



## OPEN ACCESS

## EDITED BY

Kostas Karpouzis,  
Panteion University, Greece

## REVIEWED BY

Caroline F. Keating,  
Colgate University, United States  
Giulio Jacucci,  
University of Helsinki, Finland  
Suresh Neethirajan,  
Dalhousie University, Canada

## \*CORRESPONDENCE

Björn W. Schuller

✉ schuller@ieee.org

Johanna Löchner

✉ Johanna.Loechner@med.uni-tuebingen.de

RECEIVED 31 December 2022

ACCEPTED 09 October 2023

PUBLISHED 02 November 2023

## CITATION

Schuller BW, Amiriparian S, Batliner A,  
Gebhard A, Gerczuk M, Karas V, Kathan A,  
Seizer L and Löchner J (2023) Computational  
charisma—A brick by brick blueprint for  
building charismatic artificial intelligence.  
*Front. Comput. Sci.* 5:1135201.  
doi: 10.3389/fcomp.2023.1135201

## COPYRIGHT

© 2023 Schuller, Amiriparian, Batliner, Gebhard,  
Gerczuk, Karas, Kathan, Seizer and Löchner.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Computational charisma—A brick by brick blueprint for building charismatic artificial intelligence

Björn W. Schuller<sup>1,2\*</sup>, Shahin Amiriparian<sup>1</sup>, Anton Batliner<sup>1</sup>,  
Alexander Gebhard<sup>1</sup>, Maurice Gerczuk<sup>1</sup>, Vincent Karas<sup>1</sup>,  
Alexander Kathan<sup>1</sup>, Lennart Seizer<sup>3</sup> and Johanna Löchner<sup>3\*</sup>

<sup>1</sup>EIHW – Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany, <sup>2</sup>GLAM – Group on Language, Audio, & Music, Imperial College London, London, United Kingdom, <sup>3</sup>Department of Child and Adolescent Psychiatry, German Center for Mental Health (DZPG), Eberhard-Karls University, Tübingen, Germany

Charisma is considered as one's ability to attract and potentially influence others. Clearly, there can be considerable interest from an artificial intelligence's (AI) perspective to provide it with such skill. Beyond, a plethora of use cases opens up for computational measurement of human charisma, such as for tutoring humans in the acquisition of charisma, mediating human-to-human conversation, or identifying charismatic individuals in big social data. While charisma is a subject of research in its own right, a number of models exist that base it on various "pillars," that is, dimensions, often following the idea that charisma is given if someone could and would help others. Examples of such pillars, therefore, include influence (could help) and affability (would help) in scientific studies, or power (could help), presence, and warmth (both would help) as a popular concept. Modeling high levels in these dimensions, i. e., high influence and high affability, or high power, presence, and warmth for charismatic AI of the future, e. g., for humanoid robots or virtual agents, seems accomplishable. Beyond, also automatic measurement appears quite feasible with the recent advances in the related fields of Affective Computing and Social Signal Processing. Here, we therefore present a brick by brick blueprint for building machines that can appear charismatic, but also analyse the charisma of others. We first approach the topic very broadly and discuss how the foundation of charisma is defined from a psychological perspective. Throughout the manuscript, the building blocks (bricks) then become more specific and provide concrete groundwork for capturing charisma through artificial intelligence (AI). Following the introduction of the concept of charisma, we switch to charisma in spoken language as an exemplary modality that is essential for human-human and human-computer conversations. The computational perspective then deals with the recognition and generation of charismatic behavior by AI. This includes an overview of the state of play in the field and the aforementioned blueprint. We then list exemplary use cases of computational charismatic skills. The building blocks of application domains and ethics conclude the article.

## KEYWORDS

charisma, AI, empathy, mimicry, affective computing, social signal processing

## 1. Introduction

Charisma—an irresistible force that, apart from beauty or rhetoric, captivates people; a miracle cure for professional success and an almost effortless rise to the top of power: A plethora of popular science literature, podcasts, and discussions rotate around this fascination, providing training to adopt a charismatic style—going along with the great promise of being successful and attractive to others. The fascination of the topic comes on the one hand from these promises; on the other hand, charisma remains a complex phenomenon. Similarly, intelligence and emotions are not subject to any ubiquitously valid formula, and thus this article may frustrate the reader with various approximations of a definition. Besides this great uptake, the topic of charismatic behavior also has a research tradition in sociology and psychology and is now increasingly trending in computation. This is promising, since computational charisma may be applied to numerous fields such as leadership training, mental health care, and education, and enhance outcomes in several ways: more efficient leadership, increased comfort in recipients, better teamwork, and reduction of reluctance and irritation. However, charismatic behavior is not bound to particular values and initially exists independently of an ideology. This also makes the appropriation of charisma a potential danger if it is misused for unethical purposes.

Charisma can be conveyed through all modalities—the verbal or nonverbal, vocal channel, the visual channel (facial gestures, hand gestures, body posture), and other biological signals—even through touch or smell. Yet, we want to argue that the verbal, vocal channel, i.e., speech, together with concomitant nonverbal, vocal signals, is the primary modality for conveying a “charismatic” message. A person can radiate charisma and be perceived as being most charismatic, by just standing there, or seen from a distance where the observer cannot hear whether they speak or what they say. As long as an intentional (pragmatic and semantic) message is not transmitted, this boils down to mere “being charismatic”—the same way as being (seen/perceived/assessed as) attractive/sexy, powerful, or miserable. The differences can easily be demonstrated when we see and hear a charismatic person giving a speech in a video clip, or only can hear the audio channel, or only can see the video channel. Thus, every modality can contribute to the impression of charisma, but only speech serves as the primary means for its functioning, i. e., its impact on the audience. In this contribution, we, therefore, concentrate on the single modality “speech,” i.e., the verbal, vocal channel, together with concomitant “non-verbal” signals in the non-verbal, vocal channel, as a necessary ingredient for integrating charisma into human-machine communication. This is sufficient as well when we aim at agents that communicate only via voice with the users. Note that in Section 3.3, we will shortly describe all the different channels and means that can be employed to indicate charisma.

In this article, we initially discuss charisma from a scientific, but also from a popular science perspective (functional aspects of charisma), to follow up with several layers of markers for charisma (formal aspects of charisma), and computational aspects of charisma. We finish with the motivation based on applications, and detail the ethics of computational charisma.

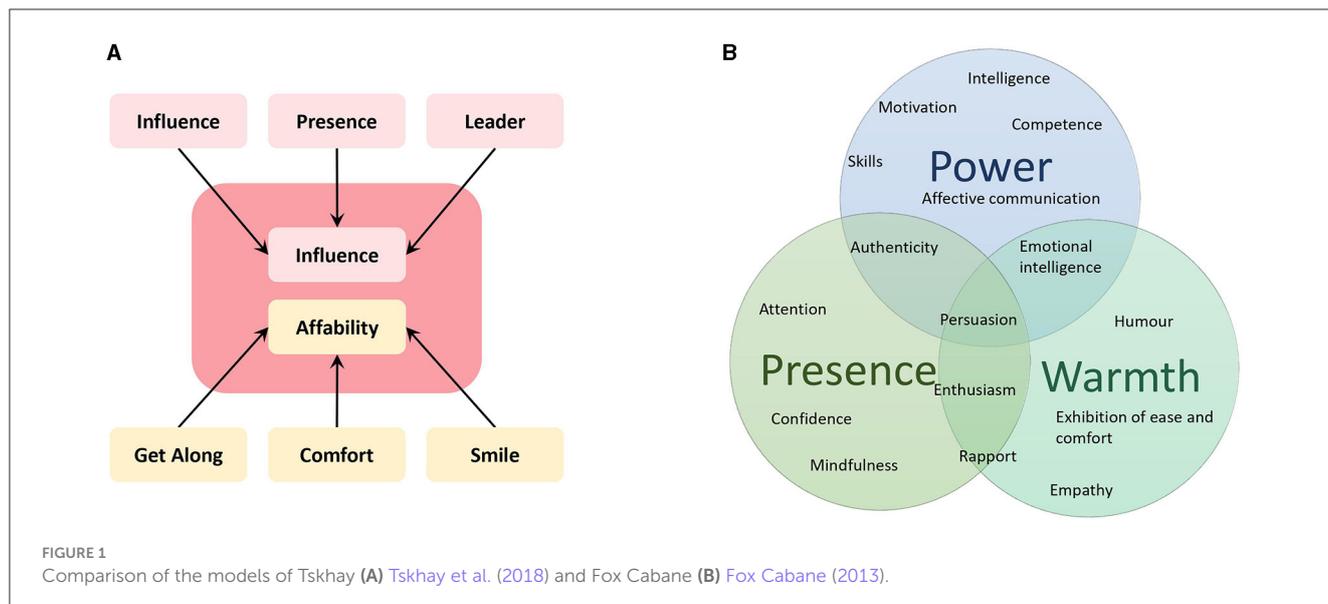
## 2. Functional aspects of charisma: psychological models

Charisma is a ubiquitous and frequently discussed phenomenon, and people seem to agree on which person displays this trait; yet, it is difficult to define it. In this first section, we discuss charisma from a sociological, psychological and also popular science perspective and aim to untidy the concrete characterisations of charisma. In contrast to other research fields, the subjective perception of charisma and popular science uptake of this phenomenon is particularly interesting, since it is part of its definition and hence, immanent.

