOI.
- [28] Hugo Gonçalo Oliveira, Brad Spendlove, Pablo Gervás, and Dan Ventura, editors. *Proceedings of the Sixteenth International Conference on Computational Creativity*. Association for Computational Creativity, 2025. URL.
- [29] Google.com. *Prompt engineering: overview and guide*. URL, consulted on July 16, 2026.
- [30] Kazjon Grace, Maria Teresa Llano, Pedro Martins, and Maria Hedblom, editors. *Proceedings of the Fifteenth International Conference on Computational Creativity*. Association for Computational Creativity, 2024. URL.
- [31] J.F. Groote and M.R. Mousavi. *Modeling and analysis of communicating systems*. MIT Press, 2014.
- [32] C. Heyes. New thinking: the evolution of human cognition. *Philosophical Transactions of the Royal Society*, pages 2091–6, 2012. August 5, 367(1599), Biological Science, DOI.
- [33] Douglas Hofstadter and Emmanuel Sander. *Surfaces and Essences. Analogy as the Fuel and Fire of Thinking*. Basic Books, 2013.
- [34] Keith J. Holyoak, Nicholas Ichien, and Hongjing Lu. Analogy and the Generation of Ideas. *Creativity Research Journal*, 36(3):532–543, 2024. DOI.

- [35] Chip Jacobs and William J. Kelly. *Smogtown: The Lung-Burning History of Pollution in Los Angeles*. Overlook Press, Woodstock & New York, 2008.
- [36] Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. Deep learning for text style transfer: A survey. *Computational Linguistics*, 48(1):155–205, 04 2022. [DOI](#).
- [37] Mark Johnson. *Embodied mind, meaning and reason. How our bodies give rise to understanding*. The University of Chicago Press, 2017.
- [38] James Joyce. *Finnegans Wake*. Wordsworth Editions, 2012 (1939).
- [39] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus & Giroux Inc, 2012.
- [40] Subbarao Kambhampati, Karthik Valmeekam, Siddhant Bhambri, Vardhan Palod, Lucas Saldyt, Kaya Stechly, Soumya Rani Samineni, Durgesh Kalwar, and Upasana Biswas. Position: Stop anthropomorphizing intermediate tokens as reasoning/thinking traces!, 2026. arXiv, 2504.09762, cs.AI, [DOI](#).
- [41] Arik Kershenbaum. *Why Animals Talk. The New Science of Animal Communication*. Penguin, 2025.
- [42] Jon Kleinberg and Sendhil Mullainathan. We built them, but we don’t understand them. In John Brockman, editor, *What to Think about Machines that Think*. Harper Perennial, 2015.
- [43] Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc., 2022. pp. 22199–22213, [URL](#).
- [44] Marloes de Koning. ChatGPT leek vol begrip over Adams zelfmoordplan. *NRC*, (September 20-21):Achtergrond pp. 4–6, 2025. In Dutch.
- [45] P. Kumar. Large language models (LLMs): survey, technical frameworks, and future challenges. *Artificial Intelligence Review*, 57:p. 260 et seq., 2024. [DOI](#).
- [46] Ray Kurzweil. *The singularity is nearer. When we merge with AI*. The Bodley Head, Penguin, 2024.
- [47] George Lakoff. *The Neural Theory of Metaphor*, 2009 (2008). UC Berkeley Previously Published Works, [URL](#).
- [48] George Lakoff and Mark Johnson. *Metaphors we live by*. The University of Chicago Press, Chicago, IL, revised edition, 2003 (1980).

