Luca Neuperti
Matriculation Number: 6666888
Major: Computer Science BA
Application Subject: Sociology
Semester: 9
luca.neuperti@stud.uni-frankfurt.de

Bachelor's Thesis

Assessing Communicative Accommodation in the Context of Large Language Models

A Semiotic Approach

Luca Neuperti

Submission Date: 30st of November

Text Technology Lab
Prof. Dr. Alexander Mehler
Prof. Dr. Alexander W. Schmidt-Catran

Acknowledgements

Humans are relational entities and as such I have been defined though a web of lovely individuals. So, I want to express my gratitude towards those who have made me who I am. Thus, in a literal sense, thank you to my parents and especially my mother who was able to make me take my mind off the thesis once in a while and focus on something else.

I want to especially thank my proofreaders, Christian, Grace, Henri, Manú, Max, Patrick and Tabea. While I love all my friends, — I am afraid of finishing this statement. Note that this paragraph doubles as a voucher for a joint dinner on me. Nevertheless, the support and care I got from *all* my friends has been truly humbling. So, thank you to Gianni, Hai Mi, Klara, Laura, Lea, another Lea, and Leonie! Another thanks is due for my friend and grandmother Barbara.

I do acknowledge that acknowledgments are rather atypical for a Bachelor's thesis but gladly break the mold for spotlighting such a group of wonderful individuals.

Abstract

Recently, significant strides have been made in the ability of transformer-based chatbots to hold natural conversations. However, despite a growing societal and scientific relevancy, there are few frameworks systematically deriving what it means for a chatbot conversation to be natural. The present work approaches this question through the phenomenon of communicative accommodation/interactive alignment. While there is existing research suggesting that humans adapt communicatively to technologies, the aim of this work is to explore the accommodation of AI-chatbots to an interlocutor. Its research interest is twofold: (1) Firstly, the structural ability of the transformer-architecture to support accommodative behavior is assessed using a frame constructed in accordance with existing accommodation-theories. This results in hypotheses to be tested empirically. (2) Secondly, since *effective* accommodation produces the same outcomes, regardless of technical implementation, a behavioral experiment is proposed. Existing quantifications of accommodation are reconciled, extended, and modified to apply them to nonhuman-interlocutors. Thus, a measurement scheme is suggested which evaluates textual data from text-only, double-blind interactions between chatbots and humans, chatbots and chatbots and humans and humans. Using the generated human-to-human convergence data as a reference, the degree of artificial accommodation can be evaluated. Accommodation as a central facet of artificial interactivity can thus be evaluated directly against its theoretical paradigm, i.e. human interaction. In case that subsequent examinations show that chatbots effectively do not accommodate, there may be a new form of algorithmic bias, emerging from the aggregate accommodation towards chatbots but not towards humans. Thus, existing, hegemonic semantics could be cemented through chatbot-learning. Meanwhile, the ability to effectively accommodate would render chatbots vastly more susceptible to misuse.

Contents

1 Introduction

1.1 The Need to Re-think Interaction in the Era of Large Language Models

With the introduction of ChatGPT by OpenAI (2022), heralded by some as the "iPhone moment of AI" (Faught 2023), public dissemination of chatbots grew abruptly. As the ChatGPT app achieves market dissemination in record speed (Duarte 2024), many people consider it their first conscious *interaction* with Artificial Intelligence (hereafter AI[1]). The commercial success and popularity of these systems is the product of rapid development in Large Language Models (hereafter LLMs), which was sparked — in part — by the transformer model architecture as introduced by Vaswani et al. (2017). This architecture presents a new approach to *translation* of text but "generalizes well to other tasks" (Vaswani et al. 2023, p. 1) of language processing[2].

Given the recent societal perfusion of and an investment into AI technologies, there is great economic and consumer-study interest in making AI more accommodating to human users (cf. Paul, Ueno, and Dennis 2023). From the user-side, the release of (often gendered) voice-interfaces for chatbots (cf. OpenAI 2023a) and increasing embodiment and multi-modality result in unprecedented similarity in form and aesthetics between human and nonhuman interlocutors. This bears the potential to further and extensive anthropomorphization. Thus, regarding day-to-day communication with chatbots, especially when used in a more embodied way, the need arises to examine which ontological status can be ascribed to these supposed interlocutors. The question of how LLM-enabled chatbots can communicate and adapt to humans is not only philosophical, however. It constitutes a nexus of societal and technical issues. Considerations regarding the human-likeness of AI influence popular discourse, policy decisions, paths in technical development and even broader phenomena such as cultural change, language shifts and more.

The ability to properly define whether LLMs in general or a given LLM are able to hold human-like conversations also provides a foundational benchmark and thus basis for more concrete evaluations. In a time of rapid progress in LLM-development, having a classification system to clearly describe such progress alongside qualities yet unattained, could prove a helpful guiding structure. The aim of this thesis is to explicitly and systematically examine the societally and scientifically highly relevant phenomenon of communicative attunement in LLM-chatbots. It hopes to inspire pioneering empirical approaches, expanding what currently is a niche field.

[1] Compare the term of artificial *interactivity*, here: AI_M, from the first letter of the cited author, Mehler (2010)

[2] For a decisively *sociological* explanatory approach of the current popular fixation towards human-like machines see Pfadenhauer (2015).

1.2 State of Research and Research Gap

1.2.1 The Entanglement of Communication and Technology

While the widespread societal dispersion of AI chatbots is a relatively recent phenomenon, there is established literature on human communication with robots and chatbots, observing general tendencies (cf. F. Liu 2023). There appears to be a rise of explicit and implicit AI anthropomorphism, i.e. "individuals' perception of [nonhuman] objects as humanlike" (Li and Suh 2022, p. 2253). Concerns have been raised that such behavior could cause societal harm and should thus be limited (Shanahan 2023; Dippold 2023, cf.).

However, maintaining awareness of the non-living, non-thinking and non-feeling ontology of LLMs might still not suffice to bar humans from adapting to nonhuman interlocutors more generally. That is, "[h]umans tend to react to this simulated behavior in similar ways as they react to human behavior" (Hildt 2021, p. 2) and engage with technology in fundamentally social ways. Social Response Theory (SRT) theoretically frames this phenomenon, showing a social response to computers that is not explainable by explicit anthropomorphization, but considered a more primal phenomenon (Nass and Moon 2000, pp. 93–94). Various studies came to similar conclusions in phenomena such as users using human self-initiated communicative repair strategies towards both embodied (cf. Gandolfi, Pickering, and Garrod 2023, pp. 4–5) and chatbot artificial interlocutors (Dippold 2023, p. 29).

While human adaptation in the direction of AI systems has been explored in the aforementioned research, there is very little inquiry into the reverse phenomenon. The question of how to assess whether an AI-system is conversationally coequal to humans and more specifically, the ability of LLM-chatbots to adapt communication *towards humans* (hereafter called *communicative attunement*) has been on the periphery of scientific interest in various fields. However, it is not in focus of specific examination and seldomly explicated using coherent terminology. Given the broad societal, ontological and political implications associated with research on the matter, this lack of focus is unforeseen. Existent efforts on research and development regarding the social aspects of AI/LLM are concentrated on other areas. These are — among others — factuality (cf. Guo et al. 2023) and speculative research on AI-consciousness or Theory of Mind (ToM) (cf. Kosinski 2023). In the seminal technical reports of prominent LLMs, prevention of "toxic" utterances and discrimination is highlighted (cf. Touvron, Martin, et al. 2023; Touvron, Lavril, et al. 2023; OpenAI 2023b; Thoppilan et al. 2022) but there is no discussion equating to AI communicative attunement.

Research in consumer studies reveals that human-modeled communicative abilities in chatbots evoke more positive responses and continued interaction[3]. Namely, "social presence", is a positive predictor of continuance intention towards an AI chatbot (Jin and Youn 2023, pp. 1879–1880). Furthermore, a chatbot's "empathy response" positively impacts conversation quality, predicting consumer trust (Chi and Hoang Vu 2023, p. 269). [4]. These studies focus largely on the *effects* of (supposed) communicative attunement and not on the abil-

[3] See Mariani, Hashemi, and Wirtz (2023, pp. 12–14) for a broad overview

[4] There might be a phenomenon structurally similar to the *uncanny valley* explaining why humans have greater adaption towards bad chatbots and very good ones but not ordinarily good ones (cf. Skjuve et al. (cf. 2019), Mariani, Hashemi, and Wirtz (cf. 2023))

ity of a chatbot to attune, lacking meaningful definitions of the concept. Moreover, there are similarly implicit normative stances present in research touching on the phenomenon. Works in consumer studies exhibit a tendency to conflate the *is* and *ought* of human-like communication. As continued engagement and a positive emotional response are axiomatically assumed good, anthropomorphizing chatbots is described as instrumental to this end (Paul, Ueno, and Dennis 2023, p. 1214). In some cases (cf. Chi and Hoang Vu 2023, p. 270; Jin and Youn 2023), studies explicitly call for considering "increasing anthropomorphism of chatbots and inducing the sense of being co-present" (Jin and Youn 2023, p. 1874) in further AI development. Across fields, instances of presenting more attunement as an end in itself (cf. Biancardi, Dermouche, and Pelachaud 2021) and using human-associated terms to describe robots (Hildt 2021, p. 2) are prevalent.

1.2.2 Nonhuman Communicative Resemblance of Humans

Given the lack of clear, descriptive terms in some works and a certain normative tainting in others, the need arises to theoretically narrow down, in a clear and descriptive way, the concept of nonhumans engaging "naturally" in conversation.

In media studies, the concept of *media richness* describes the ability of a communication-medium "to reproduce the information richness sent over to it" (Sheth et al. 2019, p. 6) (cf. Daft and Lengel 1986). While media richness has the ability to encapsulate the discussed phenomenon[5], using it presents an attempt to expand a nonhuman-modeled theory towards the human. However, since the standard of human face-to-face (hereafter HH_F[6]) communication ideal-typically underlies any discussion on the ability of chatbots to aspire to it, the reverse approach is examined.

The concept of *engagement* in the field of Human Computer Interaction (HCI) fits this directionality better as it incorporates more social elements. While HCI has encompassed work on AI throughout its existence (Dix 2017, p. 128), there are calls from within to propel the field towards "Human Centered AI (HCAI)" (Xu et al. 2023, p. 5) and to thus address issues specific to AI systems. As such, engagement in HCI could provide a theoretical frame encompassing communicative attunement. Use of the concept within HCI is broad and flexible, enrolling many different definitions. Yet, 65% of publications addressing engagement[7] do not define the concept (K. Doherty and G. Doherty 2019, pp. 3, 6). Among the prominent definition classes employed by the remaining studies, "conversational engagement" (K. Doherty and G. Doherty 2019, p. 23) most adequately fits the descriptive requirements of the communicative attunement. Nevertheless, the term remains vague and warrants questions of operationalization. More relevantly, *engagement* in HCI carries a normatively more positive ascription, given the adjacency of research to development. Another perspective approach lies in *interactivity*. While most existing conceptions of the term (cf. Braun-Thürmann 2002) have lacked selectivity, the concept bears potential for a fundamental delineation as will be demonstrated in section 2.1. It is used for the purposes of this work.

[5] Albeit with less potential for precisely classifying LLMs

[6] Analogously, abbreviations of the scheme AB_C are employed using $A, B \in \{\text{Human}, \text{Computer}\}$ and $C \in \{\text{Textual}, \text{Face-to-face}, \text{Computer-Mediated}\}$.

[7] The sample was processed in a systematic review on the matter with $n = 351$

1.3 Position of This Work and Research Questions

Research on the vague concept of human-likeness in chatbot communication has been shown in section 1.2 to largely (1) not explicate or not clearly delineate utilized terms, (2) conflate normative and descriptive stances on the matter and (3) not engage with the phenomenon in a direct, scientifically sound manner. This thesis aims to provide an account of LLM-attunability that explicitly provides clear definitions, and discusses questions of ethics, ontology, and policy on a systematically derived basis, providing a taxonomy aiding in conceptually classifying specific LLMs and LLMs as a whole. To these ends, the concept of interactivity, as alluded to in section 1.2.2, is applied.

Doing so, this work strives to balance the needs of wide-ranging attempts of embedding chatbots into semiotics, psycho-linguistics and social sciences while at the same time aiming to maintain coherence and avoiding conceptual overextension. Due to the multiple fields of science involved and the complexities within them, this thesis does not have the ambition nor ability to provide a *grand theory* of sorts. Instead, its essential aim is to highlight connections across fields and phenomena and enable further research.

In service of conducting feasible, focused and operationalizable research in the highly relevant area thus far referred to as communicative attunement, two research questions (RQs) are brought forth:

Research Question 1 *To what extent do transformer-based Large Language Models possess the ability to communicatively attune to an interlocutor in socially situated dyadic, plain-text communication?*

Research Question 2 *How could empirical research be designed to assess communicative attunement exhibited by transformer-based Large Language Models towards an interlocutor in socially situated dyadic, plain-text communication?*

Both RQs approach the same phenomenon of *communicative attunement*. The term here refers to the ability of an actant (focused on LLM-actants) to attune its utterances to an interlocutor in a specific manner. It constitutes an umbrella term without any particular theoretical allegiance. An exact account of communicative attunement will be outlined in section 2 and defined in section 3.4, incorporating existing theories.

The explicated context of communicative attunement is that of dyadic, plain-text communication, narrowing the research scope down to exclude soliloquy, group interaction and any form of phonetic, gestural or otherwise embodied communication, including virtually "embodied" agents (Biancardi, Dermouche, and Pelachaud 2021, p. 2). The notion of *social situatedness* emphasizes that no social interaction is devoid of context. Given the more solipsistic tendencies present in research on communicative attunement, this element is crucial to properly encapsulate the phenomenon in question.

The object of inquiry is transformer-based Large Language Models (LLMs). These are artificial neural networks (ANNs) based on the transformer architecture (cf. Vaswani et al. 2023) whose weights are adjusted using large amounts of textual data, usually sourced from the internet. They are typically subsequently fine-tuned, and have the purpose of solving complex tasks related to language using language coherently (Naveed et al. 2023, p. 1). Through

chatbot-interfaces, LLMs are often used in socially situated dyadic, plain-text communication with humans.

RQ$_1$ and **RQ$_2$** share the same phenomenon and object. They differ, however, in their direction. **RQ$_1$** ponders the *general ability* of these language models to attune, which should be derived *structurally and theoretically*. Complementing **RQ$_1$**, **RQ$_2$** considers the *behavior* exhibited by specific LLMs of attuning to a human interlocutor, potentially measurable *experimentally and empirically*.

Thus, it is the aim of his work to bridge the gap between theories of communicative attunement on one hand and LLM architecture and behavior on the other. In doing so, a framework of *interactivity* is developed systematically through semiotics and Actor Network Theory in section 2. This framework is then operationalized by generalizing social-psychological theories of attunement (Interactive Alignment Theory, Communication Accommodation Theory) to nonhumans in section 3. Using propositions derived from these theories, ”genotypical” and ”phenotypical” characteristics of interactive systems are proposed. Using these, the general ability of the LLM transformer architecture to support interactivity is assessed in section 4 and suggestions are undertaken to reliably evaluate the attunement behavior of any communicating AI system (but specifically LLM-chatbots) in section 5. Finally, important parts are recollected and implications for future research trajectories, ethics, and policy inferred before finally providing a broad research outlook into artificial interactivity in section 6.

2 From Artificial Interactivity to Accommodation

2.1 Artificial Interactivity Defined Through Semiotics

For addressing the **RQs**, the notion of (dyadic) "communicative attunement to an interlocutor" is central. Therefore, a semiotic[1] framework to derive attunement from will be constructed before reviewing theories on the phenomenon directly. Such an embedding aids a more systematic and deductive delineation.

As introduced in section 1.2.2, the concept of interactivity, more so than engagement or media richness, bears the potential to describe the *humanesque* features of nonhuman interlocutors. To further use this notion for deriving communicative attunement, a substantial definition of interactivity is needed.

To take on this role, "parasocial interaction" might appear a promising interactivity model. However, it lies wholly in the subjective experience of one of the interlocutors (i.e. a human interlocutor) and thus does not constitute interaction, despite its name (Sutter and Mehler 2010, p. 94). Because of the lack of reciprocity and the danger of anthropomorphization of computers when applying a parasocial schema to them, some argue against applying labels of interaction or interactivity to HC communication in general (Sutter and Mehler 2010, p. 94). This perspective, however, dismisses the chance of defining in *social* terms what effectively is already a deeply socially interwoven phenomenon. It would result in the absence of a socially-informed, tangible framework to formally explicate the intuition that conversations with some chatbots feel more "socially present" than with others (cf. Jin and Youn 2023, pp. 1878–1880).

There are many approaches to *interactivity* from Computer Linguistics and Sociology (cf. Pfadenhauer 2013, pp. 142-143 and Sutter and Mehler 2010, pp. 91-99). The amount of these which examine nonhuman interlocutors, however, is considerably smaller.

One particularly early approach is undertaken by Braun-Thürmann (2002), who regards "interactivity" as a distinct HC communication analogue to interaction, a concept in HH_F communication. He defines the term as: "Those sequential activities which take place between the technical artifact and the user" (Braun-Thürmann 2002, p. 118), elaborating that this interactivity is subject to different restrictions of structure and mode which are set about by the design of the artifact. This rather ambiguous delineation is not suitably selective for today's context of advanced chatbots as it allows for many communicative happenstances to be labeled as *interactivity* under the pretense of accounting for *design-restrictions*.

For a more selective and formalized framework re-defining interactivity semiotically, Mehler (2010) is consulted. The semiotic "stage" for his definition is set up in a paper discussing

[1] Throughout this thesis and in service of simplicity, the term *semiotics* will be used to refer to the general discipline of the science of signs. In accordance with the Encyclopedia Britannica (Duignan 2023), it will thus cover the *semeiotic* of Charles S. Peirce, the *semiology* of Ferdinand de Saussure and the structural semantics of Algirdas J. Greimas (see also Švantner 2021, p. 290).

an inquiry broader than the present **RQs**: That of the ability of artificial agents to enable interaction, or "artificial interactivity" (hereafter AI_M[2]); i.e. the ability of nonhuman artifacts to constitute thorough partners in interaction (Mehler 2010, p. 3). Mehler introduces a "strong" notion of artificial interactivity, superseding the sociological conception of Braun-Thürmann (2002) by employing Peircian (cf. Peirce 2011) semiotics in the regularized notation of situation theory (Barwise and Perry 1983). This is done for the purpose of deriving reliable, testable qualities an artifact needs to posses for the ability to provide artificial interactivity in the context of web-mediated communication (Mehler 2010, p. 3).

Referencing prior conceptions of interactivity, Mehler (2010, pp. 1–3) notes that face-to-face interaction might retain its underlying interactivity even if some of its formal constraints are subverted. Aiming at providing a rigid re-construction of the term, he begins from a working definition applying the recursive, triadic self-dependency of signs (cf. Peirce 2011) to the interactivity of a joint activity: "The joint acting [ger. Handeln] of two [actants] is interactive if it creates dispositions for interactions or modifies preexisting interaction-dispositions." (own translation Mehler 2010, p. 4).