### 2.1. Origin and definition

The word charisma originally comes from the Greek (*χάρισμα*) and means “gift of grace.” Even the ancient Greeks assumed that charisma is a gift from God that some have, and others do not. Today, the word is typically used as a descriptor for people who are attractive to others and manage to gather a following around them (for the good or bad) such as Princess Diana, Oprah Winfrey, Martin Luther King, or Adolf Hitler. Although everybody has an intuitive understanding of the concept of charisma and there is a high agreement in the population about which persons are charismatic, a scientifically sound and commonly used definition is an ongoing subject of research (Antonakis, 2012). This may partly be explained by the fact that the study of charisma is relatively young and still mostly restricted to economic psychology in terms of leadership research. Moreover, the construct charisma is complex and, as mentioned above, as difficult to grasp as phenomena such as intelligence and emotions, for which there is still no agreed upon conceptualization. However, the perception of charisma appears to be natural, coming across even on very limited databases. For instance, naturally, children are even less likely to have a defined concept of what charisma actually means; however, they are well capable of voting the “captain.” Hence, Antonakis and Dalgas asked 5–13 years-olds to rate the “captain” among a selection of pairwise displayed photographs of French candidates for the presidency, resulting in an 85 % hit rate (Antonakis and Dalgas, 2009). Although these results rely on visual data only, it was shown that attractiveness is not the key feature of charisma.

The sociologist Max Weber defined charisma at the beginning of the 20th century as an “extraordinary quality of a personality,” a “supernatural or superhuman power” (Weber, 1922). Based on this work, House (1976) provided the first operationalisation. Thereby, charisma was defined as the ability to inspire others toward a common goal and identity by appealing to their emotions and collective identity in order to impart an idealized vision to their followers, pointing out the central role of charisma in leadership research. In the following decades, more specific traits and behaviors have been associated with charisma (Antonakis, 2012). A key quality lies in the ability to connect with other people and exhibit ease, trust, and comfort in the audience paving the ground to become a leader. In addition, a charismatic person is



highly persuasive (Mhatre and Riggio, 2014). Several definitions of such phenomena are discussed in the literature including properties such as authenticity, emotional competence (e.g., understanding emotions in oneself and others, or managing own emotions), empathy, persuasiveness, spending attention to others, passion, enthusiasm, and obtaining strong opinions. However, such qualities may not only be used for charity aspects. Welsh and colleagues investigated the associations of psychopathy, charisma, and success. They found that psychopathy was positively associated with leadership charisma and the influence component of general charisma (Welsh and Lenzenweger, 2021); in addition, charisma moderated the association of psychopathic traits and perceived success in the form of evading detection and punishment.

To date, there is no single, overarching and generally accepted definition of charisma; researchers rather work with multiple different (although overlapping) operationalisations. Thereby, concepts of charisma vary for example between trait-oriented models as inherent personality characteristics (Burke and Brinkerhoff, 1981), state-oriented models in terms of observer perception and outcomes (Awamleh and Gardner, 1999), the transactional relationship between charismatic leaders and followers (Keating et al., 2020), or a combination of these (Conger et al., 2009). Interestingly, there are even neurological correlates of brain activity assessed by EEG supporting a more balanced activity in the right and left frontal hemispheres of subjects when observing charismatic vs. non-charismatic leaders (Keating et al., 2020). In the following, we will describe two of these charisma models (Fox Cabane, 2013; Tskhay et al., 2018) in detail along with associated psychological concepts that will lay the foundation for this article. Thereby, for the purpose of this article, we will favor models focusing on personality and behavior modes of charisma.

Aiming to provide a comprehensive model of charisma based on empirical data, Tskhay et al. (2018) created an empirical model of charisma: They investigated characteristics of charisma by rigorous and repeated questioning of people how they use to describe charismatic people, and subsequently applied factor

analyses to identify the most important components. Their analyses resulted in a two-factor model with one factor—*influence*—consisting of items that describe leadership ability and influence in a group setting, and another factor—*affability*—that consists of items describing a pleasant and inviting disposition toward others. The factors with more detailed descriptions and an exemplary list of behaviors that are associated with each are given in this section, with no claim to completeness. While *influence* and *affability* are separate qualities, in the combination of traits and behaviors associated with these two, charisma emerges somehow as a novel trait. Charisma is thereby defined as a multi-dimensional construct of traits and behaviors in contrast to “just” being a likable person. Similarly, Keating (2011) argued that dominant behavior triggers avoidance reactions in others, whereas emotionally warm behavior triggers approach reactions. She further claims that the perception of charisma emerges specifically through the simultaneous elicitation of avoidance and approach reactions by the combination of *influence* (dominance, power) and *affability* (emotionality, approachability) in a charismatic person.

There are several psychological constructs that may be convergent or discriminant to charisma. In a validation study of the influence-affability model, the uniqueness or relatedness of charisma to other individual difference measures was tested in multiple samples (Tskhay et al., 2018; see Table 1). Thereby, *influence* and *affability* were both found to be significantly related to emotional intelligence, i. e., the appraisal, expression, regulation, and utilization of emotions in a variety of contexts (Schutte et al., 1998). In terms of emotionality experienced by oneself, positive affect was positively related to the two charisma factors, while negative affect was negatively related to both. Political skill, defined by the four dimensions of social astuteness, interpersonal influence, networking ability, and apparent sincerity (Ferris et al., 2005), is often used as a metric of charismatic leadership; accordingly, it was found to be positively related to *influence* and *affability*. Intelligence is a trait that is often ascribed to charismatic individuals in lay theories; however, intelligence, as determined by Raven’s Matrices

TABLE 1 Detailed overview of the Tskhay et al. model.

	Model	GCI item	Associated constructs
Influence	Influence Presence Leader	Has the ability to influence people Has a presence in a room Knows how to lead a group	Convergent: emotional intelligence, positive affect, extraversion, openness, conscientiousness, political skill, competence Discriminant: negative affect, neuroticism
Affability	Get along Comfort Smile	Can get along with everyone Makes people feel comfortable Smiles at people often	Convergent: emotional intelligence, positive affect, confidence, extraversion, openness, conscientiousness, agreeableness, political skill, competence, warmth Discriminant: negative affect, neuroticism

(Raven and Court, 1998), was not associated with influence or affability, indicating that charisma may rely more on interpersonal skills in social interactions rather than intelligence. Further, the general confidence of an individual as the degree to which one feels certain about both the world and idiosyncratic surroundings and their ability to deal with stress (Keller et al., 2011) was positively related to affability, but not associated with influence. In terms of personality traits as the Big Five (McCrae et al., 1999), openness, conscientiousness, and extraversion were positively related to both influence and affability, while agreeableness was only related to affability. Neuroticism on the other hand was negatively associated with influence and affability. Of the dimensions competence and warmth—two essential elements of both social behavior and personal characteristics (Fiske et al., 2007)—influence was only related to competence, while affability was related to both warmth and competence.

Another model of charisma was proposed by Fox Cabane (2013), in which she refers to charisma as deriving from three pillars: *presence*, *power*, and *warmth*. *Presence* is displayed by dwelling in the current moment, active listening, and responding adequately. The focus of attention lies on the person one is talking to and taking an honest interest in the conversation partner. *Power* is not defined as actual power like being in a position as president or high-rank manager. It is rather understood as high competence due to certain skills, abilities, or intelligence a person obtains. *Warmth* requires a high level of empathy, openness, and positivity. The pillar *warmth* has frequently been studied as part of the two-dimensional *warmth* and *competence* (W&C) model (Fraser et al., 2021, 2022; Wang and Chanel, 2021), where *warmth* indicates the nature of the sender's intent toward the receiver, and *competence* the ability of the sender to enact this intent. The combination of these dimensions evokes emotional responses ranging from admiration to disgust (Fraser et al., 2022). Thus, *warmth* is closely related to the perceptions of attractiveness and empathy. Therefore, charismatic individuals usually radiate acceptance and friendliness that one otherwise experiences only from family members or friends. It is discussed whether one or two of the three qualities may be sufficient to appear charismatic, as Steve Jobs, for instance, scored with *presence* and *power*, but lacked *warmth*. In contrast, Martin Luther King showed all three qualities. Hence, the pillars *warmth* and *power* may relate to *affability* and to the *influence* of the two-factor model by Tskhay et al. (2018), while the pillar *presence* was discussed as a non-latent, i.e., secondary, variable in their empirically found model by a factor analysis following questioning participants (see Figure 1A).

The two models presented here aim to provide an operationalisation of charisma for this article. The first model proposed by Tskhay et al. (2018) is an empirical two-factor model with the factors of influence and affability. Influence consists of leadership ability and influence in a group setting, while affability consists of a pleasant and inviting disposition toward others. The second model proposed by Fox Cabane (2013) includes three pillars of charisma: presence, power, and warmth. Presence is the ability to dwell in the current moment and actively listen and respond adequately. Power refers to high competence due to certain skills, abilities, or intelligence, while warmth requires a high level of empathy, openness, and positivity. Overall, both models show a high level of overlap and suggest that charisma is a multi-dimensional construct of traits and behaviors that make people appear influential, likable, and captivating, and that the combination of different dimensions, such as influence, affability, presence, power, and warmth, can create a charismatic persona.