- [49] Yann LeCun. A path towards autonomous machine intelligence version 0.9.2, 2022-06-27. *Openreview.net*, 2022. [URL](#).
- [50] B.C. Lee and J. Chung. An empirical investigation of the impact of Chat-GPT on creativity. *Nature Human Behaviour*, 8:1906–1914, 2024. [DOI](#).
- [51] Taewhoo Lee, Minju Song, Chanwoong Yoon, Jungwoo Park, and Jaewoo Kang. The curious case of analogies: Investigating analogical reasoning in large language models, 2025. arXiv, [DOI](#).
- [52] Martha Lewis and Melanie Mitchell. Evaluating the robustness of analogical reasoning in large language models. *Transactions on Machine Learning Research*, 2025. OpenReview.net, [URL](#).
- [53] Belinda Z. Li, Maxwell Nye, and Jacob Andreas. Implicit representations of meaning in neural language models. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online, August 2021. Association for Computational Linguistics. pp. 1813–1827, [URL](#).
- [54] Litseller.com. *Finnegans Wake*. Consulted on April 13, 2026, [URL](#).
- [55] David Y. Liu, Aditya Joshi, and Paul Dawson. Narrative Theory-Driven LLM Methods for Automatic Story Generation and Understanding: A Survey, 2026. arXiv, [DOI](#).
- [56] Yang Liu, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin. Datasets for large language models: a comprehensive survey. *Artificial Intelligence Review*, 58, 2025. article number 403, [DOI](#).
- [57] Sebastiaan Mathôt. *Een wereld vol denkers. Op zoek naar het bewustzijn van mens, dier en plant (en misschien zelfs AI)*. Uitgeverij Balans, 2025. In Dutch.
- [58] Humberto R. Maturana and Francisco J. Varela. *The tree of knowledge: The biological roots of human understanding*. Shambhala Publications Inc., Boston (Mass.), revised edition, 1998.
- [59] Anne Sophie Meincke. Bio-Agency and the Possibility of Artificial Agents. In Alexander Christian, David Hommen, Nina Retzlaff, and Gerhard Schurz, editors, *Philosophy of Science: Between the Natural Sciences, the Social Sciences, and the Humanities*, pages 65–93. Springer International Publishing, 2018. European Studies in Philosophy of Science, volume 9, [DOI](#).
- [60] L. Meincke, G. Nave, and C. Terwiesch. Chatgpt decreases idea diversity in brainstorming. *Nature Human Behaviour*, 2025. June, 9(6):1107-1109, [DOI](#).

- [61] Melanie Mitchell. *Artificial Intelligence. A Guide for Thinking Humans*. Pelican Books & Penguin Random House, 2020 (2019).
- [62] Philipp Mondorf and Barbara Plank. Beyond Accuracy: Evaluating the Reasoning Behavior of Large Language Models – A Survey, 2024. arXiv, DOI.
- [63] Kibum Moon, Adam E. Green, and Kostadin Kushlev. Homogenizing effect of large language models (LLMs) on creative diversity: An empirical comparison of human and ChatGPT writing. *Computers in Human Behavior: Artificial Humans*, 6:100207, 2025. DOI.
- [64] Emiko J. Muraki and Penny M. Pexman. The role of emotion in acquisition of verb meaning. *Cognition and Emotion*, 39(7):1457–1464, 2025. DOI.
- [65] Tarek Naous, Michael J Ryan, Alan Ritter, and Wei Xu. Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Bangkok, Thailand, August 2024. Association for Computational Linguistics. pp. 16366–16393, DOI.
- [66] Arvind Narayanan and Sayash Kapoor. *AI Snake Oil: what artificial intelligence can do, what it can't, and how to tell the difference*. Princeton University Press, 2024.
- [67] Allen Newell and Herbert A. Simon. *Human Problem Solving*. Echo Point Books & Media, 2019 (1972).
- [68] Doroteya Nikolova. Predictive Processing Theory in Mind Studies: Cross Points with 4E Cognition and Cognitive Linguistics. *Open Journal of Philosophy*, 15(2):349–366, 2025. DOI.
- [69] Gustaw Opielka, Hannes Rosenbusch, and Claire E. Stevenson. Analogical Reasoning Inside Large Language Models: Concept Vectors and the Limits of Abstraction, 2025. arXiv, DOI.
- [70] Andrew Ortony, editor. *Metaphor and Thought*. Cambridge University Press, second edition, 1993 (1979).
- [71] Matteo Pasquinelli and Vladan Joler. The Noosphere manifested: AI as instrument of knowledge extractivism. *AI & SOCIETY*, 36:1263–1280, 2021. DOI.
- [72] P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, Amy N. Yates, Kristen Greene, David A. Broniatowski, and Mark A. Przybock. Four Principles of Explainable Artificial Intelligence. Technical Report Interagency or Internal Report 8312, National Institute of Standards and Technology, september 2021. URL.