He expands this notion into a definition, describing that a dyad of two actors A and O acting[3] jointly constitutes *interactivity*. Specifically, Mehler lays out requirements for the interrelation of the components of both actants' actions which he refers to as *action operation sequence [Handlungsoperationssequenz]* H. He proposes six criteria for interactivity (Mehler 2010, pp. 12–14):

Argumentative Pillar 2.1

AI_M1	*The situation E is able to be decomposed into constituent situations $E = S_1, \ldots, S_k$ having unique time-space attributes. These can be used to contextualize the emergence of the constituent actions performed by A (i.e. $x = a_1, \ldots, a_n$) and O (i.e. $y = b_1, \ldots, b_n$).*
AI_M2&3	*Actors A (**AI_M2**) and O (**AI_M3**) respectively shape their disposition $d_A(O, E)$ and $d_O(A, E)$ for further operations in similar situations with the same (or same type of) interlocutor. These cause them to operate in a way determined through inductive learning from H.*
AI_M4	*The effect of this inductive learning is that the relation of both actants' operations H is increasingly likely to lose complexity with increasing amount of follow-up interactions.*
AI_M5	*To interact with the other in such a continued manner, A and O need to have a memory-structure to represent their dispositions $d_A(O, E)$ and $d_O(A, E)$. This cognitive perspective means that the relation between the dispositions can be seen as a complementarity or similarity relation This alignment also holds true for the perception of the context both actors might differ in. These typing functions grow increasingly similar given successful continued interaction.*

[2]Referencing the author's name for a clear delineation from the more common term of artificial *intelligence* (AI)

[3]Here, the term *acting* is used to refer to human and nonhuman operations symmetrically instead of using *acting* and *operating* respectively as Mehler does. Analogously, *interactivity* is used to generalize the notion of *artificial interactivity* for cases in which nonhumans are not part of the dyad.

AI*$_M$*6 *The dispositions can align both inter- and intra-situationally as described by Pickering and Garrod (2004).*

It is not the aim of this thesis to assess the ability of LLMs to enable interactivity in general. Instead, it focuses only on the central phenomenon of communicative attunement which might constitute sufficiency but not necessity for artificial interactivity at large[4]. While setting the stage for systematic determination of interactive attunability is a terminal goal in its own right, the notion of artificial interactivity shows how it could also be instrumental in aiding the assessment of the broader phenomenon.

2.2 Generalization Using Actor Network Theory

Mehler introduced the conceptual requirements for a certain situation to constitute *Artificial Interactivity*. However, it might be of analytical benefit to not only semiotically define the relational class in which the requirements *are met* but also theoretically frame the situation regardless of its interactivity.

Dismissing HCI's engagement for the same reason of partiality as in section 1.2, another framework is needed for generalization. Since Science and Technology Studies (STS) could be seen as carrying less positive normativity towards technology and since their object of research is the interplay of humans and (technological) nonhumans, the field presents a suitable candidate for generalization.

Using theories and frameworks from STS to analyze human-nonhuman interaction is not unprecedented. While other approaches employ theories such as Barad's (1998) Agential Realism (cf. Holohan and Fiske 2021), a particularly useful framework for generalizing AI$_M$ in particular is that of Actor Network Theory (ANT, cf. Latour 1987; Callon 1984). Its role here is afforded through ANT's semiotic and linguistic grounding which is — crucially — expanded into socio-materiality.

Mehler (2010) and Latour (2005) both base their frameworks on a notion of semiotic structure. Whereas AI$_M$ is based on Peirce's (2011) triadic notion of the sign, Latour builds on the distinct semiotics of Greimas and Rastier (1968).[5]

As Braun-Thürmann (2002, p. 63) points out, the strength of ANT relating to interactability of technical artifacts lies in its symmetry principle (cf. Callon 1986) which does not attempt to bind (social) agency to human-centered notions such as consciousness. Therefore, Mehler's (2010, p.15) argument of potential interactivity being independent of consciousness is shared by ANT; agency[6] is afforded not exclusively to human actants.

[4]This assumption is derived through analysis of the possibility of attunement and criterion for interactivity to exist in various combinations, pair-wise considering each of the six criteria. the result is attunement implying criteria 1,2,3 and 5 while there is a bientailment between criteria 4 and 6 and attunement respectively. Thus, the statement of the presence of attunability implying interactivity but not necessarily vice versa, is derived. Due to the focus of this work, this derivation cannot be laid out in more detail.

[5]A theoretically sound delineation and reconciliation of Peircian and Greimassian semiotics in the context of Latour exceeds the scope of this thesis and perhaps even one solely dedicated to such a delineation. The endeavor could, however, be undertaken by considering approaches from Švantner (2021), Tamminen (2020), and van Wolde (1987).

[6]Using the language of nonhuman agency and the fundamental human-nonhuman symmetry principle of

While the ontological question of where to draw the border of what belongs to a specific network and what does not is notoriously difficult[7], let the situation of two actants A and O whose operations are interrelated through relation H be distilled down to a dyadic actor-network with social context embedded/inscribed in both actants. In such a network of merely two entities, the most fundamental descriptor or their interplay in terms of ANT is that of *translation*. This term has undergone what can only aptly be described as a *translation* of different meanings in the works of Latour (cf. Janicka 2023). Translation evolves from (1) a concept rooted in Greimassian semiotics (Janicka 2023, pp. 852, 851) to (2) a more general, material phenomenon "extending the overly-restrictive definitions of semiotics to things themselves" (Magne 2011, p. 132) . The concept is further refined and generalized when (3) embedded into the project of "An Inquiry Into Modes of Existence"[8]. However, since this thesis regards a very specific phenomenon, a projection of this generalized multidimensional concept onto the discussed plane of the dyadic actor-network can be made[9]. Specifically, a more semiotically grounded Latourian concept of translation is enrolled into this thesis:

> [F]irst, translation means drift, betrayal, ambiguity. It thus means that we are starting from inequivalence between interests or language games and that the aim of the translation is to render two propositions equivalent. Second, translation has a strategic meaning. It defines a stronghold established in such a way that, whatever people do and wherever they go, they have to pass through the contender's position and to help him further his own interests. Third, it has a linguistic sense, so that one version of the language game translates all the others, replacing them all with "whatever you wish, this is what you really mean." (Latour 1988, p. 253)

This definition highlights the fact that all successful translation is making two propositions more equivalent. This is to be sharply delineated from the notion of artificial interactivity in which not only is there a movement towards equivalence but it is undertaken reciprocally by *both* actants A and O (cf. **AI$_M$ 2&3**). It follows that there is translation without interactivity if only one of the actors[10] aligns to the other. The other, in this case, would take the role of an "immutable mobile" (cf. section 6.2). This asymmetry in degree of mutability of the human and the nonhuman in the dyadic actor-network lends itself to further analysis of its societal and political dynamics which is done in section 6.2. It should be noted that translation (and blackboxing) occurs on many conceptional levels in this context, e.g. in the chain of "society to data to training to machine models to machine utterances to human perception to human action to society" and elsewhere. However, the focus here lies on the human and chatbot as blackboxed dyad-constituents. These are viewed monadically until there is a problem

ANT, one might be enticed to succumb to anthropomorphization. Consequently, the distinctively nonhuman quality of humanely associated ANT terms such as "agency", "cooperation" and more needs to be stressed here.

[7] This has led to many substantial critiques on the operationalization of ANT.

[8] This is an undertaking whose discussion is beyond the scope of this thesis. See Latour (2011) and the project's online nexus, *An Inquiry into the Modes of Existence* (2013).

[9] This is not done without losing much of what Latour seeks to highlight in his later works, i.e. the material-semiotic character and mode-dependency of existence. However, as Latour (2018, p. 257) puts it: "A linguist should never circumscribe the isolated domain of "Language," unless it is to interrupt this movement of articulation for a moment, to make the analysis easier".

[10] Intuition would suggest the human actor A

or controversy that needs them to be opened (Latour 1999, pp. 183–184). It follows that in section 4, assessing *accommodatability* "genotypically", the black box of the LLM is opened and its content examined.

Re-framing dyadic communication as an actor-network allows for structural pondering of the actants' mutability, a semiotically consistent root category for interactivity and non-interactivity and hints at power dynamics at play. It allows for a fruitful re-embedding of the more regularized *artificial interactivity* into the socio-material — something between "agency" and "structure", between "global-social" and "local-social" (Figueiredo 2010, p. 43).

2.3 The Need to Incorporate Non-Aligning Attunement

After expanding the conception of *artificial interactivity* to the socio-material root category of the dyadic actor-network, an expansion into the opposite direction is needed: the creation of sub-categories. The complementarity/similarity (cf. Raible 1981) relations of the interrelation of subsequent *action operation sequences* H_i of turn i provides a high-level description of the mechanisms at play. However, only the inclusion of established theories in the respective scientific domains related to interactivity allows for a more thorough analysis as well as the inference of testable criteria for communicative attunement.

This semiotic framework[11] poses several constraints which can be operationalized through different theories of which Mehler suggests *Interactive Alignment Theory* Pickering and Garrod (2004), Pickering and Garrod (2013), Pickering and Garrod (2021), and Gandolfi, Pickering, and Garrod (2023) and *Structural Coupling* as presented in *Autopoiesis* (Maturana and Varela 1980).

IAT provides a highly specific and mechanistic account of the alignment of several linguistic levels (or semiotic channels) in continued engagement. It thus is well-equipped to operationalize the notion of alignment as defined through Mehler. However, such an account would not sufficiently consider the high-level social functions irreducible from HH_F communication, a phenomenon, which Mehler (and works on interactivity in general (Sutter and Mehler 2010, p. 90)) views as a conceptional model[12] for interactivity. Consequently, section 3 reviews behavioral theories for their ability to fully support *alignment* as framed by Mehler.

[11] With $\mathbf{AI_M\ 1}$ more defining the set of applicable situations E.

[12] original phrasing: "maßstabsbildend[...]" (Mehler 2010, p. 15)

3 Embedding Accommodation in Communication Theories

3.1 Review of Interactive Attunement Theories and Their Relation

There is an abundance of different theories which bear the potential to provide an apt systematic framework for communicative attunement as described in section 2.1 and constrained by **AI_M 2-6**. However, even given the propensity of overlapping phenomena described by the term, there is little reference in the literature across traditions. The field has long been fractured and concepts and terms appear incoherent, leading to miscommunication. This lack of attunement in the field of communicative attunement, is not a recent phenomenon (cf. Burgoon, Stern, and Dillman 1995). An non-comprehensive list of the terms denoting some variety of the concept would encompass: *convergence, accommodation, interactional synchrony, entrainment, mimicry, adaptation, repetition, alignment, resonance, parallelism* and *structural priming* [1].

Not without some simplification, various approaches on the matter can be seen as adhering more to automatic *priming* or deliberate *grounding* processes as base mechanism for communicative attunement (cf. Rasenberg, Özyürek, and Dingemanse 2020, pp. 4-5 and Oben 2015, p. 19). The idea of *grounding* (cf. CAT in section 3.3) describes the deliberate, conscious, coordinative effort undertaken by the interlocutors to establish what is called *common ground* (CG, Clark and Brennan 1991), i.e. a shared, aligned understanding of the conversation's context. This CG requires less energy to establish and maintain than a full mental model (via Theory of Mind) of the interlocutor (Pickering and Garrod 2004, p. 178). In contrast, *priming* approaches (cf. IAT in section 3.2) presuppose an automatic, unconscious process triggered by low-abstraction-level neural effects[2] which mechanistically and interactively engender what Pickering and Garrod (2004, p. 178) call "implicit common ground" (iCG): a version of CG that is cognitively less expensive than *full common ground* (Pickering and Garrod 2004, p. 178). Many approaches of communicative attunement differentiate between synchrony by linguistic level (phonetics, phonology, morphology, syntax, semantics and pragmatics) as well as within these levels. Accounts using priming effects tend to causally link alignment across levels whereas grounding-based theories do not (Rasenberg, Özyürek, and Dingemanse 2020, p. 7).

While there are various undertakings comparing different communicative attunement theories (cf. Rasenberg, Özyürek, and Dingemanse 2020; von Bergmann et al. 2015; Oehmen 2005; Pardo, Pellegrino, et al. 2022) as well as offering hybrid-approaches (cf. Lewandowski,

[1] See Oben (2015, p. 11), Kopp (2010, pp. 1–4) and von Bergmann et al. (2015, p. 2) for more detail and expansive references

[2] Alongside priming, these may invoke mirror neuron theory (cf. Rizzolatti et al. 1999) or exemplar theory (cf. Lewandowski, Schweitzer, and D. Duran 2014).

Schweitzer, and D. Duran 2014; Fusaroli, Rączaszek-Leonardi, and Tylén 2014; Pardo 2012; Babel 2012; Pardo, Urmanche, et al. 2017), this thesis focuses on two prominent approaches (IAT and CAT) because of their central role in existing research and their integrative potential regarding nonhuman attunement.

3.2 Interactive Alignment Theory: Overview and Shortcomings

3.2.1 Overview: Mechanistic Alignment via Priming

Interactive Alignment Theory (IAT)[3], also referred to as theory of *Linguistic Alignment (LA)*, *Interactive Alignment Model (IAM)* or, originally as *Interactive Alignment Account*, was introduced through the seminal publication of Pickering and Garrod (2004). It has been regularly amended since its inception (cf. Pickering and Garrod 2021; Gandolfi, Pickering, and Garrod 2023). The basis of this priming-based approach is a mechanistic understanding of most dialogue. Alignment of situational models as the presumed goal of all communication is achieved through automatic processes percolating across different linguistic levels. This inter-level alignment is argued to result from shared cognitive representations. While the "autonomous transmission account" assumes isolated comprehension and production systems in speech, in IAT, there is no "decision box" between comprehension and production. Rather, each linguistic level in one of the two systems is coupled to its counterpart in the other (Pickering and Garrod 2004, pp. 175–176). The neurological underpinning affording this phenomenon — and thus a prerequisite for full interactive alignment — is argued to be "parity" of representation between comprehension and production (Pickering and Garrod 2004, pp. 177–178).

IAT criticizes that traditional linguistic theory is based on monologue as a theoretical model for general human communication behavior (Pickering and Garrod 2004, p. 170). In contrast, it presents mechanistic *dialogue* as the "default mode" of language. Given the resource-intensity of higher-level cognition, Pickering and Garrod (2004, p. 187) postulate a "dialogic continuum", ranging from intimate one-to-one dialogue with maximal alignment to the theoretical minimum of it: An isolated speech with no audience-feedback. Communication forms between these poles include more active mental state-mapping in smaller groups and "serialized" monologue in larger groups (Pickering and Garrod 2004, p. 187).

In their seminal paper, Pickering and Garrod (2004, p. 172) present six definitory propositions for IAT:

IAT1: "Alignment of situation models (Zwaan and Radvansky 1998) forms the basis of successful dialogue;"

IAT2: "the way that alignment of situation models is achieved is by a primitive and resource-free priming mechanism;"

IAT3: "the same priming mechanism produces alignment at other levels of representation, such as the lexical and syntactic;"

[3]Crucially distinct from Interactive Adaptation Theory (also IAT) by Burgoon, Stern, and Dillman (1995).

IAT4: "interconnections between the levels mean that alignment at one level leads to alignment at other levels;"

IAT5: "another primitive mechanism allows interlocutors to repair misaligned representations interactively; and"

IAT6: "more sophisticated and potentially costly strategies that depend on modeling the interlocutor's mental state are only required when the primitive mechanisms fail to produce alignment."

While the term "alignment" in IAT— crucially — does not refer to behavior directly, it manifests into behavior such as linguistic repetition and associative priming through percolation (Pickering and Garrod 2021, p. 129).

The communicative reduction of ambiguity through context is of the proposed effects of Interactive Alignment. Pickering and Garrod (2004, pp. 184, 187) predict that context would have a stronger influence than word-frequency and thus enable ambiguity reduction. Yet, there is little focus or evidence on automatic-alignment on such a pragmatic level resolving ambiguities (cf. Roche, Dale, and Caucci 2012, pp. 12–13).

The temporary utterance-to-utterance relationship of alignment behavior described thus far is referred to as *focal alignment* (Pickering and Garrod 2021, pp. 126–127). Yet, the notion of alignment invoked by Mehler (2010, pp. 12–13) differs in scope from this. Criteria **AI_M 2-3** describe the actant acting in similar ways when in similar situations and **AI_M 4** predicts a downward trend in relational complexity with continued interaction(s). These notions are better represented in *global alignment* (Pickering and Garrod 2021, p. 127) which means for the interlocutors to "produce similar utterances under similar conditions and to interpret such utterances in similar ways". Further "interaction (in general) enhances global alignment" (Pickering and Garrod 2021, p. 128). While focal and global alignment influence each other bidirectionally, global alignment is constructed through focal alignments, e.g. lexically through complex routines with novel interpretations (Pickering and Garrod 2021, p. 128).

3.2.2 Empirical Data and Critique on Interactive Alignment Theory

In order to correctly assess part of the interactivity of LLMs, the concept of interactivity itself needs to resemble the model that is HH_F communication. If it would not, unwarranted anthropomorphization could result from intuitively applying more qualities of the — using metaphor-terminology — *source/vehicle* of HH_F communication to the *target/tenor* (cf. Mooij 1975), i.e. HC communication defined through interactivity. Despite this work not aiming to systematically review communicative attunement theories, the described "metaphoric mismatch" could increase anthropomorphization in a problematic way and thus engaging with the empirical soundness of any involved theories is paramount.

IAT itself is based on broad empirical research on behavioral attunement, which is cited in the respective works (cf. Pickering and Garrod 2004; Pickering and Garrod 2013; Pickering and Garrod 2021; Gandolfi, Pickering, and Garrod 2023).

Subsequent studies have also examined the empirical extent of various predictions made in IAT. Regarding phonetic convergence, the evidence is inconclusive because of the high inter-

person variability of the phenomenon (Pardo, Urmanche, et al. 2017). Ostrand and Chodroff (2021, p. 13) demonstrate that alignment can happen at some linguistic levels but not on others, casting doubt on the IAT idea of a fully automatic, percolative mechanism. Research on the neural representation of percolation, representational parity, indicate similarly that the extent of neural parity does not fully support IAT. Some (cf. Pardo 2012) only focusing on phonetic alignment) accounts argue not for neural parity through a mirror-neuron mechanism but for the self-regulating function of speech perception influencing the production of speech (Pardo 2012, pp. 760–761), thus arguing against a *fully automatic* alignment mechanism (Pardo 2012, pp. 763–764). Regarding syntactic alignment, there is highly contradictory evidence ranging from evidence to a complete rejection of the phenomenon, explaining it instead through the lexical boost effect (Oben 2015, pp. 22–26) There is— however — robust evidence for lexical alignment whose priming strength requires significant effort to overcome (Oben 2015, pp. 17–21).