Very similarly, the concept of *rapport* is defined and may well serve as part of charismatic behavior. Tickle-Degnen and Rosenthal (1990) conceptualized the nature of *rapport* in terms of a dynamic structure of three interrelating components: mutual attentiveness, positivity, and coordination; these are differently weighted and present over time in a relationship. Hence, *rapport* is characterized by agreement, mutual understanding, or empathy establishing ease and comfort in communication partners. In consequence, a charismatic individual is capable of establishing *rapport*.

Conclusively, charisma is a person-specific descriptor that emerges specifically in social situations through the attribution of a certain set of traits to an individual. Despite heterogeneous conceptualization and the inherent complexity, there is a consensus that charismatic individuals exert influence over others, have extraordinary social skills, comfort and connect to others, inspire followership, and are prone to leadership roles (Antonakis, 2012; Tskhay et al., 2018). Breaking charisma down to such concrete properties reveals a combination of personality traits and skills that are partly inherited, socially acquired, and trained. In social psychology, the processes that lead us to form impressions about other people are referred to as person perception (Moskowitz and Gill, 2013). Some methods of perceiving another person involve inferring details about them based on observations of their activities. Other types of personal perception happen more immediately and only need one to view another person. In building a machine that people perceive as charismatic, a bias in human inference processes can be exploited, namely the fundamental attribution error: People tend to ascribe observed behaviors to internal factors like personality or character rather than to external

factors such as situational constraints (Colman, 1982). This is an error that has previously been shown to happen also in human-machine interactions (Horstmann and Krämer, 2022). Thus, by mimicking certain appearance cues, characteristics, and behaviors programmatically to elicit the perception of charisma-associated traits, it should be possible to build a “charismatic AI.”

## 2.2. Acquisition of charismatic behavior

In consequence, charismatic behavior can be acquired and there is a plethora of training programs offered especially in the field of leadership coaching. Overall, the two key qualities, i. e., factors, introduced by Tskhay et al. (2018) may be achieved especially through confidence and skills (*influence*), emotional intelligence, and empathy (*affability*). As is conclusive from the above (see also Figure 1), they may also be complemented by a third pillar or factor suggested by Fox Cabane (2013)—mindfulness (*presence*). Following Fox Cabane (2013) and Tskhay et al. (2018), such characteristics will elicit an increased impression of attractiveness, energy, persuasiveness, power, and empathy, and establish rapport between communication partners.

Focusing on leadership training, and translating charismatic behavior into more concrete tactics, Antonakis et al. (2012) investigated twelve techniques to increase charisma—the so-called “charismatic leadership tactics” (CLTs). Similar to athletes who follow a training schedule, leaders who aim to become influential, trustworthy and “leaderlike” are recommended to practice certain tactics regularly. For this purpose, they examined the nomination speeches of all candidates for president in the U. S. between 1916 and 2008. The analysis revealed that the use of figurative language, anecdotes, proverbs, and the proper use of body language had a significant impact on the outcome of the election. Despite humor, repetition, and talking about sacrifices, such verbal and non-verbal techniques were shown to have the greatest impact in any context. The nine verbal techniques are metaphors, similes, and analogies; stories and anecdotes; contrasts; rhetorical questions; expressions of moral conviction; reflections of the group’s sentiments; three-part lists; the setting of high goals; and conveying confidence that they can be achieved. For example, the metaphor of being on a boat in a storm may serve as a metaphor for a critical period. Even without being a born raconteur one can tell the compelling story of taking a deep breath as “anchor” and visualize the north star for guidance. Another example to motivate followers through a challenging period would be an anecdote of a personal story, as climbing a mountain when a thunderstorm arises and how the team must have kept going. In addition, there are three non-verbal techniques: animated voice, facial expressions, and gestures. Keeping with voice-associated techniques to improve oneself’s charisma, or rather the perception of charisma in others, it is suggested to speak clearly, fluently, forcefully, and in an engaging manner that invokes images, energy, and action; moreover, the delivery’s pace and intonation should be varied, with a general upbeat tempo and an occasionally slowing down to create tension (Tubbs, 2019). Similarly, Fox Cabane (2013) recommends lowering the tone of one’s voice at the end of each statement and make frequent pauses while speaking. Despite these strategies to develop

or improve charisma, the debate on whether charisma can be learnt or simply is a trait with between-subject variation is still ongoing (Tubbs, 2019). It is of note that, even in human generation of charisma, an attribution error can apply: When speakers learn to speak with a “charismatic voice,” people perceive them as charismatic, even when their personality does not change.

Very concrete acquired behavior was shown to lead to a more charismatic behavior of individuals; hence, it can be installed on who- or whatever to some extent: A person might not be charismatic in themselves but may appear this way, due to, e.g., speaking in a charismatic way. Note that it was shown that appearance is not the key factor in “charismatic appearance.” Antonakis et al. (2012) observed that in eight out of ten U.S. presidential races, candidates who deployed such verbal CLTs won more often. Since communication nowadays is primarily technology-mediated, Ernst et al. (2022) investigated CLTs in a recent prospective meta-analysis on virtual charismatic leadership, yielding large to moderate effect sizes. In summary, disentangling especially phonetic and linguistic markers for charisma may be particularly beneficial in times of virtual communication in all kind of fields.

In the following section, this article focuses on the specific phonetic, linguistic, and other markers that are associated with charisma.

## 3. Formal aspects of charisma: phonetic, linguistic, and other markers

So far, we dealt with the definition of charisma and with its function, i. e., what it can be used for. In this section, we now present studies that describe the formal aspects of charisma, i. e., which phonetic, linguistic, and other features can be employed as markers to describe charisma and to extract and classify pertinent information. As the focus is not on definition and theoretic modeling, in these studies, charisma is introduced in a rather pre-theoretical fashion. References to Weber can often be found, as well as straightforward operationalisation such as: “As a preliminary conjecture, we might say that if a person is felt to be charismatic, there should be something in his/her way of being or behaving that causes others to attribute him/her with certain internal properties, properties that we might call ‘charismatic traits’” (D’Errico et al., 2013).

In the realm of charisma, we are faced with a three-way confusion: (i) the very same or highly similar markers are named differently—for example, the mutual adaptation of communication partners has been called as well (phonetic) convergence, entrainment, or even “synchronization”; (ii) many different, partly overlapping traits can contribute to the impression of charisma—for instance, traits such as warmth, affability, and likability are all different aspects of something that we could call as well “empathetic adaptation”; (iii) the phenomenon charisma is defined differently, depending on, for instance, the roles impersonated by different or the same individuals in different scenarios—for example, political leader, therapist, or advertiser. This of course impacts the roles and importance of markers and by that, features as well.

The marking of charisma is definitely multi-modal, and trading relations exist both between and within modalities—i. e., a more pronounced, but not exaggerated marking in one parameter can compensate for weak signaling in another parameter, see Niebuhr et al. (2020). Ranking the importance of modalities is futile and either based on intuition or on one or only a few studies with their specific databases, designs, and methods; see the “7 %-myth” (Mehrabian and Wiener, 1967; Schuller and Batliner, 2014, p. 14). Note that our paralinguistic taxonomy does not fully overlap with another traditional one telling apart “verbal” from “nonverbal” communication which has been established in the 1960s, as a reaction to the emphasis on verbal communication, see Jones and LeBaron (2002). Besides vocal aspects, “nonverbal” includes everything else which is not verbal, i. e., facial gestures, body postures, gait, etc. We would not favor a concept of “nonverbal” that, for instance, includes “shorter turns at talk” (Burgoon et al., 1998): it does not separate the different recording channels “acoustics” and “video,” and it camouflages the role of linguistics: shorter sentences (“verbal”) simply result in shorter turns (“non-verbal”).

In this contribution, we concentrate on the paralinguistic aspects of charisma. This includes the two modalities of spoken and written language, i. e., [+verbal, +vocal]: speech; [−verbal, +vocal]: “nonverbals” such as affect bursts, hesitations; and [+verbal, −vocal]: written language; see Batliner et al. (2022). This appears reasonable, given the focus on today’s AI that often communicates with users by this modality; in addition, also when analyzing human interaction, spoken language plays a key role. Yet, we will as well present a sketchy overview of charisma conveyed within the other modalities. We will start with phonetic markers of charismatic speech in Section 3.1; then follow linguistic markers in Section 3.2, and markers in other modalities in Section 3.3.

The intuitive understanding of charisma is mirrored in equally intuitive characterisations such as attractive, inspiring, animated, enthusiastic, warm, likable, or pleasant. In this section, we now report the state of the art in mapping these terms onto markers and by that, features, that can be measured and counted. In Figure 2, we depict some more or less prominent “sub”-traits—relevance is indicated by font size—that constitute charisma; at the same time, they represent topics on their own that have been addressed in research papers; see Section 3.1. We can relate them to the pillars and factors depicted in Figure 1: presence, influence, and power can establish leadership and persuasiveness, with competence as a default prerequisite; warmth and affability presuppose empathy, and this makes a person likable and attractive. Traits such as sexual attractiveness might not constitute charisma but co-occur. The four marginal notes in Figure 2 indicate the intricate status of these and all other traits that characterize charisma. The semantics of related terms such as “leadership,” “influence,” or “persuasion” is complex: Leadership encompassed influence which is or is not exerted intentionally; persuasion is intentional *per se*; influence can be highly intentional (think of present day internet influencer who want to sell something) but need not to be.