- [73] Roel Pieterman. *De voorzorgcultuur. Streven naar veiligheid in een wereld vol risico en onzekerheid*. Boom juridische uitgevers, 2008. In Dutch.
- [74] Andrew Piper, Richard Jean So, and David Bamman. Narrative Theory for Computational Narrative Understanding. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. pp. 298–311, [URL](#).
- [75] Aske Plaat, Annie Wong, Suzan Verberne, Joost Broekens, Niki Van Stein, and Thomas Bäck. Multi-Step Reasoning with Large Language Models, a Survey. *ACM Comput. Surv.*, 58(6):article 160, pages 1–35, December 2025. [DOI](#).
- [76] Karl Popper. *All Life is Problem Solving*. Routledge, 2001 (1999).
- [77] Chengwei Qin, Wenhan Xia, Tan Wang, Fangkai Jiao, Yuchen Hu, Bosheng Ding, Ruirui Chen, and Shafiq Joty. Relevant or Random: Can LLMs Truly Perform Analogical Reasoning? In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Findings of the Association for Computational Linguistics: ACL 2025*, Vienna, Austria, July 2025. Association for Computational Linguistics. pp. 23993–24010, [URL](#).
- [78] Herman de Regt and Hans Dooremaelen. *Het snapgevoel: Hoe de illusie van begrip ons denken gijzelt*. Boom, Amsterdam, 2015. In Dutch.
- [79] Anil Seth. *Being you. A new science of consciousness*. Faber & Faber Ltd., 2022 (2021).
- [80] Tian Shi, Yaser Keneshloo, Naren Ramakrishnan, and Chandan K. Reddy. Neural Abstractive Text Summarization with Sequence-to-Sequence Models. *ACM/IMS Trans. Data Sci.*, 2(1), January 2021. [DOI](#).
- [81] Tensorflow.org. *Tensorflow C4 dataset*. [URL](#). Consulted on April 13, 2026.
- [82] Transmathematica. *About the journal*. Consulted on May 21, 2026, see: [URL](#).
- [83] Jan Verplaetse. *Zonder vrije wil: Een filosofisch essay over verantwoordelijkheid*. Uitgeverij Nieuwezijds, 2011.
- [84] S.F.M. van Vlijmen. Exploration of Problem and Solution Types in Los Angeles County Smog Control 1943-1976. *Transmathematica*, 2025. [DOI](#).
- [85] S.F.M. van Vlijmen. Problem Positions and Nineteen Instances in the Smog Control Case Los Angeles County 1943-1976. *Transmathematica*, 2025. [DOI](#).

- [86] Mengru Wang, Yunzhi Yao, Ziwen Xu, Shuofei Qiao, Shumin Deng, Peng Wang, Xiang Chen, Jia-Chen Gu, Yong Jiang, Pengjun Xie, Fei Huang, Huajun Chen, and Ningyu Zhang. Knowledge Mechanisms in Large Language Models: A Survey and Perspective. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, Florida, USA, November 2024. Association for Computational Linguistics. pp. 7097–7135, [DOI](#).
- [87] Craig S. Webster. Natural and artificial intelligence – the psychotechnical agenda of the 21st century. *Journal of Psychology and AI*, 1(1):2491445, 2025. [DOI](#).
- [88] Wikipedia.org. *Houston, we have a problem*. [URL](#). Consulted on May 21, 2026.
- [89] Rebecca Winthrop. What 370,000 College Essays Tell Us About A.I.’s Effects on Creativity. *The New York times*, (May 27), 2026. Opinion, [URL](#).
- [90] Stephen Wolfram. *What Is ChatGPT Doing ... and Why Does It Work?* Wolfram Media, Inc., Kindle edition, 2023.
- [91] Xiao Yu, Zexian Zhang, Feifei Niu, Xing Hu, Xin Xia, and John Grundy. What Makes a High-Quality Training Dataset for Large Language Models: A Practitioners’ Perspective. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, ASE ’24. IEEE/ACM, 2024. pp. 656–668, [DOI](#).