More generally, IAT has largely been based on studies using *task-oriented*, formalized interactions. Recently, it has become increasingly challenged by studies which examine more naturally situated interactions, suggesting presence of interfering "top-down social processes" (N. D. Duran, Paxton, and Fusaroli 2019, p. 420). Yet, given the evidence against the full automaticity and the complete percolation proposed by IAT, the most recent work of this tradition by Gandolfi, Pickering, and Garrod (2023) — while introducing important contextual aspects — does still not account for a full social embedding of dialogue. Thus, a theory is required to adequately explain when non-alignment at a given level takes place, possibly through higher level "social or communicative benefit" (Ostrand and Chodroff 2021, p. 15), which is examined below (section 3.3).

3.3 Communication Accommodation Theory: Overview and Shortcomings

3.3.1 Overview: Accommodation for Social Goals

As argued in section 3.2.2, all facets of HH_F communication need to be theoretically framed. Since IAT appears unable to do so, a supplementary theory with a more complete frame is needed. This frame is provided by *Communication Accommodation Theory* (CAT), detailed by Giles, N. Coupland, and J. Coupland (1991), based on *Speech Accommodation Theory* (SAT) as first brought forth by Giles, Taylor, and Bourhis (1973).

Accommodation is defined as an attunement process which can take the form of convergence and divergence[4]. Convergence describes the adaptation of qualities of a person's speech towards their interlocutor, which is in many ways reminiscent of IAT's alignment. The reverse effect is divergence through which an individual actively stresses differences from their interlocutor by adjusting their speech characteristics away from those perceived in the other (Giles, N. Coupland, and J. Coupland 1991, pp. 5–11). Both convergence and

[4]This thesis employs the notion used by Giles, N. Coupland, and J. Coupland (1991) of accommodation as an umbrella term, encompassing convergence and divergence as well as all of their shadings. This distinction is needed because in later works, Giles and Gasiorek (2013, p. 6) refer to non convergent accommodation as nonaccommodation, which conflicts with notion used here.

divergence can be further classified into subjective and objective as well as linguistic and psychological, presenting a total of eight sub-types (cf. Giles, N. Coupland, and J. Coupland 1991, p. 36), not including more nuanced delineations considering power, modality, symmetry (Giles, N. Coupland, and J. Coupland 1991, p. 11) as well as over- and underaccommodation.

In contrast to IAT, there are *two* main functions to *accommodation*: the "cognitive function" of establishing mutual understanding and CG and the "affective function" of managing social distance (Giles 2016, pp. 41–43; Giles, Scherer, and Taylor 1979). In IAT, the motive of cognitive accommodation, intelligibility is largely assumed to be universal to all communication (Pickering and Garrod 2021, p. 68). For this purpose of interaction, especially when framed in a joint tasks, IAT provides a detailed explanatory framework grounded in empirical evidence [5] The affective motive lies outside the theoretical scope of IAT but comprises a prevalent phenomenon with empirically measurable effects (cf. Giles, Edwards, and Walther 2023). For the purpose of identity maintenance, both convergence and divergence are employed. Convergence is used in *cooperative accommodation*, through similarity attraction (cf. Byrne 1971) and divergence (in *non-cooperative accommodation*) helps to increase social distance from an outgroup-interlocutor (Giles 2016, pp. 42–43). A third form of communication used for the affective motive is that of maintenance, the "absence of accommodative adjustments by individuals, that is, maintaining their 'default' way of communicating without taking into account the characteristics of their fellow interactants" (Giles and Gasiorek 2013, pp. 6–7).

CAT stresses the intra-dialogical *continuous* malleability of dispositions and thus theoretically implements criterion **AI_M 6** of learning within *and* across encounters. Namely, the interlocutor's "initial orientation" (Giles 2016, p. 44) is translated into an continuously shifting "psychological accommodative stance" (Giles 2016, p. 45) in a process which structurally parallels translation.

CAT remains largely agnostic as to whether processes of accommodation happen consciously/deliberately or unconsciously/automatically (Giles 2016, pp. 30, 41). There are experimental approaches into understanding which processes are automatic and which are the result of conscious cognition. These examine the effect of increased cognitive load on performance of a given accommodation behavior (Giles 2016, p. 30; Giles, N. Coupland, and J. Coupland 1991, p. 24). The evidence of trials using this schema appears to point — considered very broadly and without reference to IAT — at convergence occurring more automatically while individuals are more aware of and intentional in their divergence. Yet, higher social functions and expectation are able to direct conscious attention towards any kind of accommodation (Giles, N. Coupland, and J. Coupland 1991, pp. 24–25).

Similarly to IAT, "CAT rests on a number key Principles" (Giles, Edwards, and Walther 2023, p. 11) (cf. Giles 2016, pp. 50–51)) which have been last revised by Giles, Edwards, and Walther (2023, p. 11) to include 11 statements, most of which are testable. Omitting more

[5] It is noted but not elaborated on that the cognitive motive is not congruent with Interactive Alignment as there are instances of divergent behavior decreasing complexity and increasing comprehension. For instance in a person over-pronouncing their own dialect to highlight the limits of the shared CG or in a person trying to slow down their interlocutor by diverging in speech rate (Giles 2016, p. 43). There are also cases in which the opposite of the cognitive motive holds true and divergence is used to deliberately make communication more problematic (Giles 2016, p. 43).

general principles on the fundamental nature of accommodation, social expectancies, group identities and more, there are principles useful for further discussion (Giles, Edwards, and Walther 2023, p. 11, emphasis in original):

CAT1: "The nature of communication accommodation during an interaction is a product of people's various motivations for, and abilities as well as willingness to, adjust to certain relationally-defined others as well as the topics that unfold and are managed"

CAT2: "When people wish to reduce social distance during a FtF and mediated interaction, they are more likely to engage in accommodative acts that they *believe* will facilitate this outcome"

CAT3: "When people wish to increase social distance during a FtF and mediated interaction, they are more likely to engage in [accommodative] behaviors that they *believe* will facilitate this outcome and that can, arguably, be attributed by them as successful failure"

CAT4: "Accommodation [...] can occur not only in response to pre-interactional social identities but may, subsequently, arise in discourse from which they are then created and become situationally salient"

3.3.2 Empirical Data and Critique on Communication Accommodation Theory

As with IAT, CAT is built on numerous studies which are cited in the respective works (cf. Giles, Taylor, and Bourhis 1973; Giles, N. Coupland, and J. Coupland 1991; Giles and Gasiorek 2013; Giles 2016; Giles, Edwards, and Walther 2023)[6]. Further, there are examinations regarding specific forms of accommodation. For instance, lexical accommodation seems to exhibit such a strong effect that it is even present in the non-episodic, asynchronous plain-text communication of social media, as Danescu-Niculescu-Mizil, Gamon, and Dumais (2011) demonstrate. More specific shortcomings of CAT are discussed below.

3.4 Implications of a Common Framework

As has been established, a theory of communicative attunement would need to (1) satisfy the constraints of AI_M (cf. **AI_M 2-6**) while (2) accurately describing the (interactivity-informing) interaction of humans. Concerning the latter, studies on both theories[7] have found "conflicting results, supporting either CAT or IAT" (Jiang and Kennison 2022, p. 219). Crucially, there is both evidence for automatic convergence not predicted by CAT (cf. Kwon 2021) and evidence of social effects on alignment not predicted by IAT (Babel 2012, p. 188).

While IAT and CAT approach communicative attunement from what could be seen as diametrically opposing perspectives, both theories acknowledge their respective scope and its limitations. Hunter (2020) emphasizes that IAT "does not require alignment to be entirely

[6] Giles (2016) helpfully reviews quantitative and qualitative studies which use CAT.

[7] For extensive references that surpass the scope of this work, see Jiang and Kennison (2022) and Babel (2012, p. 188) on phonetic alignment as well as Oehmen (2005, p. 224) examining speech pauses.

automatic" (Pickering and Garrod 2021, p. 129), leaving space for effects described by CAT. Meanwhile, CAT's agnostic stance towards automaticity and aforementioned studies on cognitive load favor a more automatic mechanism co-existing with a more deliberate one (Giles, N. Coupland, and J. Coupland 1991, pp. 24–25). Further, both IAT and CAT suggest convergence is the *unmarked* ("default") pattern of communication (Pardo, Pellegrino, et al. 2022, p. 2).

Thus, there are social contexts in which "the natural tendency to converge may be superseded by a strong motivation to diverge" (Pardo, Pellegrino, et al. 2022, p. 2). That is, the non-conscious, automatic process of alignment is actively counteracted through a conscious effort to maintain identity[8].

As has been shown, neither IAT nor CAT are able to fully theoretically encompass all aspects of accommodation in a detailed manner. Evidence supports the existence of both automatic and deliberative communicative attunement processes. Accordingly, there have been numerous calls for hybrid models and proposals for existing or newly developed models to be used to this end. However, for analytical structure, this work maintains the distinct competencies of IAT and CAT by augmenting and connecting them into a common framework accounting for all features of HH communication. The section of said framework that is relevant to this thesis is referred to as *communicative attunement* or *accommodation*[9] and described briefly hereafter.

At the start of communication, in the absence of a accommodation motive, effortless automatic alignment (IAT) describes the interaction. If the actant has the *cognitive* motive (CAT) for mutual understanding, convergent (CAT) or divergent (CAT) behavior can be employed depending on the circumstances. If the accommodation motive is *affective* (CAT), i.e. regulating social distance and thus identity, convergence (CAT), divergence (CAT) and maintenance (CAT) can also be used. For either motive, divergence and maintenance have to result in costly high-level cognition to override any automatic alignment mechanisms (IAT) present while convergence can use a mix of automatic alignment (IAT) and high-level cognition, depending on the needs of the situation and the mental energy required. Both convergence through deliberate strategy/automatic-alignment and divergence/maintenance trough deliberate strategy can manifest in focal (individual utterances) and global predispositions for future utterances. A focal manifestation of convergence of any kind might be behavioral repetition while divergence might manifest in a more deliberate use of different behavior (like differing phrases or anti-association). Maintenance manifests in no measurable attunement of behavior towards or away from the other.

It is argued here that with density and frequency of communication, the need for aligned mental models increases significantly. This means that with decreasing frequency of communication, there is less punishment in misaligned situational models and non-convergent strategies become more viable. Yet, with increasing frequency, the cognitive motive grows stronger and even in interactions with overall *focal* divergence[10] (e.g. familiar and opposing

[8] Among other possible high-level goals such as perception

[9] Here, the term is used a more general way than in CAT since accommodation is the root category of communicative attunementbehavior in the framework.

[10] Yet not maintenance as the absence of reciprocal influence on communicative behavior does not support any conceptual alignment.

debating partners), the situational models become increasingly aligned. Note that this does not mean every dyad will culminate in aligned situational models with time, since social motives for non-convergence can emerge from interaction and context at any time. In such a case, however, the frequency of encounters is likely to decrease due to the misalignment, supporting the proposed association of higher frequency of encounter with more aligned situational models.

For measurability, this means that in an experimental, few-encounter context, barring any communicative task, there is little incentive to adopt a certain communicative approach, i.e. focal convergence, divergence or maintenance or any combination of the three might be employed. Consequently, only convergent (be it automatic or deliberate) behaviour or divergence would prove attunability in the object of study while non-accommodation could result both from the inability to accommodate or the operational need to not accommodate. Thus, on an epistemological level, the (effective) 'accommodatability' of an LLM could only be proven but not disproven. However, since higher-level planning abilities are agreed to be beyond the scope of what (current) LLMs can support, the case of these abilities being the reason for a lack of behavioral adaption, can be dismissed, leaving the ability of an LLM to realize or simulate accommodation able to be proven or disproven based on the focal manifestations of convergence and divergence. Hence, if HC_T communication provides a similar measure of behavior matching as HH_T communication, there is evidence supporting convergence behavior. Divergent behavior might be considerably more difficult to assess and will be omitted in this account due to the lack of LLM high-level planing ability.

For this endeavor, the propositions of both IAT and CAT are collected and adjusted to be used as "genotypical" (i.e. structural) requirements (cf. section 4)) and "phenotypical" expression (cf. section 5) of LLM-accommodation. Generalizing more human-specific elements and omitting **IAT4** (percolation) given its empiric controversy, a common framework F for accommodation is derived:

Argumentative Pillar 3.1

F1: *The nature of accommodation during a communicative episode is a product of the actants' abilities and higher reasons to adjust to the relationally defined other as well as to the topics that unfold and are managed (cf. **CAT 1**)*

F2: *When an actant aims to reduce social distance (affective motive) or increase comprehension (cognitive motive) during a communicative episode, it is more likely to engage in accommodative acts that it assesses will facilitate this outcome, i.e. convergence (cf. **CAT 2**)*

- ***F2a:*** *Alignment of situational models enables cognitively successful dialogue (cf. **IAT 1**)*
- ***F2b:*** *A resource-free priming mechanism enables automatic alignment of situational models (cf. **IAT 2**)*
- ***F2c:*** *The same priming mechanism enables alignment at other levels of representation, such as the lexical and syntactic (cf. **IAT 3**)*
- ***F2d:*** *Another primitive mechanism allows interlocutors to repair misaligned representations interactively (cf. **IAT 5**)*
- ***F2e:*** *More resource-intensive and sophisticated grounding strategies can enable alignment of situational models or repair of alignment if automatic-alignment or automatic repair fails or higher social functions require it (cf. **IAT 6**)*

F3: *When an actant aims to increase social distance (affective motive) during a communicative episode, it is more likely to engage in accommodative acts that it assesses will facilitate this outcome, i.e. divergence or maintenance (cf. **CAT 3**)*

F3a: *Resource-intensive, sophisticated, non-automatic strategies enable divergence or maintenance*

F4: *Accommodation can occur not only in response to predispositions before the encounter but may, subsequently, arise in discourse from which they are then created and become situationally salient (cf. **CAT 4**)*

F5: *The global alignment of situational models or predispositions (and thus the decrease in complexity) grows with the frequency of communication employing any variation of local convergence and/or divergence*

Mehler (2010, p. 21) argues that only cognitive (i.e. structural) criteria can be used for assessing whether interactivity in a given system is *realized* rather than merely *simulated* (as differentiated by Pattee 1987). While the differentiation is epistemologically warranted, the aim of this work is to provide assessment-criteria for both the stricter *realized interactivity* and the practical *effective interactivity*. The latter being functionally identical to the former when both systems are blackboxed. Using these concepts and the dyadic actor network, a nuanced taxonomy is created.

In this taxonomy, a disjunctive decomposition of alignment as well as a composition of the dyadic communicative network aids in maintaining a concise discussion and operationalization. While only *realized intrinsic alignment* would constitute an ontologically precise category, the inclusion of simulated extrinsic and intrinsic (see discussion in section 4.4.2) alignment allows for a practical and helpful further classification where there would only have been "non-alignment". These further categories thus enable this discussion to also encapsulate more "common-sense" understandings of interactivity and the degree in which communication with an LLM-chatbot *feels* natural to the user[11]. The taxonomy's distinction between successful non-alignment (maintenance) and failed alignment (Giles and Gasiorek 2013, p. 19) might prompt questions of AI intentionality and possible benefit of such a behavior which are addressed in section 4.

[11]This experience may depend on many other factors not discussed in this thesis; see also Jin and Youn (2023).

4 Structural Conditions for Accommodation

While there are considerations on the neural basis of attunement behavior, most research to this date focused on phenotypical effects of accommodation. Thus, genotypical works on the matter (both on humans and nonhumans) can only be highly exploratory and speculative. As such, any subsequent application of said research to nonhuman cognition is only reasonably able to produce *hypotheses* regarding the 'accommodatability' of LLMs. These nevertheless hold value in their potential to shape subsequent experimental protocols to be more promising in their contribution to the field (cf. section 5).

4.1 The Viability of Structure-to-Structure Comparisons

To derive reasonable claims about behavior based on a cognitive-structural substrate in another entity, inferring similar behavior based on similar structure can — in general — be a viable path. This section briefly outlines the status of research on possible cognitive substrates for human alignment and the extent these can be mapped to the general cognitive structure of transformer models.

The travel of alignment effects across linguistic levels is a major element of IAT and explicitly predicted by it. The structural prerequisite of this percolation is a parity (or coupling) of representation for each linguistic level between comprehension and production systems, as laid out in section 3.2.1. In contrast to the "autonomous transmission account" of isolated comprehension and production systems, IAT suggests that there is no "decision box" between comprehension and production (Pickering and Garrod 2004, pp. 175–176). This makes parity a necessary, but not sufficient, condition for Interactive Alignment (Menenti, Pickering, and Garrod 2012, p. 4). Expanding on this parity, Pickering and Garrod (2013) propose that an interlocutor predicts the other's utterance while listening and then compares it to what is actually uttered. For this, the same mechanisms as for production is used in what they coin the "*simulation route* in action perception" (Pickering and Garrod 2013, p. 334, emphasis in original).

There have been many studies on different levels of alignment regarding the neural structures involved (Menenti, Pickering, and Garrod 2012). Pickering and Garrod (2004, p. 188) [1] suggest *mirror neuron* activity as a plausible neural substrate of such. Mirror neurons were first observed by Rizzolatti et al. (1999) in monkeys via direct measurement of neuron activation. These are neurons activated by perceiving an action such as grabbing in the observing monkey's motor system associated with the perceived action. While it is assumed by many that mirror neurons play a role in human cognition as well, they are only studied *indirectly* through regional activation correlations[2].

[1] See also Pickering and Garrod (2013, p. 336)

[2] Still, there are speculative hypotheses that the mirror neuron mechanism of *embodied simulation* has played

While there is evidence supporting the IAT hypothesis of shared mental representations between interlocutors when evaluated in a highly-structured dyadic online communication task of image description and recognition (W. Liu et al. 2019, p. 71), there is no neuron-activation data on real interactive dialogue (Menenti, Pickering, and Garrod 2012, p. 5). Given such inconclusive evidence of mirror neuron activity in human language processing (and alignment), a useful and more observable proxy-measure might be the phenomenon of percolation that has been linked with mirror neuron activity.

Despite studies both showing alignment at every linguistic level (Ostrand and Chodroff 2021, pp. 1–2) and demonstrating the potential for percolation on the basis of parity, there is empirical evidence against such automatic percolation. Both a corporal study measuring correlation between levels (cf. Weise and Levitan 2018) and a trial comparing phonetic and syntactic alignment variables in naturalistic image description (cf. Ostrand and Chodroff 2021) showed no clear relationship between alignment in a certain level and others. Importantly, Ostrand and Chodroff (2021, p. 16) showed that even within a given level (here: phonetics), there can be alignment in some measures while not in others as "communicative success does not require alignment within and across levels in tandem" (Ostrand and Chodroff 2021, p. 16). Consequently and despite frequent precedence in publications (Ostrand and Chodroff 2021, p. 2), inter-level as well as intra-level percolation can not be used for generalization of empirical findings. Given the current state of research, while mirror neuron activity is a valid potential explanation of alignment behavior (Oben 2015, p. 41), the evidence is inconclusive regarding its involvement at different linguistic levels.

In summary, there is evidence against the phenomenon of percolation of alignment and mixed evidence able to support parity of neural representation between comprehension and production systems. Mirror neurons are a plausible candidate to explain a part of alignment, but their existence and role in aligning representations is highly debated.