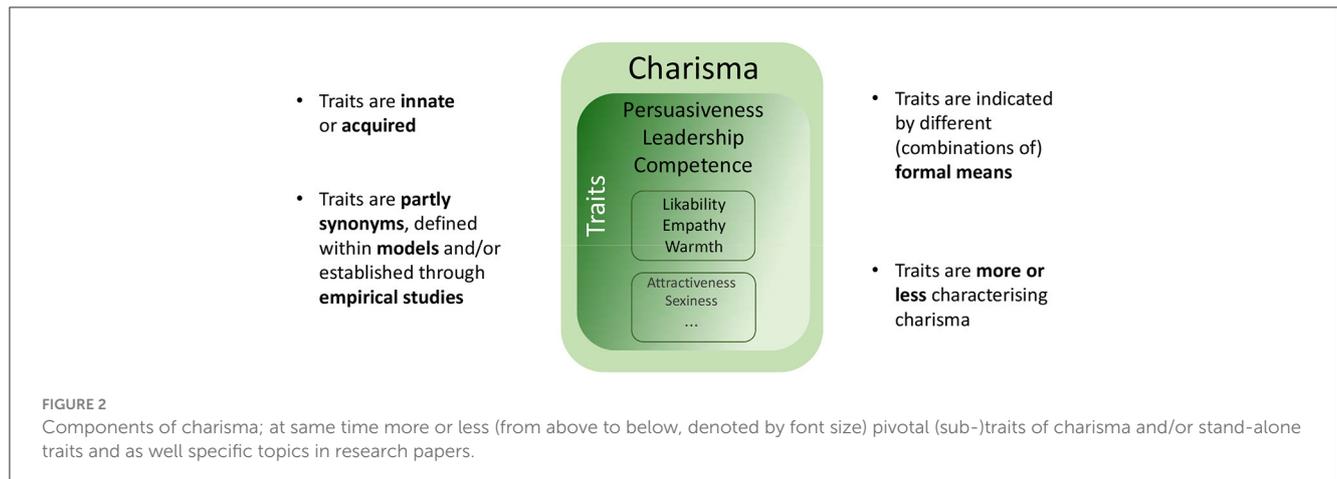
Here, we cannot aspire to completeness—there are many personality traits besides that might be seen as partly constituting the complex trait “charisma.” They have a different functional load for characterizing different charismatic roles; to mention two of the

most prototypical charismatic role models: Charismatic traits can look different, depending on whether we want to model a gifted CEO such as Steve Jobs or a political leader such as Barack Obama, vs. another type of political leaders such as Donald Trump (Lukes, 2019). And the very same political leaders change their roles and adjust the formal markers accordingly, depending on the situation, see Rosenberg and Hirschberg (2009): “Speakers were rated as more charismatic when they were delivering a stump speech (mean rating of 3.28) than when they are being interviewed (2.90).” In an interview, signaling competence in the first place might be more important than signaling charisma. Thus, charisma and by that, its characterizing formal features have to meet the expectations of the perceivers and to be adjusted to the role model in different types of interaction.

### 3.1. Phonetic markers

Arguably, Rosenberg and Hirschberg (2009) described the first sets of studies on charismatic speech. So far, most of them addressed charisma in politics (candidate speeches) and marketing (Zoghaib, 2019) and focus on prosody. Related are states and traits such as leadership (Weninger et al., 2012), competence/trustworthiness (Yang et al., 2020; Davidson, 2021), likeability (Weiss and Burkhardt, 2010; Schuller et al., 2015), and (sexual) attractiveness (Trouvain et al., 2021). Weiss et al. (2021) attribute warmth to likeability that can be used “as a synonym for social attractiveness.” Charisma can be tied to performing something, e. g., a candidate speech, and can be switched on and off; see Rosenberg and Hirschberg (2009). So, at least the “acoustics of charisma” are not in an “always one-to-one relationship” to personality. Of course, this makes it possible for it to be taught and acquired. An infamous example is Adolf Hitler, where the only recording of non-public speech<sup>1</sup> reveals a relaxed, almost likable style of speaking, much different from his rehearsed public speeches. We can distinguish “dark charisma” (Fragouli, 2018), where, e. g., anger can be strongly marked with prosodic means, when this is in accordance with the audience, from “bright charisma,” which can be rather marked prosodically (Barack Obama), or linguistically and by the context (Mahatma Gandhi); see D’Errico et al. (2013). Even psychopaths can display traits of bright charisma in discordance to their personality (Ga, 2016). Intervening factors can be gender, age, and culture (D’Errico et al., 2013). Laryngealised, “creaky” voice—that is at the same time indicating very low, but also irregular pitch—can make men more cool and attractive (Davidson, 2021), whereas a breathy voice is preferred for women (Greer and Winters, 2015); this, however, mostly holds for younger women (Anderson et al., 2014), whereas in business and academia, a creaky voice can be a sign of competence for both females and men. Klofstad et al. (2016) summarize the experiment on leadership: “... males with lower-pitched voices tend to be perceived as more attractive, physically stronger, and more “dominant.”” For females, the standard is dichotomous: Women with higher-pitched voices

1 <https://www.youtube.com/watch?v=WE6mnPmztoQ>



tend to be considered more attractive, whereas those with lower-pitched voices are perceived as more dominant”; see as well (Anderson and Klofstad, 2012; Klofstad et al., 2015; Zoghaib, 2019). Niebuhr et al. (2016) compared customer and investor keynotes of Steve Jobs and Mark Zuckerberg. Jobs, commonly perceived as the more charismatic speaker, produced a higher pitch level (even approaching that of female speakers), and almost twice the pitch range of Zuckerberg. Jobs used shorter phrases, had fewer disfluencies, and scored higher in the voice quality metrics. However, he did not exceed Zuckerberg in terms of intensity variability. Both showed significant differences when addressing customers and investors, showing again that charisma is situation-dependent.

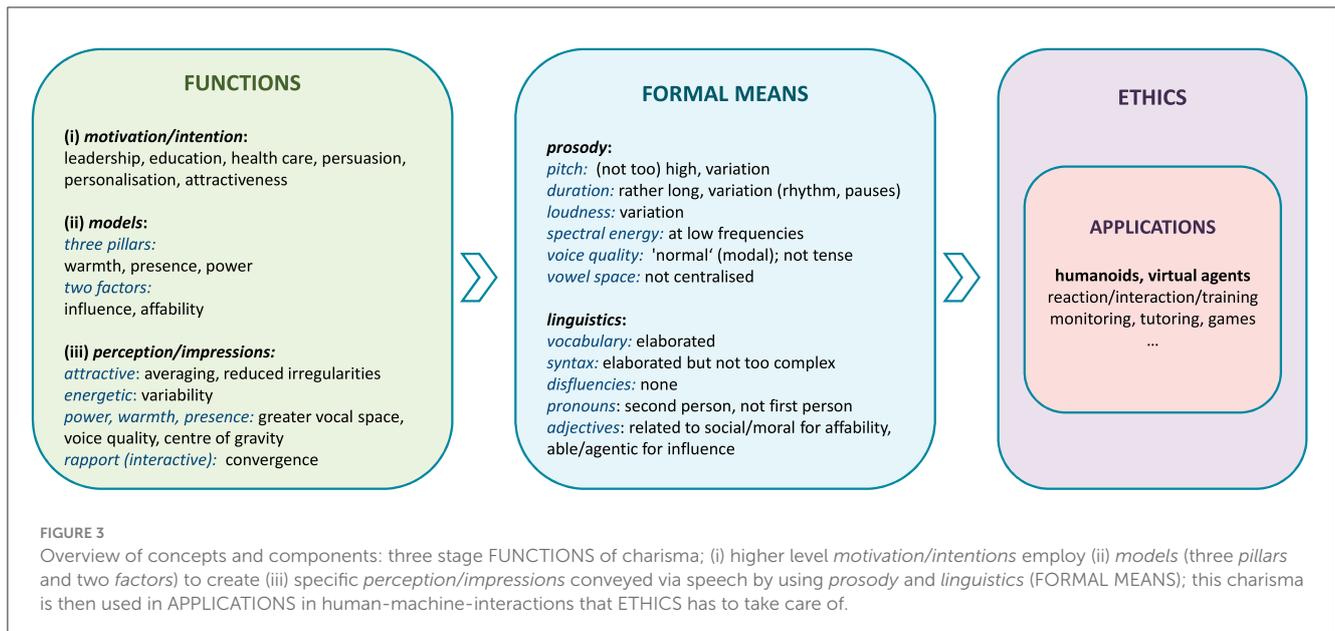
Low-level descriptors of the voice have been shown to convey perceptions of speaker personality traits (Schuller et al., 2015). The likeability of a person can be predicted using pitch (fundamental frequency  $F_0$ ), articulation rate, and spectral parameters such as MFCC (Weiss and Burkhardt, 2010). D’Errico et al. (2013) conducted a cross-cultural study showing the effects of pitch and the duration of speech pauses on the perception of two dimensions aggregated from 67 traits and conforming to proactivity-attractiveness and calm-benevolence. Pasquale et al. (2019) found a strong positive correlation between  $F_0$  standard deviation, i.e., variability, and client’s perception of therapist’s empathy in a therapeutic setting; see as well (Weiste and Peräkylä, 2014).