Neuroscience has — similarly to cognitive linguistics prior to Pickering and Garrod (2004) — focused on the isolated individual as base of analysis. It is only recently that there have been calls for "a second-person neuroscience" (Schilbach et al. 2013) with interpersonal studies on neurological alignment still severely lacking. It is this inconclusiveness of evidence that hinders a productive mapping of human alignment structures onto artificial cognitive structures. Further, there is a fundamental difference between biological neural networks (BNNs) and artificial neural networks (ANNs) by design. Bioplausibility has not been a major paradigm of modern AI development. Instead — paralleling neuroscience and cognitive linguistics —, there has been a paradigmatic "solipsistic perspective on intelligence" (Bolotta and Dumas 2022, p. 1), conceptualized as an agent interacting with objects. There are increasing calls for moving towards what Bolotta and Dumas (2022) call the "dark matter" of the field. These developments towards bioplausibility are however still in their infancy.

As follows, trying to compare cognitive structures of BNNs to ANNs would (1) not be based on a scientific consensus and (2) constitute metaphorical "overstretching" due to the non-bioplausibility of ANNs.

a significant role in the development of more abstract language and social behavior in humans (Gallese 2008)

4.2 Behavior-to-Structure Inference

The propositions of the common IAT CAT framework introduced in section 3.4 posit that certain characteristics are needed for an LLM[3] to cognitively *realize* accommodative behavior. For what is here called *minimally viable* attunability C_{min}, an LLM would need to fulfill all of:

$C_{min\ A}1$: Representational access: The ability to self-modify its representations of utterances (based on F2)

$C_{min\ A}2$: Deliberate repair The ability to— based on a higher goal — self-initiate a deliberate repair process to align misaligned representations (based on F2)

$C_{min\ A}3$: Global predisposition: The ability to access a *predisposition* shaping (and being shaped by) the current accommodation based on prior accommodations in similar contexts (based on F5).

$C_{min\ A}4$: Global alignment: The emergent tendency to increasingly align situational representations with an interlocutor intra- and inter-conversationally with increasing frequency of encounters (based on F5)

as well as at least one of the following IAT/CAT-specific criteria:

$C_{min\ B}1$: Primal alignment: The ability to — based on a higher goal or per default — *focally align* its output to the input and prior inputs through a priming mechanism on lexical, syntactic and conceptual levels (or on one of the levels if full percolation is shown) (based on F2)

$C_{min\ B}2$: Deliberate accommodation: The ability to — based on a higher goal — (**C5a**) *converge*, (**C5b**) diverge or (**C5c**) maintain the style of its output structure depending on the (total prior) input through deliberate action on lexical, syntactic and conceptual levels (based on F2)

For the *realization* of *fully* human-like accommodative behavior, it would need to fulfill all criteria $C_{min\ A} \cup C_{min\ B} \cup C_{full}$ with C_{full} being the set containing the additional criterion of

$C_{full}1$: Primal repair: The ability to self initiate a primal repair mechanism to align misaligned representations (based on F2)

Note that the *higher goal* for convergence and divergence does not have to be emergent but could reasonably also be hard-coded[4]. Further, an LLM depicting only deliberate maintenance through a hard-coded goal of the LLM not taking into account the interlocutor's language would not have 'accommodatability' according to this definition since behavior of only maintenance cannot lead to *global alignment*.

[3] or any other communicating entity

[4] As one could argue the affective and cognitive goals are in humans as well.

4.3 A Cognitive Account of Transformer Architecture

The foundational transformers architecture $T(i) = o$, as introduced by Vaswani et al. (2017) describes a model taking a certain language input i and gives out o which relates to i in a certain manner. To make this relation[5] more human-like, certain characteristics of the input i need to be inferred. Quantization of the semantic meaning of tokens (base constituent of i, e.g. words) is undertaken using word embeddings, i.e. converting sequences of characters into representational vectors using a function/neural network with consistent weights. Similarly, the position of tokens is encoded (Vaswani et al. 2023, pp. 5–6).

The complete contextual reliance on an *attention mechanism* presents the pivotal characteristic of transformers, differentiating them from the previously prevalent models using *long short-term memory (LSTM)* (cf. Hochreiter and Schmidhuber 1997). While semantic information can be represented using the semantic spaces spanned by embedding functions, sentences are not only an aggregation of word-meanings. They contain implicit structures informing differing significance of tokens, such as syntax. To be able to be trained on recognizing these structures, transformers possess "self-attention". For each token, its relation to all other tokens is represented through calculation of similarity values between tokens using what is called "queries" and "keys" and abstracting a self-attention score using "values" (Vaswani et al. 2023, pp. 3–7). Through "residual connections", Vaswani et al. (2023, p. 3) are able to maintain the identity of the embeddings for each token and its self-attention values. This enables the learning of common structures in human utterances.

While the original paper by Vaswani et al. (2017) envisioned a separate encoder and decoder (related through encoder-decoder attention), there are now many prominent models which employ only decoders (OpenAI 2023b, cf. e.g.) and which employ only encoders (Devlin et al. 2019, cf. e.g.). While the former has proven effective in language production tasks, the latter is typically only used for language comprehension tasks.

4.4 Examination of Structural Criteria

For any kind of accommodative output for a given input, the linguistic levels of accommodation need to be captured by the transformer. While this is trivial for lexical information, as words are embedded and thus represented through the embedding semantic space, the concept of syntax is less directly present in the architecture. Yet, there is conclusive evidence of activation patterns of the self-attention mechanism corresponding to syntax (Mareček and Rosa 2019), also suggesting a context-dependent representation of concepts. With the possibility for representation given, the behavioral arguments are examined regarding transformers.

4.4.1 Priming and Deliberation in Accommodation

Primal alignment presupposes the existence of a priming mechanism as mentioned by Pickering and Garrod (2004). Transformer models of both discussed architectures can be considered

[5] In the context of its original proposal, the aim was better implementation for language translation tasks

to employ a priming mechanism since their output is mechanistically shaped by the input. The attention mechanism allows for certain parts of the input to be more influential in shaping the output which can be likened to human priming. Given this, the (H1) *existence of measurable focal alignment in LLMs* is a reasonable hypothesis.

Contrasting priming, assessing the utility of convergence, divergence or maintenance for a terminal goal requires high-level decision making processes. Since there is no dedicated component for such assessments in transformer architecture, this requirement could be assumed to be not met. However, the emergence of properties observed with the scaling of LLMs could make these models exhibit pseudo-decision-making based on intricate pattern-repetition (cf. Zarzà et al. 2023)[6]. Regardless, any decision-making would also require a terminal goal which LLMs are not equipped with. Thus, it is proposed that (H2) *only the (automatic) alignment of LLMs can be both expected and tested in current-stage LLMs.*

4.4.2 Repair and Global Predispositional Alignment

Priming and deliberate repair strategies can be considered a bridge from focal to global alignment as they are needed to (1) recognize misalignment of representations and then (2) change these representations accordingly, resulting in more global and subsequently more focal alignment. Whereas focal alignment through priming appears plausible in LLMs, this is not the case with repair processes. Transformer-based LLMs lack the proactivity and the recognition function (as both are not part of the architecture) needed to self-initiate repair. More centrally, to complete even other-initiated repair, a model would need to be able to change its representations of the tokens involved (e.g. when confronted with the ambiguity of the word "book" regarding a digital or paper version) according to the situation. As discussed above, the semantic representation of the word at hand is in its entirety defined as the weights the transformer uses for word-embedding, i.e. the embedding function which, more generally, defines the semantic space the transformer operates in. Changing this embedding outside of dedicated training and using only the conversational context — while technically implementable[7] — is not part of the transformer design. Even when encountering out-of-vocabulary words, transformers use pre-generated, sub-word embeddings to derive a word's meaning[8]. The reason for this fixedness could be found in issues of containability, safety, consistency and power, all requiring this technology to be an *immutable mobile* (cf. section 2.2). Transformers consequently have what Lücking and Mehler (2013, p. 3) refer to as "extrinsic" semantics as opposed to the "intrinsic" semantics of humans[9]. Thus, transformers lack (1) the recognition function and (2) proactivity for self-repair[10] and — crucially — (3) malleability of

[6]The possibility of decision-making emergence, including schemes such as chain-of-thought (cf. Wei et al. 2022), cannot be discussed in-depth in this thesis.

[7]although challenging given the extremely small amount of data compared to regular pre-training data sets

[8]There are also language models based on characters rather than words (Likhomanenko, Synnaeve, and Collobert 2019).

[9]It should be noted that the semantic embeddings of transformers such as GPT are trained based on conversational data. However, this is done (1) in-transparently, (2) globally for the entire model and (3) incrementally and manually as opposed to automatically and continuously.

[10]Even without self-initiated repair, one could ascribe a chatbot limited 'accommodatability' since repair exchanges could still be undertaken, if only started by the user. As has been shown, this is not the case with

its representation in form of the word embedding function. Similarly, any more complex global disposition would need to reside in a preserved and editable basis which is not given in current transformer architecture. This could either take the form of a dedicated memory module or be the product of emergence on the basis of embedding-weights, which should be editable for this, as well. Despite this lack of continuous, inductive semantic learning, there might be an effective approximation of repair and alignment contained by the attention-window of the model. Since the individual word's embeddings are (through residual connections) combined in a structure-preserving, additive way with the values generated by the attention mechanism, the resulting representation of a given word in a certain context is different from the same word in another context. Given the context of an entire conversation, the *effective* embedding of a word can thus, be perceived to change on the basis of attention. This might include the context of an other-initiated repair exchange, thus simulating (but not realizing) conceptual alignment[11]. Thus, it is proposed that (H3) *there might be additional simulated global alignment within LLM-chatbot's context window* and (H4) *given prolonged conversations, there should be a recognizable decrease in global alignment, corresponding with context window size.*

4.5 Possible Transformer-Based Implementations of 'accommodatability'

There have been two broad directions of development towards "more social" AI. Firstly, there is the goal of bioplausibility, i.e. making ANNs resemble BNNs more (cf. Voelker 2015; Bolotta and Dumas 2022; Borzenko 2010; Tognoli, Dumas, and Kelso 2018; Bal and Sengupta 2023) as well as artificial mirror neurons (cf. Borenstein and Ruppin 2005; Shapshak 2018; M. Stamenov, Gallese, and M. I. Stamenov 2002)). Secondly, there it the attempt of "LLM-emergence-wrapper" models using prompt engineering to elicit certain emergent behavior (cf. Lin et al. 2023; Deng et al. 2023). Given the controversial state of research on social cognition, bioplausible approaches cannot effectively be evaluated for their potential in facilitating any aspect of AI_M. Instead, prospects of possible transformer-architecture-modifications[12] are presented. As elaborated, there are several demands for *minimally viable, realized* 'accommodatability' — specifically, a realized form of alignment behavior. Current transformers are lacking structure-dependent abilities A:

Argumentative Pillar 4.1

A1: *Long-term memory*

A2: *A misalignment recognition function*

A3: *Proactivity*

current architecture.

[11]This *attention-simulated alignment* is structurally similar to the *context window* itself, simulating conversational memory through appending conversational context to the input.

[12]The opposite approach, not in the focus of this work, might be to increase the language abilities of ANNs that are already capable of inductive learning.

Suggestions for consideration when implementing these demands are touched on here. Only appending an external long-term memory module (cf. Wang et al. 2023)[13] would allow for a certain "workaround" regarding the context window (i.e. the transformer maximum token amount) limitation of simulated focal alignment but not change the aforementioned extrinsicality. Having the extent of aligned conversation be a function of the transformer's token limit instead of the memory module's capacity would make it immensely more computationally expensive, not modularly adaptable in-deployment and not indefinitely maintainable[14]. Thus, a model might be needed which is able to self-adjust its embedding weights, enabling a change of representations. In two transformers with external memory modules of which one also has adjustable embeddings, the simulation of alignment of the former would reach its limits and thus show a stark, noticeable behavioral contrast to the latter. The implementation of such a self-adjusting procedure would need misalignment recognition which could be a machine learning algorithm comparing the pure word embedding of a word with its attention-embedding and noticing a mismatch as manifested through a discrepancy. Such a function might be trained in a GAN-eske set-up in which the extended transformer acts as a generator with only the weights of the misalignment recognition function adjusted in training. Existent ML-based tools for quantifying alignment in textual data (cf. section 5) might be used as a discriminator.

Regarding proactivity, there are attempts to use "chain-of-thought prompting" (Naveed et al. 2023, p. 8) to simulate proactivity and elicit better clarification requests (i.e. repair initiations) (cf. Deng et al. 2023). Further, such prompting approaches on existing LLMs could also be used to introduce the ability to deliberately accommodate to LLMs. Given the fact the differentiation between automatic alignment and deliberate accommodation resembles that of dual-process theory in psychology, the former can be classified as system 1 and the latter as system 2. Lin et al. (2023) propose a prompting framework implementing the decision between system 1 and system 2 given a certain context. This framework could be adapted to instead let the actant "decide" between utterances to use obviating automatic alignment in favor of deliberate communicative accommodative strategies.

Goebel, Siekmann, and Wahlster (n.d., pp. 292–299) attempt to model alignment in AI, focusing on a joint attention mechanism but disregarding alignment on many other levels

4.6 Conclusions on LLM 'accommodatability'

In summary, the analysis of human cognitive structures involved in alignment, such as mirror neuron systems, has not yet progressed to a point at which structurally comparing human and transformer cognition would be fruitful. Regardless of evidence, the equivalences made in such a comparison would be at risk of metaphorically misrepresenting the relevant parts of human language cognition. More reasonable is the examination of LLM-structures to feasibly

[13] Such an implementation, capable of storing attention data, could even enhance the simulation of alignment while not being able to realize it.

[14] For every token, the amount of calculations to be performed is a linear function of the number of other tokens. Thus, the resource requirement for increasing the context window grows quadratically.

support the largely *behavioral* combined principles of IAT and CAT laid out in section 3.4. This exercise points to four testable hypotheses H on the 'accommodatability' of (current) transformer-based LLMs:

Argumentative Pillar 4.2

H1: *There is measurable focal alignment in LLMs regardless of conversation length*

H2: *Only the (automatic) alignment of LLMs can be both expected and tested in current-stage LLMs*

H3: *There might be additional simulated global alignment within the LLM-chatbot's context window*

H4: *Given prolonged conversations, there should be a recognizable decrease in global alignment, corresponding with context window size*

Further, there are four structure-enabled abilities **A** needed for minimal viable alignment missing in (current) transformer models. Possible implementation of those using current technologies were alluded to: (**A1**) long-term memory , (**A2**) misalignment recognition, (**A3**) proactivity and (**A4**) the ability to continuously change semantic representation.

These conclusions show which specific aspects of chatbot-behavior should be empirically examined to infer the possibility of not only simulated but *realized* accommodation. The possible design of such studies as well as studies measuring *effective* accommodation is discussed in section 5.

5 A new Approach to Accommodation Measurement

Whereas the previous chapter discussed the more fundamental architectural underpinnings of simulated (distinct from realized) 'accommodatability', the phenotypical examination of LLM utterances has the starting point of existing work on accommodation-measurement. Based on this, schemes for empirical research are suggested, taking into account the findings in section 4. While *deliberately* accommodating LLM-systems are possible, current LLM-chatbots are incapable of such behavior (cf. section 4). Thus, in this section, only the quantitative effect of *effective convergence* behavior will be assessed[1].

When approaching existing research on convergence, two classes of examination need to be differentiated: interventional *trials* and observatory *corporal studies*. While the former, and preferred form of IAT research, enable the tight control of conversational structure and experimental set-ups designed to specifically enable alignment on a certain level, they are not without flaws. Generalizing from very restricted circumstances for communication to broader sociolinguistic phenomena has been criticized in the past. Moreover, predictions stemming from IAT have often times not manifested when tested in more naturalistic settings. These more day-to-day forms of communication are better observed without the behavior-influencing environment of scientific experimentation. Undertaking analyses of such data are corporal studies, which more often employ CAT. Yet, these observational studies present issues regarding the clear separation of factors leading to observed behavior[2].

Crucially — especially given the broadness of the phenomena — the behavioral predictions of CAT and IAT for convergent/aligning behavior do not differ meaningfully. This enables here the examination of two specific models, citing either tradition, which both pioneered the use of semantic spaces for convergence measurement in plain-text data.

5.1 The Use of High Dimensional Semantic Spaces in Convergence Measurement

Traditional studies on attunement use various approaches, including the simple measure of recurrence (Fusaroli, Rączaszek-Leonardi, and Tylén 2014, p. 151), correlation-based measures (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011, p. 2) and *difference in difference* (as critiqued by Cohen Priva and Sanker 2019). While these approaches have been found to be effective, they rely on human approximations on how to best describe accommodation effects and cannot properly mathematically frame phenomena such as similarity of concepts *in context*.

[1] Given (simulated) intentionality, more intricate studies would have to be designed to discern various additional factors.

[2] It should be noted that many studies fall between these ideal types.

5.1.1 ALIGN: Assessing Alignment of Lexis, Syntax and Concepts

To better approach these challenges, N. D. Duran, Paxton, and Fusaroli (2019, pp. 420–422) introduce *ALIGN*[3], a framework to "quantitatively and reproducibly measure turn-by-turn alignment across syntactic, lexical, and conceptual levels of language" (N. D. Duran, Paxton, and Fusaroli 2019, p. 422) considering time as well as directionality of alignment[4]. Assuming pre-structured data, this is done in two phases.

First, *ALIGN* pre-processes the text through parameterized removal and correction of all elements that are not standardized words (N. D. Duran, Paxton, and Fusaroli 2019, p. 423). Subsequently, it tokenizes and optionally lemmatizes each word to then tag part-of-speech (PoS). In a second phase, these pre-processed lexical and PoS sequences are converted into n-grams (within a set range) whose frequency within each turn level (not in the conversation) is represented as a vector. **Lexical alignment** is calculated using turn-by-turn cosine similarity of the lexical n-gram vectors while **syntactic alignment** results from the same process using only some of the PoS n-gram vectors[5](N. D. Duran, Paxton, and Fusaroli 2019, p. 424).

Conceptual alignment requires additional steps for the semantic contextualization of concepts. This is done by selecting a pre-generated high-dimensional semantic space (HDSS) or creating one given sufficient corpus size. Through this HDSS, words are each converted into high-dimensional semantic vectors (HDSV) which are then additively composited into a vectorial representation on the level of an utterance. Words of especially high- and low-frequency are not included as to prevent noise in the representation. The resulting utterance-vectors can then be compared using cosine similarity (CoS) to calculate conceptual alignment (N. D. Duran, Paxton, and Fusaroli 2019, pp. 424–426). An issue in measuring attunement is that both interlocutors might be similarly influenced by the task instead of by each other (Cohen Priva and Sanker 2019, p. 3). A solution to such biases is the creation of synthetic "control dialogues" through random recombination of utterances and 'utterers', resulting in a baseline measurement (cf. Oben 2015; Danescu-Niculescu-Mizil, Gamon, and Dumais 2011, p. 25). N. D. Duran, Paxton, and Fusaroli (2019, p. 426) generate such synthetic dialogue, with the restraint of preserving turn order. This modification is used to prevent certain structural parallels, which occur with the progression of conversation, to be misclassified as alignment.