As far as prosody is concerned, we can sum up with Yang et al. (2020): “... voices that are louder, higher, faster, and with greater fluctuation in pitch were rated as more charismatic.” Now, we “only” have to define the acceptable range of these prosodic varieties; too great an intensification will certainly yield undesirable consequences such as the impression of distortion or a lower discriminability, see Hamilton and Stewart (1993) and Holz et al. (2021). Moreover, suitable ranges depend on which aspect of charisma is emphasized, whether it is enthusiasm and engagement via high-pitched, dynamic speech, or competence and dominance via lower-pitched voice; see again Rosenberg and Hirschberg (2009), Anderson et al.

(2014), and Davidson (2021). Other prosodic parameters as well, and other acoustic parameters such as spectral distributions, favorable for conveying charisma, can be described as “well-balanced” and “well-shaped”: neither too integrating nor too isolating prosodic phrasing—i.e., not too many, but not too few pauses; more spectral energy at low frequencies (“full voice”); and more precise articulation (no centralization of vowels). A charismatic voice is definitely not characterized by dysphonia, i.e., disordered voice (hypophonia, i.e., soft voice, or the opposite, hyperphonia, i.e., tense, harsh voice). Based on all these findings, Niebuhr et al. (2020) describe a system for “charisma profiling” by establishing an “Acoustic Voice Profiling”—a program for acquiring charismatic speech by an iterative training of relevant vocal parameters, followed by an assessment of students’ speeches.

Voice quality—such as modal, creaky, or breathy voice—contributes to perceived personality traits and by that, to perceived charisma of the speakers; see Voße et al. (2022) and Pearsell and Pape (2023). Yet, Bono and Judge (2004) assume rather weak associations between the big five personality traits and charisma. In the same way, Michalsky et al. (2020) argue against a too close correlation of charismatic personality with charismatic speech, pointing out that charismatic speech can be learnt. Thus, we should not try to establish too close links between single personality traits and charisma. It might be better to establish a direct link between features and charisma as a complex trait, with components (sub-traits) such as persuasiveness that display themselves characteristic acoustic-prosodic feature values.

Figure 3 summarizes the formal acoustic aspects dealt with in this subsection and the linguistic aspects described in Section 3.2. At the same time, it relates these formal aspects to the functional aspects: the motivation behind creating charismatic agents; the models employed by us; and the perception and impression that such charismatic agents have on the human interaction partners. These components are employed to create applications where charisma is harnessed to achieve their specific goals. Ethics has to assess and possibly restrict the use of charisma in these applications, see Section 6.



### 3.2. Linguistic markers

As far as linguistic markers are concerned, the use of informal language, high occurrence of pronouns, and avoidance of synonyms can be used to elicit greater warmth, while the opposite holds for formal, complex language. For pronouns, those that involve the audience, e.g., *we* and *you*, are useful for creating a better first impression (Biancardi et al., 2019). In addition, Rosenberg and Hirschberg (2009) found that using first-person pronouns positively correlated with the charisma ratings of political candidates in spoken but not in written form.

Adjectives can serve as markers for the charismatic content of language. They can be clustered via concepts such as sociability and morality for warmth or ability and agency for competence (Fraser et al., 2021). The usage of adjectives, as opposed to nouns in describing persons, affects the generated impressions, with nouns conveying a greater sense of defining immutable traits (Fraser et al., 2022). When referring to groups of people, the choice of descriptors can evoke various impressions of warmth and competence via associated stereotypes; consider, e.g., the differences between *the elderly*, *old people*, *old folks* and *senior citizens* (Fraser et al., 2022).

The clarity of the intended message also affects the perception of charisma. A lower amount of disfluencies may make a speaker appear more confident and focused (Kirkland et al., 2023). The negative effect of disfluency is more pronounced for linguistics than for prosody according to a comparison between speech and transcripts by Rosenberg and Hirschberg (2009), possibly because the audience may expect it in spoken but not in written form. Regarding the content of a message, conveying more information is not necessarily beneficial from a charisma perspective; speakers using more content words (relative to the number of function words) are rated as being less charismatic, possibly due to higher rhetorical complexity (Rosenberg and Hirschberg, 2009).

Charisma is closely related to being able to influence others; thus, we also examine linguistic markers of persuasion. Guerini et al. (2003) propose a taxonomy resting on four pillars: cognitive

state, social relations, emotional state, and interaction context. Here, cognitive elements refer to goals and beliefs of agents and concepts related to them, social elements deal with power dynamics between relevant persons, emotional elements can be used to enhance or diminish a message, and contextual elements can add useful information. Persuasion strategies are then grouped by their objective: inducing a change in beliefs, and inducing a change in actions. The former can be achieved by appealing, e.g., to the opinions of experts, to public opinion, or to empirical evidence. The latter may follow a social strategy by appealing to someone from whom the target derives standards or morals, or to the target's social image at large. Another option would be to present imaginary consequences, either positive via promises or negative via threats. A charismatic agent may select and modify these strategies to improve the success rate.

### 3.3. A note on other modalities or: the fields of para-linguistics and non-verbal communication

These two fields (scientific disciplines) deal with phenomena that are relevant for indicating and modeling charisma: "paralinguistics" and "non-verbal communication." Figure 4 presents an overview with their intersection, indicated by green color in the paralinguistics box. Both have in common that they are defined and have been evolved as "being not" something else: "para-"linguistics in the middle of the last century as something important indicated and denoted by written and spoken language, besides its grammatical structure and semantic meaning; "non-"verbal communication in the 1960s, "largely in reaction to the overwhelming emphasis placed upon verbal behavior in the field of communication" (Jones and LeBaron, 2002). More details can be found in Batliner et al. (2022) and Schuller and Batliner (2014, p. 7–10); see as well above Section 3. Both fields overlap

with the philosophical/linguistic field of pragmatics (Levinson, 1983; Korta and Perry, 2020). The features [vocal] and [verbal] have already been explained in Section 3. They characterize vocal and verbal behavior of the “subject of interest” and are therefore [-distant]—they so to say “belong to” the subject; they can be [ $\pm$ dynamic] because they can be short-term (accentuation, intonation contour) or long-term (pitch range, stable voice quality). For non-verbal, i. e., body(-to-body) communication, we establish four different feature constellation domains which are detailed in Figure 4 in the “sub-domain/phenomena” column:

- Body outfits as [-dynamic,-distant]; this means that these characteristics do not change during the communication, and that they belong to (are closely tied to) the “subject of interest”;
- Body extensions do as well normally not change during communication—at least, they are stable and do only change when the subjects moves. However, they are not tied to the subject but more or less distant from it: [-dynamic, +distant].
- Body distances can be either/or, i. e., [ $\pm$ dynamic,  $\pm$ distant]: either close (haptics/touch) to the communication partner, or more or less distant; this can but need not change dynamically in the course of the communication;
- Body kinesics characterizes (parts of) the body’s movements of the subject and is therefore [+dynamic, -distant].

First and foremost, the feature values for our four features define the senses and the modality, and by that the processing channels: within paralinguistics, [+vocal] the audio channel, [-vocal] written language, i. e., natural language processing (NLP). Within non-verbal communication, [-dynamic] could do with only image processing but not [+dynamic]. [-distant] means for both fields that we can concentrate on the subject of interest whereas [+distant] means that we have to protocol its environment. Note that the terms in the columns “non-verbal communication, domain and sub-domain/phenomena” are largely but not fully taken from Burgoon et al. (2022, p. 22–23). They specify and exemplify the respective domains.

We now want to exemplify the four main domains of non-verbal communication with studies that address specific sub-domains and phenomena, either directly for modeling charisma, or for modeling user states and traits that can be employed for modeling charisma; note that sometimes, studies span over two domains and are dealt with in only one of them. Body kinesics is arguably the most explored area; accordingly, more studies can be found on this domain.

Body outfits: Maran et al. (2021) found out that “leaders can manipulate their style of attire to actively shape their followers’ impressions of themselves.” In the same vein, Lennon (1990) determined the effects of clothing on attractiveness. In the study of Prehn-Kristensen et al. (2009), emotional contagion seemed to be mediated by the olfactory system.

Body extensions: Personal bookshelves as background in video conferences attracted attention in the time of COVID-19 and the quarantine it caused; its carefully selected content can contribute to the perceived competence of the speaker (Towheed, 2022)—or to the opposite impression when inappropriate publications can be detected.

Body distances: Appropriate proxemic behavior of the physician (such as directly facing the patient, legs uncrossed, moderate eye contact) results in high rapport between them and the patients (Harrigan et al., 1985) which is, of course, a necessary prerequisite for a beneficial treatment. The same way, touch is a powerful tool establishing human connections (Kelly et al., 2020), crucial for social development, and necessary for children in order to grow up healthy (Weiss et al., 2000; Van Erp and Toet, 2015). It is our primary channel for expressing intimate emotions and can effortlessly establish social presence (Van Erp and Toet, 2015). In addition to distinct emotions like love, anger, and fear, touch can also convey more complex emotional patterns such as trust, receptivity, and affection (Burgoon, 1991; Hertenstein et al., 2006, 2009; Van Erp and Toet, 2015). As previously mentioned, charismatic persons radiate characteristics like trustability, presence, and warmth, which makes affective touch an essential modality next to audio—if appropriate in the specific situation.