ALIGN offers a means of assessing lexical, syntactic and conceptual alignment which is both reliable and uses (synthetic) control data. Yet, there are elements of this approach which might distort the results. Importantly, in conceptual alignment, the use of a single HDSS (be it pre-trained or trained on the data in focus) together with the additive composition of utterance does not consider relational context, thus constituting fixed semantics. The alignment of indexicality (cf. Pfadenhauer 2013, p. 142) of certain expressions is not well encapsulated because the semantics enrolled by ALIGN are extrinsic[6]. Furthermore, the n-gram approach limits the extent of analysis possible regarding lexical and syntactic alignment.

[3]Openly available on https://github.com/nickduran/align-linguistic-alignment.

[4]ALIGN only assesses one manifestation of alignment in behavior, i.e. "linguistic repetition" but not other measurables (cf. Pickering and Garrod 2021, p. 129).

[5]This is done to prevent correlational effects between both, specifically "lexical boosting".

[6]The authors acknowledge the issue in passing but do not relate it to (pragmatic) alignment as a concept (N. D. Duran, Paxton, and Fusaroli 2019, p. 426).

5.1.2 QuantCAT: Assessing Convergence Using BERT

An especially recent approach for quantifying attunement measurement is found in *QuantCAT*[7] by Rosen (2023), referencing CAT as a central theory. They define *convergence* as in-group attunement[8] and *accommodation* as the more general phenomenon describing sequential attunement of language to an interlocutor (Rosen 2023, p. 61). Instead of quantifying lexical, syntactic and conceptual attunement, the focus of this framework lies solely in convergence of different concepts of the same lexical word. In contrast to ALIGN, these words need to be determined manually before processing; The model only compares occurrences of this set. Rosen (2023, pp. 63–65) utilizes the language representation model *Bidirectional Encoder Representations from Transformers* (BERT) (cf. Devlin et al. 2019). The proposed algorithm first represents the selected pivotal words in vectors using BERT (let $E_{x,w'}$ be the vector matching keyword w' in sentence x) and then pair-wise compares them using the probabilistic complement of CoS, i.e. "cosine error", here $CoE(E_{x,w'}, E_{y,w'}) = 1 - CoS(E_{x,w'}, E_{y,w'})$ with x and y being different sentences.[9]. CoE is used to infer a confidence of similarity $P(E_{x,w'}|E_{y,w'})$ using a half-Gaussian distribution function:

$$P(E_{x,w'}|E_{y,w'})_\sigma = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{CoE(E_{x,w'}, E_{y,w'})^2}{2\sigma^2}\right) \tag{5.1}$$

with σ as a manually set scaling factor determining the slope of the confidence distribution (Rosen 2023, p. 64). These comparisons are represented in a "similarity matrix" $M^{X;Y}$ with every cell as similarity confidence between the sentences with data set positions x and y (Rosen 2023, p. 65). This matrix is subsequently used to calculate convergence and accommodation scores. For in-group convergence, the mean of intra-group similarity is divided by the mean of inter-group similarity (Rosen 2023, p. 65). Rosen (2023) defines accommodation only as a measure that, requiring similarity functions for key-word w', relates an utterance a by actant α to an utterance b by actant β with $b \in B_{\text{prior}}$ and B_{prior} as the set of all utterances uttered by β and before a. To foreclose measurement of convergence where there is uniform likeness, the similarity measure of a and b is divided by the mean of the pair-wise similarity of prior utterance a to all other utterances before a:

$$A^{X;Y} = \left\{(a;b) : |B_{\text{prior}}| \left(\frac{M^{X;Y}(a;b)}{\sum_{B_{\text{prior}}} M^{X;Y}(B_{\text{prior}}, b)}\right)\right\} \tag{5.2}$$

QuantCAT avoids the semantics issue of ALIGN by using the attention mechanism of transformers where ALIGN assumes a fixed conceptual localization of each single word, then composed through addition (N. D. Duran, Paxton, and Fusaroli 2019, p. 425). Yet, the use of pre-trained semantic representations[10] likewise constitutes an extrinsic semantic space. This allows for an (albeit smaller) semantic bias to remain in the model[11]. For instance, given a

[7] Openly available on GitHub

[8] similar to what Danescu-Niculescu-Mizil, Gamon, and Dumais (2011, p. 4) call *cohesion*

[9] For this, see the file "probability.py" in the model's repository

[10] which are derived from the 7th hidden layer of BERT (Rosen 2023, p. 70).

[11] enabled by the implicit assumption that all language is generalizable

certain embedding function similar to one actant's semantic space (and less so to the other's) might provide a more nuanced representation of this actant's utterances, introducing distortion into the pair-wise comparisons and thus all subsequent accommodation scores.

5.2 A New Quantified Assessment Framework

As has been demonstrated, both ALIGN (cf. N. D. Duran, Paxton, and Fusaroli 2019) and QuantCAT (cf. Rosen 2023) introduce more versatile measurement schemes for convergence than traditional or probabilistic (cf. Danescu-Niculescu-Mizil, Gamon, and Dumais 2011) approaches. However, ALIGN's usage of an (albeit potentially corpus-specific) global, context-less embedding function and QuantCAT's usage of pre-trained BERT word embeddings, introduces systemic distortion into the data.

Its focus on keywords makes QuantCAT a suitable tool for analyzing accommodation *in discourse*. However, since the aim of this proposal is to quantify a *general* sense of attunement between two people, two adjustments are needed: (1) The ability to automatically select the pivotal words while (2) retaining a fully retrospective analysis as opposed to the turn-by-turn analysis of ALIGN. The latter modification is needed because alignment may manifest between certain patterns of non-contiguous turns, not only to the other actant's last utterance[12]. Similarly, the n-gram approach to lexical and syntactic alignment by N. D. Duran, Paxton, and Fusaroli (2019), while providing an approximation of alignment, is not equipped to detect more complex patterns.

To address these issues, a new approach is proposed, based on the previous vectorial approaches and extending them. The following section details the steps of insight-generation for this approach: (1) data generation (resourcing), (2) Natural Language Processing (NLP), (3) modeling and evaluation.

5.2.1 Experimental Data Generation

While many of the methods detailed below can also be used on corporal data, the subject of HC communication strongly suggests measurement under controlled circumstances, enabling the derivation of statistically sound conclusions.

To these ends, a trial would be conducted online or in a controlled environment, guaranteeing for isolation of each subject. The (operationalization-dependently) selected set of participants $P = P_H \cup P_L$ consisting on humans capable of textual communication P_H and LLM-instances with chat interface P_L accounts for $n = |P| = |P_H| + |P_L|$ involved actants.

As CAT describes, even in HH communication the extent and direction of accommodation is heavily mediated through social perceptions and expectations of perception. The communication with an AI system could be expected to be distorted by prior assumptions about chatbots (cf. Hildt 2021). To eliminate this factor, omnidirectional *ontological blindness* is facilitated. Through the interactive isolation, i.e. solely text-based communication and no additional declarations of (ontological) identity, the only information permeating between two actants is that that is captured and analyzed. To prevent any suspicions of artificiality

[12] As well as the non-measurable disposition and any other possible factors.

from arising in human subjects, the timing of message submission is considered. The nonhuman communicators could be assigned a random realistic typing speed v_t[13] to be multiplied by message length to generate the delay in response[14]. This too addresses the element of communicative *indexicality* (Pfadenhauer 2013, p. 142), which is vital for interactivity. The indexing quality of certain phrases is contained within the examined text and should therefore be captured by the conceptual convergence's usage of context.

While the exact prompting of conversation topics or joint tasks falls within concrete experimental design to accommodate a specific research question, there are considerations relating to the "naturalness" of communication. The aim of enabling quantification of accommodation in *socially situated* dyadic, plain-text communication necessitates a certain freedom of communication not possible within the strictly task-focused experiments reviewed. It would thus be advised to prompt more organic, spontaneous communication. Accordingly, the need for strong accommodation effects would be lessened by the *relational* interpretability of results for LLM-'accommodatability' using the other ontological pairs as reference.

This approach aims to asses attunement within the dimensions (cf. Rasenberg, Özyürek, and Dingemanse 2020, pp. 8–10) of "sequence", "meaning" and "form", yet not "modality" and "time". Both incorporating reaction time[15] and modality would introduce unnecessary and potentially counterproductive complexity to the research.

5.2.2 Natural Language Processing

The data generated experimentally are uniform because of the total control of input modality. This obviates the need for them to be cleaned or (potentially introducing distortive assumptions) parsed into speech turns (cf. N. D. Duran, Paxton, and Fusaroli 2019, p. 428)).

The data are stored in database DB_{raw} with utterances/turns as a base unit: Directly contiguous utterances of the same person are precluded by study design. Each utterance has a value for a unique `UtteranceID`, a connecting `ConversationID`, `UtteranceContentRaw`, `AntecedentUtteranceID` providing the ID of the last prompt to the actant, `ActantType` which is either human or nonhuman, and `TurnNumber`[16], defining the turn number of the utterance within the conversation with `(ConversationID, TurnNumer)` and `UtteranceID` being unique keys for each other (constituting a bijective relation).

After data collection is completed, any non-word elements are removed. Then, different NLP operations are performed leading to different database versions[17] for subsequent analysis.

Word tokenization and POS tagging form database DB_{syn} and word tokenization with stop word removal results in database DB_{lex}. Meanwhile, concept tokenization without any Named Entity Recognition, stemming, lemmatization or lowercasing is performed to result in concept database DB_{con}.

[13] Requiring exact operationalization

[14] The change of message length, however, could be considered as an accommodative behavior and thus not warrant mitigation

[15] As would be very feasible.

[16] Despite being logically redundant, this is included for clarity of representation.

[17] Regardless of implementation, separate databases is chosen for notation purposes.

This overall data structure can be queried to provide the basis of all further refinement and analysis.

5.2.3 Modeling of Convergence

Given actant α with utterances $a \in A$ and actant β with utterances $b \in B$ having engaged in dialogue, relating the complete sets of utterances of α and β $A \sim B$, there are four modeling-decisions laid out below. For each linguistic dimension dim $\in$ {lexical (lex), syntactical (syn), conceptual (con)}, these are in increasing level of abstraction:

1. Declaring formalized analytical base-units on which the similarity functions can operate.

2. Defining an adirectional similarity function for utterances, $\text{sim}_{\text{dim}}(a, b) = \text{sim}_{\text{dim}}(b, a)$

3. Defining a directional convergence measure on *utterance level*, utt: $\text{conv}^{\text{utt}}_{\text{dim}}(a, B_{\text{prior}})$ or $\text{conv}^{\text{utt}}_{\text{dim}}(a, A_{\text{prior}})$ respectively, using utterance a of actant α and a subset $B_{\text{prior}} \subset B$ or $A_{\text{prior}} \subset A$ of all previous utterances of either β or α itself[18].

4. Defining a directional measure of convergence on an *actant level* act: $\text{conv}^{\text{act}}_{\text{dim}}(\alpha, \beta)$ between actant α with either β or with α itself, taking into account either actant's utterances and their dialogical relation.

Shaping the resolvement of these modeling necessities are three goals: (1) the aim of reducing influence of semantic extrinsicality while (2) enabling complex contextual analysis and (3) adhering to the common propositions defined in section 3.4 as well as the expectations expressed in section 4.6.

Base Units

Different abstraction levels can be considered for the base-unit of this framework: characters (cf. Likhomanenko, Synnaeve, and Collobert 2019), tokens, words, utterances, and larger structures. In accordance with QuantCAT and ALIGN, *utterances* are chosen for their ability to fulfill the criterion of decomposability of conversation (cf. **AI$_{\text{M}}$ 1**)[19] without the risk of noise from possible intra-utterance structures. Given the clear advantages of vectorial representations (of frequency and context) in the existing methods, the utterances are encoded as vectors.

Since the linguistic levels require different representation, there are differing vector representation functions vec() needed. Utterance $a = a_1, a_2, \ldots, a_n, i \in \{1, \cdots, |a|\}$ presents a sequence of $|a|$ tokens a_i. Following N. D. Duran, Paxton, and Fusaroli (2019), n-grams, i.e. contiguous sequences of n items, are created from a. For lexical and syntactic alignment, the tokens in a are turned into a multiset of n-grams $G_a = g^1_1, \cdots, g^n_{f(n)-1}, g^n_{f(n)}$ with n unique

[18] This option in both convergence measures is for increased statistical rigor. A similar concept is β_{S_b} by Cohen Priva and Sanker (2019, p. 2).

[19] To fully assess interactivity, the study would have to be repeated with the same subjects to conform to both parts of **AI$_{\text{M}}$ 6**.

n-grams $g^i, i \in \{1, \cdots, n\}$, each having $f(n)$ occurrences in the multiset. For one utterance and each n-gram length n, a has one lexical representation $\vec{a}_{\text{lex}} = \text{vec}_{\text{lex}}(a, n)$ and one syntactic representation $\vec{a}_{\text{syn}} = \text{vec}_{\text{syn}}(a, n)$[20]. For lexical alignment, frequencies $f(g^i)$ of n-gram's g^i occurrence in G_a are represented as

$$\vec{a}_{\text{lex}} := [f(g^1), f(g^2), \cdots, f(g^n)] \tag{5.3}$$

Similarly, the equivalent representation of syntactic alignment is based on $f'(g^1)$ which yields the number of occurrences of g^i in G_a only counting cases in which two equal syntactic n-grams do not share the same lexical n-gram, to prevent lexical boosting effects (cf. N. D. Duran, Paxton, and Fusaroli 2019, p. 424):

$$\vec{a}_{\text{syn}} := [f'(g^1), f'(g^2), \cdots, f'(g^n)] \tag{5.4}$$

For concepts, a different approach is needed. Contrasting with the other two levels, *aggregate* utterance vectors as a base unit of *conceptual* alignment do not enable optimal representation. As discussed before, ALIGN additively composes semantic vectors to create a utterance-level vector, thereby misrepresenting relational aspects. To avoid this, the inclusion of context is crucial. Let the foundational unit for assessing conceptual structure be that of the concept. Procedurally, specific tokenization[21] would be used so that phrases of different forms (e.g. "rail", "railway" "rail pass", "Japan Railways") are each represented as one conceptual token. Let I be the set of all conceptual tokens generated and $K \subset I$ a set of key-concepts. Given a key-concept $\kappa \in K$, each a has $k = |\kappa : \kappa \in u|$ conceptual representations $\vec{a}_{\text{con}} = \text{vec}_{\text{con}}(a, \kappa)$. The generation of conceptual vectors $\vec{a}_\kappa$ could utilize BERT or another transformer model capable of implementing $\text{vec}_{\text{con}}(a, \kappa)$, returning a vectorial representation of κ, *independently* representing both the global[22], semantic and the local (utterance a), contextual dimensions of κ.

$$\vec{a}_\kappa := \text{vec}_{\text{con}}(a, \kappa) \tag{5.5}$$

Thus, the vectorial representations on the level of utterance a, $\vec{a}_{\text{lex}}$, $\vec{a}_{\text{syn}}$ and $\vec{a}_\kappa$, present the basis for all further analysis of accommodation.

Similarity Norms

The vectorial representation of utterance base-units suggests CoS as a similarity function. CoS has been thoroughly established in the field of NLP (cf. Sitikhu et al. 2019) as a (semantic) difference norm and will therefore be used here. Given utterance vectors $\vec{a}$ and $\vec{b}$, hereafter a and b when irrespective of linguistic level, the similarity function $\text{sim}(a, b)$ is introduced. It is important to note that, unlike ALIGN, QuantCAT does not only use CoS (CoE) as measure

[20] The ability to choose a range of n-gram lengths is considered in section 66.

[21] But not lemmatization, since the context of application and thus the inferred meaning of words differs by morphology and necessitates differing— albeit closely related —concepts. Based on these considerations, *morphological alignment* might be assessed, as well.

[22] At this point, the use of a static and potentially extrinsic semantic space, introduces complications for measurement.

but further transforms the result using a (half) Gaussian distribution to infer a measure of confidence. However, this step requires assumptions about the distribution of accommodation behavior. Since these cannot be thoroughly substantiated at this point (especially given the nonhuman part of the statistical population), the similarity of utterances a and b given a vectorial representation $\text{vec}(a) = \vec{a}$ is defined as:

$$\text{sim}(a,b) := \text{CoS}(\vec{a},\vec{b}) = \cos(\theta) = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}, \tag{5.6}$$

Since the vectorial representation functions $\text{vec}_{\text{lex}}(a,n)$ and $\text{vec}_{\text{syn}}(a,n)$ require n-gram length n as input and $\text{vec}_{\text{con}}(a,\kappa)$ requires key-word κ, $\text{sim}(a,b)$ given a specific vector function is written as $\text{sim}(a,b,n)$ and $\text{sim}(a,b,\kappa)$ respectively.

Note that the range of similarity on lexical and syntactic levels differs from that on the conceptual level. Reason for this are the non-negative values of frequency vectors $\vec{a}_{\text{lex}}$ and $\vec{a}_{\text{syn}}$, contrasting semantic-contextual vectors $\vec{a}_{\text{con}}$:

$$\text{sim}(\vec{a}_{\text{lex}},\vec{b}_{\text{lex}}), \text{sim}(\vec{a}_{\text{syn}},\vec{b}_{\text{syn}}) \in [0,+1] \tag{5.7}$$

$$\text{sim}(\vec{a}_{\text{con}},\vec{b}_{\text{con}}) \in [-1,+1] \tag{5.8}$$

Utterance Level Convergence Measures

Thus far introduced are a universal utterance similarity function $\text{sim}()$ for differently generated utterance vectors of lexis $\vec{a}_{\text{lex}}$, syntax $\vec{a}_{\text{syn}}$ and concepts $\vec{a}_{\text{con}}$. Given these foundations, there are several options for designing utterance-level convergence measures conv_{dim} of linguistic dimension dim. N. D. Duran, Paxton, and Fusaroli (2019, pp. 424–425) provide $\text{conv}(a,b) = sim(a,b)$ with $(a,b) \in C_{a,b}$, only examining the similarity of a to *antecedent* utterance b within the ordered sequence of utterances $C_{\alpha,\beta} = A_{\text{prior}} \sim B_{\text{prior}}$. Contrasting this and in accordance with Rosen (2023), utterance-convergence[23] is here defined between *any* utterance a by α and b by β, accounting for overall utterance similarity. It utilizes the conceptual vector function $\text{vec}_{\text{con}}(a,\kappa)$ and is thus contingent on κ. Given all prior utterances by β, $b_{\text{prior}} \in B_{\text{prior}}$, utterance-level convergence of concepts is defined as:

$$\text{conv}^{\text{utt}}_{\text{con}}(a,b,\kappa) = \frac{\text{sim}(a,b,\kappa)}{\bar{x}_{b_{\text{prior}}\in B_{\text{prior}}}(\text{sim}(a,b_{\text{prior}},\kappa))} \tag{5.9}$$

This measure is, however, still dependent on key-concept κ. Yet, enabling further aggregation, an independent measure of utterance-to-utterance convergence is needed. For this, there needs to be a selection mechanism for key-concepts $K \subset I$. This could involve using Extractive Summarization (e.g. through BERT (cf. Y. Liu 2019)) to collect the concept-tokens with most conversational relevance. Thus, the κ-independent measure of conceptual convergence is:

$$\text{conv}^{\text{utt}}_{\text{con}}(a,b) = \bar{x}_{\kappa\in K}(\text{conv}(a,b,\kappa)) \tag{5.10}$$

[23] In his terms, accommodation

For lexical and syntactic convergence, the measures of utterance-convergence are respectively $\text{conv}_{\text{lex}}(a, b) = \text{sim}(a, b)$ and $\text{conv}_{\text{lex}}(a, b) = \text{sim}(a, b)$.