Body kinesics: Pauser et al. (2018) found that “certain arm actions, arm postures, and action functions have a significant effect on charismatic appearance and can in turn produce favorable attitudes toward the salesperson.” In Koppensteiner et al. (2016), body postures of “maximal expansiveness” were found to be strong predictors of perceived dominance. Cuddy et al. (2008) investigated how warmth and competence are perceived based on behavior at interpersonal and intergroup levels. Smiling, as well as engaging gestures, touch, and mirroring were found to increase the impression of presence and warmth, while disengagement and creating physical distance by leaning back or turning away decrease it. Expansive and open body poses, suggesting power and dominance, resulted in higher impressions of competence. For hand gestures, the use of object adaptors and ideationals (relating to spoken words) improved the speaker competence, while self-adaptors decrease it (Biancardi et al., 2017). In general, a speaker’s delivery can have a great influence on their credibility, i. e., a strong delivery is more likely to lead to high credibility than a weak one (Holladay and Coombs, 1993); factors contributing to a good delivery include eye contact, gestures, and facial expressions (Holladay and Coombs, 1993). This is not surprising as gestures and facial expressions can innately radiate charisma (Towler, 2003). When a person’s gaze is focused on the conversation partner, this is a sign of attention and shows both interest in the conversation and commitment to the conversation partner (Knight and Simmons, 2013; Freeth and Bugembe, 2019). That is, if the gaze is wandering through the surroundings it may evoke the impression that a person is not fully listening and wants to distract themselves with seemingly more interesting things. In leaders, eye-directed gaze has been found to be linked to their charisma. Maran et al. (2019) used eye-tracking during a simulated leadership scenario and found the frequency and duration of eye-contact to be positively correlated with both self- and follower-perception of a leader’s charisma. Furthermore, facial expressions and gaze patterns are channels through which charisma can exert its contagious effects (Cherulnik et al., 2001).

This overview of studies on (possible) “non-verbal,” i. e., body markers of charisma cannot be exhaustive; it is intended to exemplify the various sub-domains and phenomena depicted in

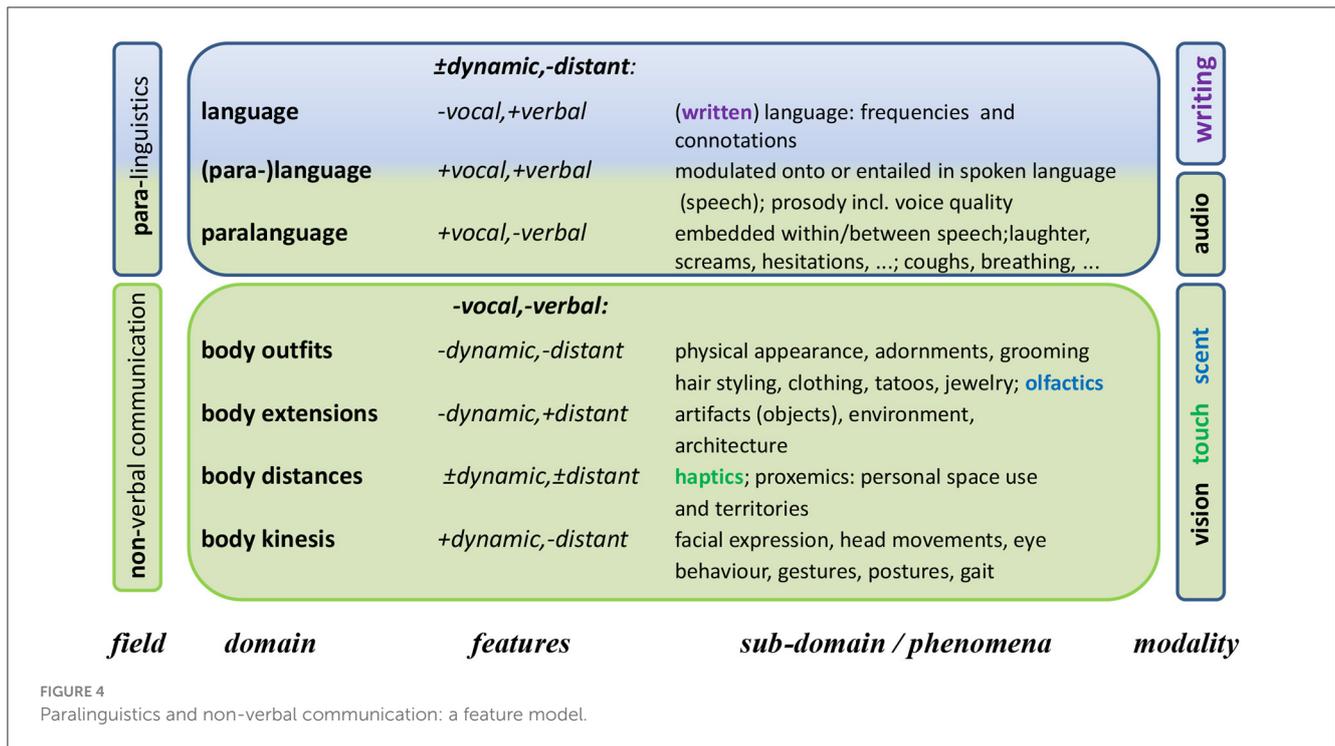


Figure 4. For a full account of non-verbal communication, we refer to Burgoon et al. (2016, 2022).

### 3.4. Different cultures, different groups

So far, we reported the state-of-the-art formal markers of charisma that are based on the well-known *Western, Educated, Industrialized, Rich, and Democratic* (WEIRD) (Henrich et al., 2010) population, with the equally well-known biases, as far as gender, class, and ethnicity is concerned. To the best of our knowledge, cross-(sub-)cultural aspects of formal markers of charisma have not yet been dealt with. We mentioned a few studies on prosody and gender or cultural differences in Section 3.1. Gender differences in body kinesis (dominance, trustworthiness, and competence) were addressed in Koppensteiner et al. (2016). Social signals (Vinciarelli and Mohammadi, 2014) in small group interaction and different roles have been addressed in Charfuelan and Schröder (2011) and Sapru and Bourlard (2015); for instance, project managers often speak with a louder voice than the average (Charfuelan and Schröder, 2011). Meyer (2014) describes the different roles a leader of a meeting has to play in different cultures (“leadership style”: “egalitarian” vs. “hierarchical” leadership); a chapter is titled: “Strategies for persuading across cultures.” We surely can expect different acoustic-prosodic and linguistic markers of such leadership in different cultures. In the same way, dialogue strategies differ across cultures (Lugrin and Rehm, 2021): for instance, speaker overlap can be modeled as a strong sign of conflict in the “Anglo” style (Grèzes et al., 2013), but as sign of interest in the “Latin” style, see (Trompenaars and Hampden-Turner, 1998; Fitzgerald, 2002; Hilton, 2016).

## 4. Computational aspects of charisma: modeling

After analyzing the formal markers of charisma in Section 3, we now deal with the modeling of charisma from a computational perspective. Automatic recognition of charisma describes the detection of the sociological markers, referring to the society, and of the psychological markers, referring to the individual, for charismatic behavior using machine learning (ML) approaches. Similarly, the automatic generation of charisma outlines methods for generating auditory or visual charismatic traits.

### 4.1. Automatic recognition of charisma

Charisma can be registered via a wide range of modalities, ranging from facial movements and gestures to speech and physiological attributes like heart rate and skin conductance. Since charisma is an interpersonal phenomenon, computational analysis can focus either on the sender projecting charisma, on a receiver forming an impression, or on dyadic interactions between the two (Wang and Chanel, 2021). Here, we take up the stated sociological, psychological, and popular science perspectives and translate them into computational aspects of phonetics, linguistics, and other modalities in automatic recognition of charisma.

#### 4.1.1. Audio

Power, described by Fox Cabane (2013) as high competence due to skills, abilities, or intelligence, can be detected from audio by analysis of features related to fluency, such as speech rate and pauses. Luzardo et al. (2014) perform an automated

evaluation of student presentation skills and found a formant-based detection of filled pauses useful for classifying the overall quality of presentations. Further, they observed that speech rate is positively correlated with a speaker's self-confidence. A similar approach based on detecting filled pauses is taken by [Ochoa et al. \(2018\)](#) in the audio modality of their automatic feedback system for presentation skills.

As charismatic speakers often attain special positions in multi-person settings, the automatic analysis of group impressions, such as emergent leaders or perceived personality traits, is highly related to the recognition of charisma. [Beyan et al. \(2019\)](#) apply unsupervised deep learning (DL) models to analyse temporal dependencies in speech and video signals for the detection of emergent leaders in small groups. When dealing with multiple speakers, charismatic traits can often be detected more easily when interpersonal feature dependencies are considered. [Okada et al. \(2019\)](#) utilize an approach for co-occurrence mining of audio-visual events to explicitly model inter-person relationships in dyadic and group settings.

In a similar vein, the mimicry of a conversation partner can help establish a connection in dyadic interactions. This entrainment may happen either subconsciously, or deliberately to project greater warmth and presence. The automatic detection of this phenomenon has been of special interest in psychotherapy where it is highly related to empathy as an interpersonal behavior, specifically between the therapist and the patient. [Xiao et al. \(2013\)](#) take a computational approach of modeling entrainment by obtaining statistical information of the acoustic similarity between consecutive turns of the dyadic partners. They empirically confirm the link between expert-rated empathy and vocal adaptation. More recently, [Nasir et al. \(2022\)](#) utilized deep unsupervised learning to generate speech representations containing relevant information for vocal entrainment.

Outside of the context of psychotherapy, [Amiriparian et al. \(2019\)](#) investigate "synchronization" (i. e., entrainment) in such dyadic conversations, using acoustic and linguistic features on a dataset with 394 speakers of six different cultures. For the acoustic analysis, both handcrafted EGEMAPS ([Eyben et al., 2016](#)) and deep DEEPSPECTRUM ([Amiriparian et al., 2017](#)) features are extracted. Autoencoders are then used to measure the degree of entrainment by training on one person and then reconstructing on their conversation partner. As the conversation continues, the reconstruction error tends to decrease across the six cultures, indicating that speakers are mutually adapting to each other.

The entrainment phenomenon is finally highly related to the concept of "presence" as described in Section 2. According to [Fox Cabane \(2013\)](#), it is characterized by dwelling in the moment, active listening and focusing on one's conversational partner. Recognition approaches for this aspect of charisma can therefore additionally aim to detect these behavioral cues.

In the early years of prosody research in automatic speech processing ([Batliner and Möbius, 2020](#)), the focus was on detecting and classifying linguistic phenomena such as phrase accents, boundaries, disfluencies, sentence modality, and dialogue acts; such explicit modeling was then superseded by implicit modeling in AI approaches. Yet, it might gain momentum in our context, when we want to model markers for charisma. Asking questions during

a conversation indicates that a person is listening and interested in what the conversation partner says; thus, it can indicate the *presence* in a conversation. In this context, acoustic and phonetic features are deployed, at which lexical features can also be crucial for the correct identification of declarative questions ([Ando et al., 2018](#)). Furthermore, recurrent neural networks (RNNs) are applied in order to obtain the high-level contextual information over time ([Tang et al., 2016](#)).

Before asking a question, it can also be beneficial to make a short pause, in order to show that one thinks about what the conversation partner has said, before giving an answer. This can convince the other person that one is listening carefully and is present in the conversation. [Trouvain and Werner \(2022\)](#) define these types of pauses as "gaps at turn changes in conversations" and do not regard them as typical speech pauses that are defined as "pauses in connected speech section." Regarding speech production and the temporal structure of speech, pauses also play a crucial role ([Trouvain and Werner, 2022](#)). We have to distinguish between silent pauses and filled pauses such as "uh" or "uhm" ([Batliner et al., 1995](#); [Bilac et al., 2017](#); [Trouvain and Werner, 2022](#)). [Bilac et al. \(2017\)](#) for instance extract Mel-frequency cepstral coefficients (MFCC) audio features and apply support vector machines (SVMs) and random forests (RFs) as classification methods. In general, automatic detection of silent speech pauses and audio silence has been possible for quite some time ([Iqbal et al., 2018](#); [Xu et al., 2020](#)).

#### 4.1.2. Language

A textual analysis lacks the information of prosody from a speech signal and must instead focus on linguistic cues. For purely text-based empathy and warmth recognition, we highlight two applications here: mental health support and stereotyping in social media. For presence, we also examine entrainment in conversations.

[Sharma et al. \(2020\)](#) investigated empathy in the context of seeker-response interactions on text-based support platforms. Their framework, adapted for asynchronous communication, includes three mechanisms: emotional reactions to the seeker, interpretations conveying understanding, and explorations to improve understanding. A dataset of interactions collected from TalkLife and mental health subreddits was annotated in terms of empathy and rationales (text snippets motivating the empathy annotation). Then, a multi-task model based on two pre-trained ROBERTA ([Liu et al., 2019](#)) encoders acting on seeker and response posts and a single attention layer combining their embeddings was proposed. ROBERTA (A Robustly Optimized BERT Pretraining Approach) is a transformer-based language model (LM) which builds upon the popular BERT model by modifying some key hyperparameters and training with a larger training data size ([Liu et al., 2019](#)). The inclusion of seeker post and attention was found beneficial while fine-tuning the encoders and adding the rationale task gave minor improvements.

Stereotypes are frequently encountered in social media posts, and may positively or negatively shape opinions on groups. [Fraser et al. \(2022\)](#) apply the warmth and competence model to stereotype identification by constructing a synthetic training set and building a model that can identify stereotypes in crowd-sourced data. First,

using a seed lexicon, polar directions for warmth and competence are defined in a word embedding subspace. Then, sentences are created via templates filled with words of known polarity from the lexicon. For the word embeddings, ROBERTA is used, with GLOVE vectors serving as the baseline. A combination of ROBERTA embeddings with intermediate dimensionality reduction via partial Least-Squares performed best. Also, the generation of sentences combining both warmth and competence-associated words improved accuracy by increasing the orthogonality of training pairs.

Amiriparian et al. (2019) use WORD2VEC to extract embeddings. The dyadic conversations, in which the conversation partners discuss their view on an advertisement spot from the 90s, are split into two parts, and the cosine similarity between their embeddings is computed. In addition, the co-occurrence of words between subjects in each part is counted and normalized with the total number of words. The word embeddings showed little entrainment (i.e., mutual adaptation of conversation partners) compared to the audio features, possibly indicating that the effect was happening too gradually on the linguistic level to be measured during the short conversations. Word usage showed a clearer correlation but strongly differed across the six cultures (Chinese, Hungarian, German, British, Serbian, and Greek) of the utilized database, being most pronounced in British subjects.

#### 4.1.3. Other modalities

While we focus on verbal aspects of charisma, non-verbal cues, such as those described in Section 3.3, can be a target for automatic recognition approaches as well. In the following, we will present a few computational approaches that exploit other modalities, such as videos and images, to obtain and analyse these markers. In this context, videos and images might be especially beneficial for recognizing how present and involved a person is in conversations by considering eye contact and facial expressions.

Some studies already try to use eye tracking in order to analyse attention and gaze patterns during social interactions (Rogers et al., 2018; Vehlen et al., 2021). Rogers et al. (2018) also point out that some people show a preference toward mouth gaze, some for eye gaze, and others tend to vary the extent of their gaze between eyes and mouth. The authors apply a standard remote eye tracker, consisting of an infrared sensor and a corresponding camera. Vehlen et al. (2021), on the other hand, employ special eye-tracking glasses enabling the opportunity for real-world experiments. In order to avoid expensive high-end processing devices, Zdarsky et al. (2021) introduce a convolutional neural network (CNN) relying on video frames from low-cost web cameras. Another study also aims at making eye tracking available for everyone owning a mobile device with a camera (Krafka et al., 2016).

Besides our eyes, facial expressions are a very important tool to express excitement and emotions. Erez et al. (2008) state that charismatic leaders exhibit more aroused behaviors than non-charismatic leaders. For instance, smiles are arguably among the most visible and frequent markers and can convey a feeling of warmth and intimacy but also of fear or compliance (Awamleh, 2011). There are already several approaches for automatic facial expression recognition (FER), most of them utilizing DL and some

sort of CNN in particular (Wang et al., 2020; Minaee et al., 2021; Revina and Emmanuel, 2021; Li and Deng, 2022). The rough pipeline is to feed an input face image to a trained network and obtain a probability for a certain emotion category, such as happy or sad. In addition to employing CNNs as the basic architecture blocks, there are extensions to improve performance, e.g., adding an attention mechanism to the network (Li et al., 2019).

## 4.2. Automatic generation of charisma

We can approach the task of automatically generating charisma in two different ways. On the one hand, we can use an approach to try to imitate the charismatic characteristics of people previously defined in the literature. Charismatic persons are—as outlined above—characterized by a certain way of speaking (e.g., pitch, duration, or rhythm during the conversation). The characteristics assigned to charismatic individuals have been established in prior research (Klofstad et al., 2016; Davidson, 2021), among numerous others mentioned earlier. These collectively form an expert-based definition. Leveraging this knowledge alongside generative ML techniques allows for the targeted incorporation of these traits during speech generation (Baird et al., 2019; Amiriparian et al., 2023), ultimately leading to an enhanced perception of charisma.

### 4.2.1. Expert-based generation of charisma

To simulate charismatic behavior, the previously identified building blocks must be taken into account when generating spoken language (or other modalities). In recent years, progress has been made in two main areas: first, generative methods (Borsos et al., 2023) for creating completely new audio outputs, and second, constrained audio generation, as well as style transfer approaches, e.g., Manzelli et al. (2018), Zhang et al. (2019), and Huzafah Bin Md Shahrin and Wyse (2020), where an existing audio file is stylistically adapted to pre-defined properties.