Conversation Level Convergence Measures

In the terms of IAT, the thus far established measure of convergence could be called *focal* convergence. With the aim of assessing accommodation as defined through AI_{M} 2-6, a measure of *global*, aggregate convergence is needed.

The choice of how to define conversation/actant level convergence[24] $\text{conv}^{\text{act}}_{\text{lex}}$, $\text{conv}^{\text{act}}_{\text{syn}}$ and $\text{conv}^{\text{act}}_{\text{con}}$ respectively contains assumptions about the precise effect mechanisms underlying the three possible distinct (cf. Ostrand and Chodroff 2021, p. 2) phenomena and should therefore be explicated in the discussion of findings. For two actants α and β, ALIGN defines $\text{conv}^{\text{act}}(\alpha, \beta) = \bar{x}_{a \in A}(\text{conv}^{\text{utt}}(a, b_{\text{antecedent}}))$ as the mean utterance-convergence over all utterances A by α and the respectively antecedent utterance $b_{\text{antecedent}}$ of β. This only considers direct convergence effects, and is this not suitable for more complex convergence patterns. Given this limitation of ALIGN and the keyword-dependency of QuantCAT, a new approach is considered. To achieve a conversation level-convergence score, there are two aggregations needed: $\text{conv}^{\text{utt'}}(a, B_{\text{prior}})$, measuring aggregate convergence of a fixed a to all respectively preceding utterances B_{prior} of β and conv^{act} averaging $\text{conv}^{\text{utt'}}$ over all utterances A of α.

The former depends on the question of how to weigh previous utterances based on their recency. Calculating the arithmetic mean would imply that all previous utterances evoke convergence equally. This conflicts with the perspective of CAT and especially IAT, whose priming mechanism relies on temporal proximity. Thus, a more sophisticated weighting function $\omega(d) = w$ is introduced, giving out a weight w, given turn-distance function between utterances $\text{dist}(a, b) = d$ with $d \in \mathbb{N}^+$. Given prior utterances B_{prior} by β:

$$\text{conv}^{\text{utt'}}_{\text{dim}}(a, B_{\text{prior}}) = \frac{\sum_{b \in B_{\text{prior}}} \omega(\text{dist}(a, b)) \cdot \text{conv}^{\text{utt}}_{\text{dim}}(a, b)}{\sum_{b \in B_{\text{prior}}} \omega(\text{dist}(a, b))} \tag{5.11}$$

The exact specification of $\omega(d)$ depends on the theoretical assumptions underlying accommodation, as well as research questions. Without any assumptions, an exponential decay $\omega(t) = e^{-\lambda t}$ is suggested, whose rate of decay λ is to be found through testing values.

The weighted convergence scores $\text{conv}^{\text{utt'}}(a, B_{\text{prior}})$ of utterance a to all previous utterances B_{prior} by β can at this point be used to lastly infer the conversation/actant level convergence $\text{conv}^{\text{act}}_{\text{dim}}$ using averaging:

$$\text{conv}^{\text{act}}_{\text{dim}}(\alpha, \beta) = \bar{x}_{a \in A}(\text{conv}^{\text{utt'}}_{\text{dim}}(a, B_{\text{prior}})) \tag{5.12}$$

Aggregate Convergence Measures

The measure of directional actant-level convergence of dimension dim $\text{conv}^{\text{act}}_{\text{dim}}((\alpha, \beta)$ can be aggregated into several derivative measures, given all possible, ordered actant-pairs $(\alpha, \beta) \in$

[24] In this context, both are identical since there is only one conversation.

P[25]. The most central measurements are:

- mean actant-to-actant convergence

$$\text{conv}_{\text{dim}}^{\text{act}\to\text{act}}(\alpha,\beta) = \frac{1}{2}(\text{conv}_{\text{dim}}^{\text{act}}(\alpha,\beta) + \text{conv}_{\text{dim}}^{\text{act}}(\beta,\alpha)) \tag{5.13}$$

- mean actant-to-group convergence

$$\text{conv}_{\text{dim}}^{\text{act}\to\text{group}}(\alpha,Q) = \bar{x}_Q(\text{conv}_{\text{dim}}^{a\to a}(\alpha,q)) \tag{5.14}$$

- mean group-to-group convergence

$$\text{conv}_{\text{dim}}^{\text{group}\to\text{group}}(R,Q) = \bar{x}_{R,Q}\text{conv}_{\text{dim}}^{a\to a}(r,q) \tag{5.15}$$

With $Q \subset P$, $R \subset P$ being groups (such as humans and LLM-instances) and $q \in Q$, as well as $r \in R$ being individuals of two respective groups.

The mean group-tp-group convergence $\text{conv}_{\text{dim}}^{\text{group}\to\text{group}}(R,Q)$ can be used to compare the convergences of humans and LLMs to each other, as detailed in section 5.2.4.

In addition to conv_{lex} and conv_{syn}, the conceptual convergence measure conv_{con} can be used to infer both convergence, maintenance/non-accommodation and divergence. Consider actant $\alpha \in P$, having encountered actant $\beta \in B$ and not having encountered actant $\gamma \in \Gamma$, with the sets $\Gamma \in P$ and $B \in P$ denoting all (non-)encountered actants relative to α. Given a certain degree of selectivity (here, $\ll$ and $\gg$), there are three possible conclusions, depending on the convergence-scores:

- α conceptually diverges from β:

$$\text{conv}_{\text{con}}(\alpha,\beta) \ll \bar{x}_\Gamma(\text{conv}_{\text{con}}(\alpha,\gamma)) \tag{5.16}$$

- α maintains concepts towards β:

$$\text{conv}_{\text{con}}(\alpha,\beta) \approx \bar{x}_\Gamma(\text{conv}_{\text{con}}(\alpha,\gamma)) \tag{5.17}$$

- α conceptually converges to β:

$$\text{conv}_{\text{con}}(\alpha,\beta) \gg \bar{x}_\Gamma(\text{conv}_{\text{con}}(\alpha,\gamma)) \tag{5.18}$$

It should be noted that instead of analyzing the degree of convergence/accommodation in all levels independent of time, a progression of accommodation can also be extrapolated, e.g. enabling research on $\mathbf{AI_M}$ 4 with regard to decreasing complexity over time.

[25]For ease of reading, the round parentheses indicating that the pairs are ordered rather than being a set, i.e. $\{\alpha,\beta\}$, are omitted.

5.2.4 Signification of Central Measurements

As highlighted before, surrogate dialog-pairs *with conserved order* are essential in clearing out structural effects and providing a baseline. They also are thought to reduce bias that might arise through actant α's language behavior being situated particularly close to its interlocutor β, which could otherwise minimize measured convergence (Cohen Priva and Sanker 2019, p. 4). For this purpose, human-centered approaches only generate two classes of communicative dyads: $(\alpha_H, \beta_H) \in C$, i.e. pairs of a human actant and an *encountered* human actant as well as $(\alpha_H, \gamma_H) \notin C$, i.e. pairs of a human actant and a *non-encountered* human actant, with C denoting the set of all actual conversational pairs (as opposed to surrogate pairs).

The addition of LLM-instance α_L increases the amount of classes to eight. These are the actual pairs

$$(\alpha_L, \beta_H), (\alpha_H, \beta_L), (\alpha_L, \beta_L) \in C \tag{5.19}$$

as well as the surrogate pairs

$$(\alpha_L, \gamma_H), (\alpha_H, \gamma_L), \alpha_L, \gamma_L) \notin C. \tag{5.20}$$

Measuring convergence in these dyad-classes can be undertaken — among other avenues — via the mean group-to-group convergence $\text{conv}_{\text{dim}}^{\text{group}\to\text{group}}(R, Q)$. It enables the assessment of a facet of interactivity, i.e. $\text{conv}(\alpha_L, \beta_H)$, relative to its foundational concept of interaction, i.e. $\text{conv}(\alpha_H, \beta_H)$. Moreover, it offers baselines for accommodation in both directionalities of HC_T communication. The human accommodation towards LLM-chatbots $\text{conv}(\alpha_H, \beta_L)$ can be measured relatively to that towards humans $\text{conv}(\alpha_H, \beta_H)$. Meanwhile, the LLM accommodation towards humans $\text{conv}(\alpha_L, \beta_H)$ has as its reference LLM-to-LLM accommodation $\text{conv}(\alpha_L, \beta_L)$. Since data for all six dyad classes (eight directionalities) are generated from the same conditions and subjects are ontologically blind, there is optimal comparability.

5.3 Outlook

In summary, based on analysis of the convergence-measurement approaches of N. D. Duran, Paxton, and Fusaroli (2019) and of Rosen (2023), an experimental set-up and methods for data processing measuring human-nonhuman convergence, and ultimately nonhuman-'accommodatability', has been constructed. Reductively, the experimental set-up may be expressed as:

$$\text{CONVERGE}(P_L, P_H, \omega, v_t, z) = \text{ data on 'accommodatability' of } P_L \tag{5.21}$$

The subjects P are comprised of a set of LLM-instances P_L whose accommodation is in focus and a set of humans P_H whose sampling depends on implementation. Additional parameters include a function of divergence decay ω, an assumption about the distribution of human typing speed v_t and a conversational prompt z for all subject-pairs.

The empirical framework presented here (and every subset of it) can be used and modified to aid pioneering systematic, experimental research on a novel approach to qualifying LLM-based chatbots to be considered *conversational equals* ("vollgütige Interaktionspartner"

(Mehler 2010, p. 3)) to humans. Yet, it could also be applied, if LLMs lack effective accommodation, to serve in the detection of artificially generated dialogue-data.

While it is the aim of this work to minimize the distorting factor of extrinsic, static semantic spaces, the suggested methods nevertheless include pre-trained word-embeddings in their use of BERT-class transformers. Even accounting for the fact of contextual attention making conceptual embeddings more intrinsic to the conversation, a certain degree of distortion remains. For lexical and syntactic alignment, n-grams are used, as introduced by N. D. Duran, Paxton, and Fusaroli (2019). Yet, these structures remain unable to capture more intricate patterns of lexical and syntactic repetition. Subsequent methodical research could address these problems to increase the value of the framework proposed here.

6 Conclusion

6.1 Summary

With the regular use of chatbots, the perception that some artificial communication is more natural than others has been prevalent. Yet, despite a clear push in the development of Large Language Models (LLMs) towards more human-likeness, there are few approaches to systematically *evaluate* such a quality. To enable such examinations, this thesis constructed a theoretical basis of communicative accommodation/alignment which can be applied to non-humans as well.

There are two central **RQs** put forth and evaluated throughout this work: The question of (**RQ$_1$**) LLMs' ability to maintain *realized* accommodation and the question of the design of (**RQ$_2$**) empirical assessment of *effective* accommodation in LLMs.

What serves as a starting point is the notion of artificial interactivity, as developed by Mehler (2010), employing the semiotics of Peirce (2011). While this concept puts forth requirements for AI-communication to constitute interactivity, its more concrete implementation is left unexplicated. This thesis uses the notion of an dyadic actor network Latour (cf. 2005) to expand the taxonomy of interactivity in a more general direction. To make the conditions of interactivity more grounded in social-psychological theory and to enable concrete inquiry and testing, two prominent theories are employed: Interactive Alignment Theory (cf. Pickering and Garrod 2004) and Communication Accommodation Theory (cf. Giles, N. Coupland, and J. Coupland 1991). While the priming-mechanism-oriented IAT and the social-regulation-oriented CAT might seem irreconcilable, a common framework is able to be created as both theories recognize their scope of description. This framework proposes that there is an automatic component making aligning easier but which can be overwritten if there is reason to instead diverge (or use maintenance).

Along this framework, the structural capacity of transformer models to support accommodation is pondered, addressing **RQ$_1$**. Given the extended framework of accommodation, incorporating IAT and CAT in keeping with the model of HH communication, it can be argued that current transformers lack the structural capacity to *realize* convergence and the affective intentionality to *simulate* divergence or maintenance. While the semantic space of transformer models is extrinsic, not allowing for any global alignment of concepts, this may be "compensated" by the contextual embeddings of the attention mechanism. It is thus expected to see LLM-based chatbots be able to accommodate, albeit in a technically limited scope. Minimal cognitive criteria (cf. **A 1-4**) enabling 'accommodatability' as well as hypotheses (cf. **H 1-4**) for research on the matter are proposed.

Lastly, addressing **RQ$_2$**, experimental measurement of such "effective" (as opposed to "realized") accommodation is undertaken in order to be able to test hypotheses such as the aforementioned. For this, the IAT-based ALIGN (cf. N. D. Duran, Paxton, and Fusaroli 2019) as well as the CAT-based QuantCAT (cf. Rosen 2023) algorithms for accommodation quan-

tification are introduced, critiqued and used as the basis of a proposed experimental set-up to compare human-to-human with LLM-to-human accommodatability. This novel empirical scheme is able to generate comparable data via a controlled experiment involving human and LLM subjects, enforcing text-only communication and *ontological blindness*. It thus allows to measure accommodation as a prominent facet of artificial interactivity in direct relation to human interaction, the phenomenon with bears the basis of interactivity.

6.2 Ethical Considerations

Large Language Models and other consumer-oriented AI technologies are at a critically malleable point in history. They still possess what Bijker (1995, p. 27) refers to as "interpretative flexibility" and can thus be formed by people as a socially constructed technology. Because of this, the question of accommodation in artificial chat interlocutors is pivotal, as this thesis aims to not only deliver an *assessment* of the current state of transformer-based LLMs. It also seeks to provide a basis for forming in which ways they are going to progress[1].

This work only introduces a measurement scheme but does not engage in empirical examination itself. Nevertheless, there are potential ethical consequences to be considered regarding accommodative AI-chatbots.

There are several positive and negative use cases being aided by chatbots being able to effectively, globally accommodate. Negative uses could include emotional manipulation tactics (e.g. in social engineering attacks) while positives might be better education technology systems. Yet, there are less individualistic effects to be considered which might arise from the inclusion of a mechanism of intrinsic semantic space, enabling inductive learning. After all, the attempt at controlling toxic behavior in LLM development is greatly aided by the fixed, "immutable mobile" nature of LLMs.

On the other hand, it is this fixedness that allows for a subtle aggregate effect, changing society and language asymmetrically. The norms and *majority truth* of a given society, through (Latourian) translation, are materialized as data which in turn are used to train LLMs. In what is often referred to as *Algorithmic Bias* (Favaretto, De Clercq, and Elger 2019), these specific societal conceptions are thus inscribed into the underlying transformer's semantic space.

In the instance of non-accommodating chatbots engaging in conversation with humans throughout society, the respective dyadic actor-network contains one accommodating actant and one immutable mobile actant. As has been shown (cf. Dippold 2023, p. 29), humans tend to accommodate and use repair strategies towards non-aligning artificial entities. Thus, the predisposition d_H of the human actant adapts towards the semantic space of the LLM. However, in case of a non-accommodative chatbot, its predisposition d_L remains static. In aggregate, this could result in a recursive enforcement of societal semantic conceptions, cementing the truths of groups in power while attenuating those of minorities[2]. The view of a seemingly isolated dyad can thus be expanded to connect its asymmetry of communication to

[1] This could be described by the concept of "democratic interventions" in the social construction of technology, as introduced by Feenberg (2010, p. 58).

[2] There is similar effect of one-sided accommodation on a *vocal* level. Specifically, consider accent-dependent disparities in voice recognition quality, affecting minority groups (cf. Mengesha et al. 2021).

institutionalized power on a societal level. Like any artifact, LLMs make non-negotiable what previously was negotiable, rendering societal relations more rigid. Considering linguistics, since language and meaning are closely intertwined, individual accommodation processes may coalesce into the development and shift of languages (cf. Pleyer 2023; M. Stamenov, Gallese, and M. I. Stamenov 2002). Thus, the resulting societal-language shift would tend to asymmetrically flow into the direction of semantics which already hold political hegemony.

In conclusion, the prolonged, widespread use of non-accommodative AI-chatbots could effectively augment a societal shift towards reinforcing and technologically inscribing inequalities. Meanwhile, the use of semantically extrinsic chatbots, as an immutable mobile, allows for better control of toxic behavior. It thus should be stressed that for either possible result of examining chatbot-accommodatability, there are a plethora of negative and positive effects to be considered.

6.3 Research Outlook

This thesis hopes to have delivered a "complete package" to assess the accommodation of any given chatbot system whose instances can convert plain-text inputs into outputs in a dialogical manner. Especially given a renewed societal and scientific interest in the assessment of human-likeness in AI communication, this systematically derived framework provides a more sociologically and psychologically grounded approach.

The discussion, adjustment, and implementation of these suggestions bears the potential to enable more research projects in the promising field of artificial interactivity. Applications could include automatic detection of AI-generated dialogue, creating benchmarks for accommodation and providing possible trajectories of further development of language models.

Bibliography

An Inquiry into the Modes of Existence (2013). National Foundation of Political Science. URL: http://modesofexistence.org/ (cit. on p. 9).

Babel, Molly (Jan. 2012). "Evidence for Phonetic and Social Selectivity in Spontaneous Phonetic Imitation". In: *Journal of Phonetics* 40.1, pp. 177–189. ISSN: 00954470. DOI: 10.1016/j.wocn.2011.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0095447011000763 (visited on 10/07/2023) (cit. on pp. 12, 16).

Bal, Malyaban and Abhronil Sengupta (Aug. 21, 2023). *SpikingBERT: Distilling BERT to Train Spiking Language Models Using Implicit Differentiation*. arXiv: 2308.10873 [cs]. URL: http://arxiv.org/abs/2308.10873 (visited on 10/18/2023). preprint (cit. on p. 25).

Barad, Karen (1998). "Getting Real: Technoscientific Practices and the Materialization of Reality". In: *differences: A Journal of Feminist Cultural Studies* 10.2, pp. 87+. ISSN: 10407391. URL: https://link.gale.com/apps/doc/A54772266/AONE?u=anon~6d9e2d85&sid=googleScholar&xid=871ca6bb (visited on 11/30/2023) (cit. on p. 8).

Barwise, Jon and John Perry (1983). *Situations and Attitudes*. Cambridge, Mass: MIT Press. 352 pp. ISBN: 978-0-262-02189-0 (cit. on p. 7).

Biancardi, Beatrice, Soumia Dermouche, and Catherine Pelachaud (Aug. 12, 2021). "Adaptation Mechanisms in Human–Agent Interaction: Effects on User's Impressions and Engagement". In: *Frontiers in Computer Science* 3, p. 696682. ISSN: 2624-9898. DOI: 10.3389/fcomp.2021.696682. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2021.696682/full (visited on 10/07/2023) (cit. on pp. 3, 4).

Bijker, Wiebe E. (1995). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge MA: The MIT Press (cit. on p. 41).

Bolotta, Samuele and Guillaume Dumas (May 6, 2022). "Social Neuro AI: Social Interaction as the "Dark Matter" of AI". In: *Frontiers in Computer Science* 4, p. 846440. ISSN: 2624-9898. DOI: 10.3389/fcomp.2022.846440. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2022.846440/full (visited on 10/07/2023) (cit. on pp. 21, 25).

Borenstein, Elhanan and Eytan Ruppin (Sept. 2005). "The Evolution of Imitation and Mirror Neurons in Adaptive Agents". In: *Cognitive Systems Research* 6.3, pp. 229–242. ISSN: 13890417. DOI: 10.1016/j.cogsys.2004.11.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389041704000750 (visited on 10/18/2023) (cit. on p. 25).

Borzenko, Alexander (Dec. 2010). "A Neural Mechanism for Human Language Processing". In: *Neurocomputing* 74.1-3, pp. 104–112. ISSN: 09252312. DOI: 10.1016/j.neucom.2009.11.051. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925231210002675 (visited on 11/15/2023) (cit. on p. 25).

Braun-Thürmann, Holger (2002). *Künstliche Interaktion: wie Technik zur Teilnehmerin sozialer Wirklichkeit wird*. 1. Aufl. Studien zur Sozialwissenschaft. Wiesbaden: Westdt. Verl. 205 pp. ISBN: 978-3-531-13849-7 (cit. on pp. 3, 6–8).

Burgoon, Judee K., Lesa A. Stern, and Leesa Dillman (1995). *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge ; New York: Cambridge University Press. 334 pp. ISBN: 978-0-521-45120-8 (cit. on pp. 11, 12).

Byrne, Donn Erwin (1971). "The Attraction Paradigm". In: *(No Title)* (cit. on p. 15).

Callon, Michel (1984). "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay". In: *The Sociological Review* 32.1_suppl, pp. 196–233. DOI: 10.1111/j.1467-954X.1984.tb00113.x. eprint: https://doi.org/10.1111/j.1467-954X.1984.tb00113.x. URL: https://doi.org/10.1111/j.1467-954X.1984.tb00113.x (cit. on p. 8).

– (1986). "The Sociology of an Actor-Network: The Case of the Electric Vehicle". In: URL: https://api.semanticscholar.org/CorpusID:140482468 (cit. on p. 8).

Chi, Nguyen Thi Khanh and Nam Hoang Vu (Mar. 2023). "Investigating the Customer Trust in Artificial Intelligence: The Role of Anthropomorphism, Empathy Response, and Interaction". In: *CAAI Transactions on Intelligence Technology* 8.1, pp. 260–273. ISSN: 2468-2322, 2468-2322. DOI: 10.1049/cit2.12133. URL: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12133 (visited on 10/07/2023) (cit. on pp. 2, 3).

Clark, H.H. and S.E. Brennan (1991). "Grounding in communication". In: *Perspectives on socially shared cognition* 13.1991, pp. 127–149 (cit. on p. 11).

Cohen Priva, Uriel and Chelsea Sanker (Sept. 11, 2019). "Limitations of Difference-in-Difference for Measuring Convergence". In: *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10.1, p. 15. ISSN: 1868-6354, 1868-6354. DOI: 10.5334/labphon.200. URL: https://www.journal-labphon.org/article/10.5334/labphon.200/ (visited on 10/07/2023) (cit. on pp. 28, 29, 33, 38).

Daft, Richard L. and Robert H. Lengel (1986). "Organizational Information Requirements, Media Richness and Structural Design". In: *Management Science* 32.5, pp. 554–571. ISSN: 00251909, 15265501. URL: http://www.jstor.org/stable/2631846 (visited on 03/19/2024) (cit. on p. 3).

Danescu-Niculescu-Mizil, Cristian, Michael Gamon, and Susan Dumais (Mar. 28, 2011). "Mark My Words! Linguistic Style Accommodation in Social Media". In: *Proceedings of the 20th International Conference on World Wide Web*, pp. 745–754. DOI: 10.1145/1963405.1963509. arXiv: 1105.0673 [cs]. URL: http://arxiv.org/abs/1105.0673 (visited on 10/17/2023) (cit. on pp. 16, 28–31).

Deng, Yang et al. (Oct. 14, 2023). *Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration*. arXiv: 2305.13626 [cs]. URL: http://arxiv.org/abs/2305.13626 (visited on 11/20/2023). preprint (cit. on pp. 25, 26).

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (May 24, 2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs]. URL: http://arxiv.org/abs/1810.04805 (visited on 11/29/2023). preprint (cit. on pp. 23, 30).

Dippold, Doris (Jan. 2023). ""Can I Have the Scan on Tuesday?" User Repair in Interaction with a Task-Oriented Chatbot and the Question of Communication Skills for AI". In: *Journal of Pragmatics* 204, pp. 21–32. ISSN: 03782166. DOI: 10.1016/j.pragma.2022.12.004.

URL: https://linkinghub.elsevier.com/retrieve/pii/S0378216622002910 (visited on 10/07/2023) (cit. on pp. 2, 41).
Dix, Alan (Oct. 2017). "Human–Computer Interaction, Foundations and New Paradigms". In: *Journal of Visual Languages & Computing* 42, pp. 122–134. ISSN: 1045926X. DOI: 10.1016/j.jvlc.2016.04.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1045926X16300088 (visited on 10/12/2023) (cit. on p. 3).
Doherty, Kevin and Gavin Doherty (Sept. 30, 2019). "Engagement in HCI: Conception, Theory and Measurement". In: *ACM Computing Surveys* 51.5, pp. 1–39. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3234149. URL: https://dl.acm.org/doi/10.1145/3234149 (visited on 10/07/2023) (cit. on p. 3).
Duarte, Fabio (2024). *Number of CHATGPT users (Apr 2024)*. URL: https://explodingtopics.com/blog/chatgpt-users#how-many (cit. on p. 1).
Duignan, Brian (2023). *"semiotics"*. In: *Encyclopaedia Britannica* (cit. on p. 6).
Duran, Nicholas D., Alexandra Paxton, and Riccardo Fusaroli (Aug. 2019). "ALIGN: Analyzing Linguistic Interactions with Generalizable techNiques—A Python Library." In: *Psychological Methods* 24.4, pp. 419–438. ISSN: 1939-1463, 1082-989X. DOI: 10.1037/met0000206. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000206 (visited on 10/09/2023) (cit. on pp. 14, 29–35, 38–40).
Faught, Andrew (2023). *Nvidia CEO Jensen Huang on Inventing New Markets | Haas News | Berkeley Haas — Newsroom.Haas.Berkeley.Edu*. URL: https://newsroom.haas.berkeley.edu/nvidia-ceo-jensen-huang-on-inventing-new-markets/ (cit. on p. 1).
Favaretto, Maddalena, Eva De Clercq, and Bernice Simone Elger (Dec. 2019). "Big Data and discrimination: perils, promises and solutions. A systematic review". In: *Journal of Big Data* 6.1, p. 12. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0177-4. URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0177-4 (visited on 07/01/2022) (cit. on p. 41).
Feenberg, Andrew (2010). *Between reason and experience: Essays in technology and modernity*. Ed. by Wiebe E. Bijker. The MIT Press (cit. on p. 41).
Figueiredo, José (2010). "How to Recognize an Immutable Mobile When You Find One". In: (cit. on p. 10).
Fusaroli, Riccardo, Joanna Rączaszek-Leonardi, and Kristian Tylén (Jan. 2014). "Dialog as Interpersonal Synergy". In: *New Ideas in Psychology* 32, pp. 147–157. ISSN: 0732118X. DOI: 10.1016/j.newideapsych.2013.03.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0732118X13000342 (visited on 10/10/2023) (cit. on pp. 12, 28).
Gallese, Vittorio (Sept. 2008). "Mirror Neurons and the Social Nature of Language: The Neural Exploitation Hypothesis". In: *Social Neuroscience* 3.3-4, pp. 317–333. ISSN: 1747-0919, 1747-0927. DOI: 10.1080/17470910701563608. URL: http://www.tandfonline.com/doi/abs/10.1080/17470910701563608 (visited on 10/09/2023) (cit. on p. 21).
Gandolfi, Greta, Martin J. Pickering, and Simon Garrod (Feb. 13, 2023). "Mechanisms of Alignment: Shared Control, Social Cognition and Metacognition". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 378.1870, p. 20210362. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2021.0362. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0362 (visited on 10/07/2023) (cit. on pp. 2, 10, 12–14).

Giles, Howard (2016). *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts*. Cambridge, United Kingdom: Cambridge University press. ISBN: 978-1-107-10582-9 (cit. on pp. 15, 16).

Giles, Howard, Nikolas Coupland, and Justine Coupland (Sept. 27, 1991). "Accommodation Theory: Communication, Context, and Consequence". In: *Contexts of Accommodation*. Ed. by Howard Giles, Justine Coupland, and Nikolas Coupland. 1st ed. Cambridge University Press, pp. 1–68. ISBN: 978-0-511-66367-3. DOI: 10.1017/CBO9780511663673.001. URL: https://www.cambridge.org/core/product/identifier/CBO9780511663673A008/type/book_part (visited on 10/07/2023) (cit. on pp. 14–17, 40).

Giles, Howard, America L. Edwards, and Joseph B. Walther (Sept. 2023). "Communication Accommodation Theory: Past Accomplishments, Current Trends, and Future Prospects". In: *Language Sciences* 99, p. 101571. ISSN: 03880001. DOI: 10.1016/j.langsci.2023.101571. URL: https://linkinghub.elsevier.com/retrieve/pii/S0388000123000360 (visited on 10/07/2023) (cit. on pp. 15, 16).

Giles, Howard and Jessica Gasiorek (2013). "Parameters of Non-Accommodation: Refining and Elaborating". In: (cit. on pp. 14–16, 19).

Giles, Howard, Klaus R. Scherer, and Donald M. Taylor (1979). "Speech markers in social interaction". In: URL: https://api.semanticscholar.org/CorpusID:158998601 (cit. on p. 15).

Giles, Howard, Donald M. Taylor, and Richard Bourhis (1973). "Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data". In: *Language in Society* 2.2, pp. 177–192. DOI: 10.1017/S0047404500000701 (cit. on pp. 14, 16).

Goebel, Edited R, J Siekmann, and W Wahlster (n.d.). "Lecture Notes in Artificial Intelligence". In: () (cit. on p. 26).

Greimas, A. J. and Francois Rastier (1968). "The Interaction of Semiotic Constraints". In: *Yale French Studies* 41, p. 86. ISSN: 00440078. DOI: 10.2307/2929667. JSTOR: 2929667. URL: https://www.jstor.org/stable/2929667?origin=crossref (visited on 10/07/2023) (cit. on p. 8).

Guo, Biyang et al. (Jan. 18, 2023). *How Close Is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection*. arXiv: 2301.07597 [cs]. URL: http://arxiv.org/abs/2301.07597 (visited on 10/07/2023). preprint (cit. on p. 2).

Hildt, Elisabeth (July 5, 2021). "What Sort of Robots Do We Want to Interact With? Reflecting on the Human Side of Human-Artificial Intelligence Interaction". In: *Frontiers in Computer Science* 3, p. 671012. ISSN: 2624-9898. DOI: 10.3389/fcomp.2021.671012. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2021.671012/full (visited on 10/12/2023) (cit. on pp. 2, 3, 31).

Hochreiter, Sepp and Jürgen Schmidhuber (Dec. 1997). "Long Short-Term Memory". In: *Neural computation* 9, pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735 (cit. on p. 23).

Holohan, Michael and Amelia Fiske (Sept. 27, 2021). ""Like I'm Talking to a Real Person": Exploring the Meaning of Transference for the Use and Design of AI-Based Applications in Psychotherapy". In: *Frontiers in Psychology* 12, p. 720476. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2021.720476. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.720476/full (visited on 10/12/2023) (cit. on p. 8).

Hunter, Philip (Dec. 3, 2020). "Understanding the Human Language Processor: Neurobiology's Insights into the Human Brain's Semantic Processing of Information Could Inform

Education and Therapies for Language Disorders". In: *EMBO reports* 21.12, e52028. ISSN: 1469-221X, 1469-3178. DOI: 10.15252/embr.202052028. URL: https://www.embopress.org/doi/10.15252/embr.202052028 (visited on 11/15/2023) (cit. on p. 16).

Janicka, Iwona (June 3, 2023). "Processes of Translation: Bruno Latour's Heterodox Semiotics". In: *Textual Practice* 37.6, pp. 847–866. ISSN: 0950-236X, 1470-1308. DOI: 10.1080/0950236X.2022.2056765. URL: https://www.tandfonline.com/doi/full/10.1080/0950236X.2022.2056765 (visited on 10/08/2023) (cit. on p. 9).

Jiang, Fan and Shelia Kennison (Feb. 2022). "The Impact of L2 English Learners' Belief about an Interlocutor's English Proficiency on L2 Phonetic Accommodation". In: *Journal of Psycholinguistic Research* 51.1, pp. 217–234. ISSN: 0090-6905, 1573-6555. DOI: 10.1007/s10936-021-09835-7. URL: https://link.springer.com/10.1007/s10936-021-09835-7 (visited on 10/07/2023) (cit. on p. 16).

Jin, S. Venus and Seounmi Youn (May 28, 2023). "Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction". In: *International Journal of Human–Computer Interaction* 39.9, pp. 1874–1886. ISSN: 1044-7318, 1532-7590. DOI: 10.1080/10447318.2022.2129277. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2129277 (visited on 10/07/2023) (cit. on pp. 2, 3, 6, 19).

Kopp, Stefan (June 2010). "Social Resonance and Embodied Coordination in Face-to-Face Conversation with Artificial Interlocutors". In: *Speech Communication* 52.6, pp. 587–597. ISSN: 01676393. DOI: 10.1016/j.specom.2010.02.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167639310000312 (visited on 11/23/2023) (cit. on p. 11).

Kosinski, Michal (Aug. 29, 2023). *Theory of Mind Might Have Spontaneously Emerged in Large Language Models*. arXiv: 2302.02083 [cs]. URL: http://arxiv.org/abs/2302.02083 (visited on 10/07/2023). preprint (cit. on p. 2).

Kwon, Harim (Sept. 2021). "A Non-Contrastive Cue in Spontaneous Imitation: Comparing Mono- and Bilingual Imitators". In: *Journal of Phonetics* 88, p. 101083. ISSN: 00954470. DOI: 10.1016/j.wocn.2021.101083. URL: https://linkinghub.elsevier.com/retrieve/pii/S0095447021000589 (visited on 11/25/2023) (cit. on p. 16).

Latour, Bruno (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass.: Harvard University Press (cit. on p. 8).

– (1988). *The pasteurization of France*. Cambridge, Mass: Harvard University Press. 273 pp. ISBN: 978-0-674-65760-1 (cit. on p. 9).

– (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, Mass: Harvard University Press. 324 pp. ISBN: 978-0-674-65336-8 (cit. on p. 10).

– (2005). "Reassembling the Social". In: (cit. on pp. 8, 40).

– (2011). *Summary of the AiME Project -An Inquiry into Modes of Existence*. URL: http://www.bruno-latour.fr/node/328.html (cit. on p. 9).

– (2018). *An Inquiry into Modes of Existence: An Anthropology of the Moderns*. Trans. by Catherine Porter. First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Harvard University Press. 486 pp. ISBN: 978-0-674-98402-8 (cit. on p. 9).

Lewandowski, Natalie, Antje Schweitzer, and Daniel Duran (Mar. 2014). "An Examplar-Based Hybrid Model of Phonetic Adaptation". In: (cit. on p. 11).

Li, Mengjun and Ayoung Suh (Dec. 2022). "Anthropomorphism in AI-enabled Technology: A Literature Review". In: *Electronic Markets* 32.4, pp. 2245–2275. ISSN: 1019-6781, 1422-8890. DOI: 10.1007/s12525-022-00591-7. URL: https://link.springer.com/10.1007/s12525-022-00591-7 (visited on 11/30/2023) (cit. on p. 2).

Likhomanenko, Tatiana, Gabriel Synnaeve, and Ronan Collobert (Sept. 15, 2019). "Who Needs Words? Lexicon-Free Speech Recognition". In: *Interspeech 2019*, pp. 3915–3919. DOI: 10.21437/Interspeech.2019-3107. arXiv: 1904.04479 [cs]. URL: http://arxiv.org/abs/1904.04479 (visited on 11/02/2023) (cit. on pp. 24, 33).

Lin, Bill Yuchen et al. (May 27, 2023). *SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks*. arXiv: 2305.17390 [cs]. URL: http://arxiv.org/abs/2305.17390 (visited on 10/08/2023). preprint (cit. on pp. 25, 26).

Liu, Fanjue (July 4, 2023). "Hanging Out with My Pandemic Pal: Contextualizing Motivations of Anthropomorphizing Voice Assistants during COVID-19". In: *Journal of Promotion Management* 29.5, pp. 676–704. ISSN: 1049-6491, 1540-7594. DOI: 10.1080/10496491.2022.2163031. URL: https://www.tandfonline.com/doi/full/10.1080/10496491.2022.2163031 (visited on 11/30/2023) (cit. on p. 2).

Liu, Wenda et al. (Sept. 2019). "Shared Neural Representations of Syntax during Online Dyadic Communication". In: *NeuroImage* 198, pp. 63–72. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2019.05.035. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811919304276 (visited on 10/19/2023) (cit. on p. 21).

Liu, Yang (2019). *Fine-tune BERT for Extractive Summarization*. arXiv: 1903.10318 [cs.CL] (cit. on p. 35).

Lücking, Andy and Alexander Mehler (2013). "On Three Notions of Grounding of Artificial Dialog Companions". In: URL: https://api.semanticscholar.org/CorpusID:51773603 (cit. on p. 24).

Magne, Laurent (Oct. 4, 2011). "On the Modern Cult of the Factish Gods". In: *Society and Business Review* 6.3, pp. 295–298. ISSN: 1746-5680. DOI: 10.1108/17465681111171037. URL: https://www.emerald.com/insight/content/doi/10.1108/17465681111171037/full/html (visited on 10/20/2023) (cit. on p. 9).

Mareček, David and Rudolf Rosa (2019). "From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions". In: *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. Florence, Italy: Association for Computational Linguistics, pp. 263–275. DOI: 10.18653/v1/W19-4827. URL: https://www.aclweb.org/anthology/W19-4827 (visited on 11/24/2023) (cit. on p. 23).

Mariani, Marcello M., Novin Hashemi, and Jochen Wirtz (June 2023). "Artificial Intelligence Empowered Conversational Agents: A Systematic Literature Review and Research Agenda". In: *Journal of Business Research* 161, p. 113838. ISSN: 01482963. DOI: 10.1016/j.jbusres.2023.113838. URL: https://linkinghub.elsevier.com/retrieve/pii/S0148296323001960 (visited on 10/07/2023) (cit. on p. 2).

Maturana, Humberto R. and Francisco J. Varela (1980). *Autopoiesis and Cognition: The Realization of the Living*. Vol. 42. Boston Studies in the Philosophy and History of Science. Dordrecht: Springer Netherlands. ISBN: 978-94-009-8947-4. DOI: 10.1007/978-94-009-8947-4.

URL: http://link.springer.com/10.1007/978-94-009-8947-4 (visited on 10/19/2023) (cit. on p. 10).

Mehler, Alexander (2010). "Artifizielle Interaktivität. Eine semiotische Betrachtung". In: *Medienwandel als Wandel von Interaktionsformen*. Ed. by Tilmann Sutter and Alexander Mehler. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 107–134. ISBN: 978-3-531-15642-2. DOI: 10.1007/978-3-531-92292-8_6. URL: http://link.springer.com/10.1007/978-3-531-92292-8_6 (visited on 10/07/2023) (cit. on pp. 1, 6–8, 10, 13, 19, 39, 40).

Menenti, Laura, Martin J. Pickering, and Simon Garrod (2012). "Toward a Neural Basis of Interactive Alignment in Conversation". In: *Frontiers in Human Neuroscience* 6. ISSN: 1662-5161. DOI: 10.3389/fnhum.2012.00185. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2012.00185/abstract (visited on 11/15/2023) (cit. on pp. 20, 21).

Mengesha, Zion, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman (2021). ""I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans". In: *Frontiers in Artificial Intelligence* 4. ISSN: 2624-8212. DOI: 10.3389/frai.2021.725911. URL: https://www.frontiersin.org/articles/10.3389/frai.2021.725911 (cit. on p. 41).

Mooij, J.J.A. (1975). "Tenor, vehicle, and reference". In: *Poetics* 4.2, pp. 257–272. ISSN: 0304-422X. DOI: https://doi.org/10.1016/0304-422X(75)90084-4. URL: https://www.sciencedirect.com/science/article/pii/0304422X75900844 (cit. on p. 13).

Nass, Clifford and Youngme Moon (Jan. 2000). "Machines and Mindlessness: Social Responses to Computers". In: *Journal of Social Issues* 56.1, pp. 81–103. ISSN: 0022-4537, 1540-4560. DOI: 10.1111/0022-4537.00153. URL: https://spssi.onlinelibrary.wiley.com/doi/10.1111/0022-4537.00153 (visited on 11/21/2023) (cit. on p. 2).

Naveed, Humza et al. (Nov. 2, 2023). *A Comprehensive Overview of Large Language Models*. arXiv: 2307.06435 [cs]. URL: http://arxiv.org/abs/2307.06435 (visited on 11/18/2023). preprint (cit. on pp. 4, 26).

Oben, Bert (2015). "Modelling Interactive Alignment: A Multimodal and Temporal Account". In: URL: https://api.semanticscholar.org/CorpusID:184659382 (cit. on pp. 11, 14, 21, 29).

Oehmen, Raoul (2005). "The Temporal Segmentation of Dialogue as a Basis for a Multivariate Analysis of Speech Convergence". In: (cit. on pp. 11, 16).

OpenAI (2022). URL: https://openai.com/blog/chatgpt (cit. on p. 1).

– (2023a). *Chatgpt can now see, hear, and speak*. URL: https://openai.com/blog/chatgpt-can-now-see-hear-and-speak (cit. on p. 1).

– (Mar. 27, 2023b). *GPT-4 Technical Report*. arXiv: 2303.08774 [cs]. URL: http://arxiv.org/abs/2303.08774 (visited on 10/07/2023). preprint (cit. on pp. 2, 23).

Ostrand, Rachel and Eleanor Chodroff (Sept. 2021). "It's Alignment All the Way down, but Not All the Way up: Speakers Align on Some Features but Not Others within a Dialogue". In: *Journal of Phonetics* 88, p. 101074. ISSN: 00954470. DOI: 10.1016/j.wocn.2021.101074. URL: https://linkinghub.elsevier.com/retrieve/pii/S0095447021000462 (visited on 11/20/2023) (cit. on pp. 14, 21, 36).

Pardo, Jennifer S. (Dec. 2012). "Reflections on Phonetic Convergence: Speech Perception Does Not Mirror Speech Production: Reflections on Phonetic Convergence". In: *Language and Linguistics Compass* 6.12, pp. 753–767. ISSN: 1749818X. DOI: 10.1002/lnc3.367. URL: https:

//onlinelibrary.wiley.com/doi/10.1002/lnc3.367 (visited on 10/10/2023) (cit. on pp. 12, 14).
Pardo, Jennifer S., Elisa Pellegrino, Volker Dellwo, and Bernd Möbius (Nov. 2022). "Special Issue: Vocal Accommodation in Speech Communication". In: *Journal of Phonetics* 95, p. 101196. ISSN: 00954470. DOI: 10.1016/j.wocn.2022.101196. URL: https://linkinghub.elsevier.com/retrieve/pii/S0095447022000717 (visited on 10/07/2023) (cit. on pp. 11, 17).
Pardo, Jennifer S., Adelya Urmanche, Sherilyn Wilman, and Jaclyn Wiener (Feb. 2017). "Phonetic Convergence across Multiple Measures and Model Talkers". In: *Attention, Perception, & Psychophysics* 79.2, pp. 637–659. ISSN: 1943-3921, 1943-393X. DOI: 10.3758/s13414-016-1226-0. URL: http://link.springer.com/10.3758/s13414-016-1226-0 (visited on 10/10/2023) (cit. on pp. 12, 14).
Pattee, Howard H. (1987). "Simulations, Realizations, and Theories of Life." In: DOI: 10.13140/2.1.1871.8723. URL: http://rgdoi.net/10.13140/2.1.1871.8723 (visited on 11/22/2023) (cit. on p. 19).
Paul, Justin, Akiko Ueno, and Charles Dennis (July 2023). "ChatGPT and Consumers: Benefits, Pitfalls and Future Research Agenda". In: *International Journal of Consumer Studies* 47.4, pp. 1213–1225. ISSN: 1470-6423, 1470-6431. DOI: 10.1111/ijcs.12928. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijcs.12928 (visited on 10/07/2023) (cit. on pp. 1, 3).
Peirce, Charles S. (2011). *Phänomen und Logik der Zeichen*. Ed. by Helmut Pape. [Nachdr.], 1. Aufl. Suhrkamp-Taschenbuch Wissenschaft 425. Frankfurt am Main: Suhrkamp. 182 pp. ISBN: 978-3-518-28025-6 (cit. on pp. 7, 8, 40).
Pfadenhauer, Michaela (Apr. 14, 2013). "On the Sociality of Social Robots A Sociology-of-Knowledge Perspective". In: 10.1, pp. 135–153 (cit. on pp. 6, 29, 32).
– (May 27, 2015). "The Contemporary Appeal of Artificial Companions: Social Robots as Vehicles to Cultural Worlds of Experience". In: *The Information Society* 31.3, pp. 284–293. ISSN: 0197-2243, 1087-6537. DOI: 10.1080/01972243.2015.1020213. URL: http://www.tandfonline.com/doi/full/10.1080/01972243.2015.1020213 (visited on 10/11/2023) (cit. on p. 1).
Pickering, Martin J. and Simon Garrod (Apr. 2004). "Toward a Mechanistic Psychology of Dialogue". In: *Behavioral and Brain Sciences* 27.02. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X04000056. URL: http://www.journals.cambridge.org/abstract_S0140525X04000056 (visited on 10/07/2023) (cit. on pp. 8, 10–13, 20, 21, 23, 40).
– (Aug. 2013). "An Integrated Theory of Language Production and Comprehension". In: *Behavioral and Brain Sciences* 36.4, pp. 329–347. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X12001495. URL: https://www.cambridge.org/core/product/identifier/S0140525X12001495/type/journal_article (visited on 11/25/2023) (cit. on pp. 10, 13, 20).
– (Jan. 7, 2021). *Understanding Dialogue: Language Use and Social Interaction*. 1st ed. Cambridge University Press. ISBN: 978-1-108-46193-1. DOI: 10.1017/9781108610728. URL: https://www.cambridge.org/core/product/identifier/9781108610728/type/book (visited on 10/08/2023) (cit. on pp. 10, 12, 13, 15, 17, 29).
Pleyer, Michael (Jan. 2023). "The Role of Interactional and Cognitive Mechanisms in the Evolution of (Proto)Language(s)". In: *Lingua* 282, p. 103458. ISSN: 00243841. DOI: 10.1016/

j.lingua.2022.103458. URL: https://linkinghub.elsevier.com/retrieve/pii/S0024384122002224 (visited on 10/07/2023) (cit. on p. 42).

Raible, Wolfgang (1981). "Von der Allgegenwart des Gegensinns (und einiger anderer Relationen)". In: *Zeitschrift für romanische Philologie (ZrP)* 97.1-2, pp. 1–40. ISSN: 0049-8661, 1865-9063. DOI: 10.1515/zrph.1981.97.1-2.1. URL: https://www.degruyter.com/document/doi/10.1515/zrph.1981.97.1-2.1/html (visited on 10/07/2023) (cit. on p. 10).

Rasenberg, Marlou, Asli Özyürek, and Mark Dingemanse (Nov. 2020). "Alignment in Multimodal Interaction: An Integrative Framework". In: *Cognitive Science* 44.11, e12911. ISSN: 0364-0213, 1551-6709. DOI: 10.1111/cogs.12911. URL: https://onlinelibrary.wiley.com/doi/10.1111/cogs.12911 (visited on 10/07/2023) (cit. on pp. 11, 32).

Rizzolatti, G., L. Fadiga, L. Fogassi, and V. Gallese (May 1999). "Resonance Behaviors and Mirror Neurons". In: *Archives Italiennes De Biologie* 137.2-3, pp. 85–100. ISSN: 0003-9829. pmid: 10349488 (cit. on pp. 11, 20).

Roche, Jennifer M., Rick Dale, and Gina M. Caucci (Jan. 2012). "Doubling up on Double Meanings: Pragmatic Alignment". In: *Language and Cognitive Processes* 27.1, pp. 1–24. ISSN: 0169-0965, 1464-0732. DOI: 10.1080/01690965.2010.509929. URL: http://www.tandfonline.com/doi/abs/10.1080/01690965.2010.509929 (visited on 10/17/2023) (cit. on p. 13).

Rosen, Zachary P (Jan. 2023). "A BERT's Eye View: A Big Data Framework for Assessing Language Convergence and Accommodation". In: *Journal of Language and Social Psychology* 42.1, pp. 60–81. ISSN: 0261-927X, 1552-6526. DOI: 10.1177/0261927X221095865. URL: http://journals.sagepub.com/doi/10.1177/0261927X221095865 (visited on 10/07/2023) (cit. on pp. 30, 31, 35, 38, 40).

Schilbach, Leonhard et al. (Aug. 2013). "Toward a Second-Person Neuroscience". In: *Behavioral and Brain Sciences* 36.4, pp. 393–414. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X12000660. URL: https://www.cambridge.org/core/product/identifier/S0140525X12000660/type/journal_article (visited on 10/18/2023) (cit. on p. 21).

Shanahan, Murray (Feb. 16, 2023). *Talking About Large Language Models*. arXiv: 2212.03551 [cs]. URL: http://arxiv.org/abs/2212.03551 (visited on 10/19/2023). preprint (cit. on p. 2).

Shapshak, Paul (Jan. 31, 2018). "Artificial Intelligence and Brain". In: *Bioinformation* 14.01, pp. 038–041. ISSN: 09738894, 09732063. DOI: 10.6026/97320630014038. URL: http://www.bioinformation.net/014/97320630014038.htm (visited on 10/18/2023) (cit. on p. 25).

Sheth, Amit, Hong Yung Yip, Arun Iyengar, and Paul Tepper (Mar. 1, 2019). "Cognitive Services and Intelligent Chatbots: Current Perspectives and Special Issue Introduction". In: *IEEE Internet Computing* 23.2, pp. 6–12. ISSN: 1089-7801, 1941-0131. DOI: 10.1109/MIC.2018.2889231. URL: https://ieeexplore.ieee.org/document/8700320/ (visited on 10/07/2023) (cit. on p. 3).

Sitikhu, Pinky, Kritish Pahi, Pujan Thapa, and Subarna Shakya (2019). "A Comparison of Semantic Similarity Methods for Maximum Human Interpretability". In: *2019 Artificial Intelligence for Transforming Business and Society (AITB)*. Vol. 1, pp. 1–4. DOI: 10.1109/AITB48515.2019.8947433 (cit. on p. 34).

Skjuve, Marita, Ida Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg (Feb. 2019). "Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in

human-chatbot interaction". In: *Human Technology* 15, pp. 30–54. DOI: 10.17011/ht/urn.201902201607 (cit. on p. 2).

Stamenov, Maksim, Vittorio Gallese, and Maxim I. Stamenov, eds. (2002). *Mirror Neurons and the Evolution of Brain and Language ; [Selected Contributions to the Symposium on "Mirror Neurons and the Evolution of Brain and Language" Held on July 5 - 8, 2000 in Delmenhorst, Germany]*. Advances in Consciousness Research 42. Amsterdam: John Benjamins Pub. 390 pp. ISBN: 978-1-58811-215-6 (cit. on pp. 25, 42).

Sutter, Tilmann and Alexander Mehler, eds. (2010). *Medienwandel als Wandel von Interaktionsformen*. 1. Aufl. Wiesbaden: VS, Verlag für Sozialwissenschaften. 289 pp. ISBN: 978-3-531-15642-2 (cit. on pp. 6, 10).

Švantner, Martin (2021). "Agency as Semiotic Fabrication: A Comparative Study of Latour's ANT". In: *The American Journal of Semiotics* 37.3, pp. 289–315. ISSN: 0277-7126. DOI: 10.5840/ajs202231579. URL: http://www.pdcnet.org/oom/service?url_ver=Z39.88-2004&rft_val_fmt=&rft.imuse_id=ajs_2021_0037_0003_0289_0315&svc_id=info:www.pdcnet.org/collection (visited on 10/07/2023) (cit. on pp. 6, 8).

Tamminen, Herman (Dec. 31, 2020). "Body Ground Red – Integrating Peirce, Kristeva and Greimas". In: *Sign Systems Studies* 48.2-4, pp. 368–391. ISSN: 1736-7409, 1406-4243. DOI: 10.12697/SSS.2020.48.2-4.09. URL: https://ojs.utlib.ee/index.php/sss/article/view/SSS.2020.48.2-4.09 (visited on 10/07/2023) (cit. on p. 8).

Thoppilan, Romal et al. (Feb. 10, 2022). *LaMDA: Language Models for Dialog Applications*. arXiv: 2201.08239 [cs]. URL: http://arxiv.org/abs/2201.08239 (visited on 11/18/2023). preprint (cit. on p. 2).

Tognoli, Emmanuelle, Guillaume Dumas, and J. A. Scott Kelso (Feb. 2018). "A Roadmap to Computational Social Neuroscience". In: *Cognitive Neurodynamics* 12.1, pp. 135–140. ISSN: 1871-4080, 1871-4099. DOI: 10.1007/s11571-017-9462-0. URL: http://link.springer.com/10.1007/s11571-017-9462-0 (visited on 10/27/2023) (cit. on p. 25).

Touvron, Hugo, Thibaut Lavril, et al. (Feb. 27, 2023). *LLaMA: Open and Efficient Foundation Language Models*. arXiv: 2302.13971 [cs]. URL: http://arxiv.org/abs/2302.13971 (visited on 10/07/2023). preprint (cit. on p. 2).

Touvron, Hugo, Louis Martin, et al. (July 19, 2023). *Llama 2: Open Foundation and Fine-Tuned Chat Models*. arXiv: 2307.09288 [cs]. URL: http://arxiv.org/abs/2307.09288 (visited on 10/07/2023). preprint (cit. on p. 2).

Van Wolde, E J (1987). "A SEMIOTIC ANALYTICAL MODEL Proceeding from Peirce's and Greimas' Semiotics". In: (cit. on p. 8).

Vaswani, Ashish et al. (2017). "Attention Is All You Need". Version 7. In: DOI: 10.48550/ARXIV.1706.03762. URL: https://arxiv.org/abs/1706.03762 (visited on 10/16/2023) (cit. on pp. 1, 23).

– (Aug. 1, 2023). *Attention Is All You Need*. arXiv: 1706.03762 [cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 10/07/2023). preprint (cit. on pp. 1, 4, 23).

Voelker, Aaron Russell (2015). "A Biologically Plausible Sum-Product Network for Language Modeling". In: (cit. on p. 25).

Von Bergmann, Kirsten, Kirsten Bergmann, Holly P. Branigan, and Stefan Kopp (May 18, 2015). "Exploring the Alignment Space – Lexical and Gestural Alignment with Real and

Virtual Humans". In: *Frontiers in ICT* 2.7, pp. 1–11. DOI: 10.3389/fict.2015.00007 (cit. on p. 11).

Wang, Weizhi et al. (June 12, 2023). *Augmenting Language Models with Long-Term Memory*. arXiv: 2306.07174 [cs]. URL: http://arxiv.org/abs/2306.07174 (visited on 10/07/2023). preprint (cit. on p. 26).

Wei, Jason et al. (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., pp. 24824–24837. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf (cit. on p. 24).

Weise, Andreas and Rivka Levitan (June 2018). "Looking for Structure in Lexical and Acoustic-Prosodic Entrainment Behaviors". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Ed. by Marilyn Walker, Heng Ji, and Amanda Stent. New Orleans, Louisiana: Association for Computational Linguistics, pp. 297–302. DOI: 10.18653/v1/N18-2048. URL: https://aclanthology.org/N18-2048 (cit. on p. 21).

Xu, Wei, Marvin J. Dainoff, Liezhong Ge, and Zaifeng Gao (Mar. 6, 2023). *Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI*. arXiv: 2105.05424 [cs]. URL: http://arxiv.org/abs/2105.05424 (visited on 10/12/2023). preprint (cit. on p. 3).

Zarzà, I. de, J. de Curtò, Gemma Roig, Pietro Manzoni, and Carlos T. Calafate (2023). "Emergent Cooperation and Strategy Adaptation in Multi-Agent Systems: An Extended Coevolutionary Theory with LLMs". In: *Electronics* 12.12. ISSN: 2079-9292. DOI: 10.3390/electronics12122722. URL: https://www.mdpi.com/2079-9292/12/12/2722 (cit. on p. 24).

Zwaan, Rolf A. and Gabriel A. Radvansky (1998). "Situation Models in Language Comprehension and Memory." In: *Psychological Bulletin* 123.2, pp. 162–185. ISSN: 1939-1455(Electronic), 0033-2909(Print). DOI: 10.1037/0033-2909.123.2.162 (cit. on p. 12).