The latest results of generative methods for speech such as AudioLM are almost indistinguishable from real speech by humans (Borsos et al., 2023). The high audio quality of the generated samples paves the path for further charismatic audio generation. Based on this, style control and style transfer approaches can be used to change certain features of the voice (Zhang et al., 2019; Huzafah Bin Md Shahrin and Wyse, 2020). For example, Baird et al. (2019) analyzed if deep generative audio can be emotional. In doing so, they changed pitch as an important speech characteristic. In a similar way, other features can be adapted, leading to a more charismatic voice.

In addition to the audio modality, this approach can be extended to other modalities, such as video or text. Based on findings from previous work investigating which features are perceived as particularly charismatic in the respective modality, these constraints can be considered in generative methods. For example, Ghorbani et al. (2023) explore gesture generation from speech which can be used as an additional modality, resulting in an overall charismatic perception of an AI.

#### 4.2.2. Learning-based generation of charisma

Automatically generating charisma can also be formulated as a weakly supervised ML task. For example, reinforcement learning (RL) methods have become increasingly popular in recent years in audio processing (and beyond) and are based on rewarding desired and punishing undesired behaviors (Latif et al., 2023).

Applied to charisma generation, various characteristics of speech are exploratively tried during generation. In doing so, a human-in-the-loop (Abel et al., 2016; Liang et al., 2017) RL approach can be employed such that the reward function includes feedback from users on how charismatic the generated output is perceived. For example, pitch shifting, ranging from a low up to a very high pitch, can be explored. Taking user feedback into account, the optimal pitch that is perceived as most charismatic (or seems to be so, as it best solves a task that benefits from high charisma) can be determined. In addition to direct user feedback, automatic charisma recognition approaches can be applied as a reward function to evaluate whether the generated behavior is charismatic. In the context of generating emotional speech, Liu et al. (2021) present such a paradigm. In an RL setting, they train an automatic text-to-speech model to generate speech with emotions that can be discriminated by an automatic speech emotion recognition model. Another advantage of RL is that new, so far unknown, charismatic traits can be discovered using this method. This could range from obvious charismatic traits to entirely new charismatic behaviors that are as yet undiscovered.

Obviously, in addition to the audio modality, RL for charisma generation can be similarly applied to video and text and beyond. For instance, it might be beneficial for robots or virtual humans/agents to imitate charismatic gestures and appearance in general (Pentland, 2010). Additionally, Won et al. (2021) have already physically simulated humanoids performing competitive two-player sports, boxing and fencing, in a high degree-of-freedom environment. The applied control policies generated responsive and natural-looking behaviors (Won et al., 2021).

Furthermore, if we organize the process of charisma generation into levels of abstraction, methods such as hierarchical RL (Saleh et al., 2020; Rothenpieler and Amiriparian, 2023) can offer a valuable approach for modeling charismatic behaviors at various levels of complexity. It can begin from overarching attributes like empathy and humor (Christ et al., 2022a; Kathan et al., 2022) (high-level) to specific behaviors such as affective storytelling (Christ et al., 2022b) and body language (Wu et al., 2022) (mid-level), down to fine-grained details like speech patterns (Niebuhr et al., 2016; Amiriparian and Schuller, 2021) and linguistics (Rosenberg and Hirschberg, 2009; Fraser et al., 2022) (low-level). This multi-level approach could enhance the comprehensiveness of charisma generation and capture its multifaceted nature.

## 5. Applications of computational charisma

As outlined above, charisma is often associated with charm and “magnetism,” empowering to influence and inspire. Even if AI is not “experiencing” or showing charisma as humans could, it can be trained or programmed to appear charismatic. This leads to a plethora of interesting use cases which, however, often come with

a number of risks and dangers. We now list examples, broadly grouped.

### 5.1. Communication

Let us first consider different aspects of communication: AI empowered with charisma may lead conversations with humans in potentially more convincing ways, engaging them, and potentially influencing them toward decisions in favor of the AI's goals. Similarly, it can simulate empathy and in general be perceived as more intelligent due to socio-emotional intelligence skills. Charismatic techniques may improve everyday conversations by creating an emotional connection within communication partners or followers, and make someone appear more powerful, competent, and worthy of respect (Antonakis et al., 2012). This holds in conversations, but also in public speeches via means of AI in the future—potentially to large audiences via the internet or in real-life settings. In automatic translation, AI could help translate charismatically or preserve charismatic traits in the target language. If AI is used for communication analysis, such as in mediation between human conversational partners, it can sense who is being more charismatic, more influencing, or more affected by the other party in a somewhat quantitative and subjective neutral manner.

### 5.2. Leadership

Let us now turn to aspects of leadership: most use cases in research but also real-world applications pertain to the training of charisma for enhanced leadership. More specifically, leaders in a variety of fields as politics, religion, or business highly benefit from being persuasive and likable to achieve specific team goals and overcome potential resistance within employees or party members. Leaders may benefit greatly to win the trust of followers, manage delicate operations, punish and reward, and achieve their goals (Antonakis et al., 2012). Levay (2010) even argues that charismatic leaders are invariably proponents of change. Charisma may help AI inspire and motivate teams—including encouraging and guidance through difficult challenges. Moreover, charisma can help AI to build and bind teams and communities. Especially in times of remote work and digitally conveyed social interaction as video calls and emails, team coherence and trust and commitment are highly welcome. More specifically, team leaders may be fed back easily in their communication style with an AI that recognizes charismatic language and enhances team outcomes, putting the team members at ease and comfort by simultaneously persuading them of a new idea. Yet, we have to be aware that charisma can be abused for unethical goals; to enforce red lines against such a use will be both mandatory and challenging. Simple ethics washing has to be avoided (Wagner, 2018; Bietti, 2020; Ressayguier and Rodrigues, 2020; Batliner et al., 2023); see Section 6. Furthermore, a charismatic AI could also recruit new staff potentially more successfully than a non-charismatic one. However, it can also help it in negotiations within teams or with other parties in persuasive ways. If AI assists in decision making, charisma may help it to get the decisions across to the humans it presents it to.

## 5.3. Healthcare

Healthcare is another suited field for charismatic AI: be it for mental health or general health care—charismatic AI could provide a compassionate, empathetic, and reassuring support providing communication and assistance during diagnosis or interventions and therapy. Recognizing charisma can also provide benefits for improving mental health. Mental health issues are widespread, affecting nearly one in five people worldwide (Holmes et al., 2018) and incurring enormous human suffering and economic costs. The majority of patients prefer therapy to medication (Holmes et al., 2018), requiring the development of new solutions to improve the effectiveness of treatment. Depending on the application, those systems may either work in real-time for live session support or in an offline fashion for reviewing progress. A system that can process interactions and estimate the charismatic content may be a valuable tool for training professional practitioners. An effective therapist or counselor can provide a sense of engagement and empathy to the patient, which in terms of charismatic dimensions involves both warmth (to show sympathy) and competence (to comprehend and engage with a patient's problems). In particular, this holds in case patients are reluctant or not sufficiently committed to overcome adverse feelings that sometime go along with behavior change as, for instance, confronting oneself with an anxiety-eliciting situation or the acquisition of new behaviors that feel initially uncomfortable. Both therapists and clients could benefit from the rapport, empathy, believe of the ability to help, or persuasiveness of the therapist that go along with charisma. An AI solution that can help review and train these skills will likely also be beneficial for the increasingly popular digital support platforms, to improve the qualification of responders (Sharma et al., 2020). Therapists are usually trained in providing ease, comfort, and being empathetic and receive supervision in doing so; yet, there is great potential in improving speech markers of influence and affability. Providing AI-generated automated feedback on the interaction and conversation style between therapist and client may improve the communication style and hence, therapy outcomes to a great extent. Considering the human-to-human conversation mediation alluded to above, this could include couple or other counseling endowed with active listening, and empathetic moderation.

Beside such fields, numerous other applications may benefit from charisma:

## 5.4. Other fields of application of computational charisma

### 5.4.1. Education

In education, a charismatic AI may be more engaging and captivating for students taught by it. Furthermore, as a coach and mentor, charismatic AI could be more motivating. An AI that can sense and measure charisma can also help in tutoring about charisma, i.e., teach humans to be

charismatic by monitoring their progress (Antonakis et al., 2011).

### 5.4.2. Customer service, marketing, and sales

Further, in customer service, marketing and sales, charismatic AI could provide friendly and likable service, but also convince customers.

### 5.4.3. Social media

In social media, charisma-empowered AI could interact with the social media users and generate a large group of followers, e.g., to prime opinions or provide positive brand connotation. This would include charismatic interaction with followers, responses to public reactions, or creation of new content in engaging ways. AI that has an understanding of charisma can also analyse charismatic behavior and skills of users—be it, e.g., for scientific analyses or identifying potential influencers early on.

### 5.4.4. Gaming

As to gaming, non-player-characters (NPCs) driven by AI could be charismatic in oncoming games leading to an increasingly immersive gaming sensation.

### 5.4.5. Events and hospitality

When it comes to event management and hospitality, charisma-empowered AI could provide engaging interaction with attendees of future virtual events. These may encompass a wide variety of events reaching from online conferences and workshops to virtual fair trades, virtual tours of galleries, museums, real-estate properties, or virtual concerts, festivals, parties, and other entertainment events. In particular, this could even include fundraising at suited events, where a charismatic AI could interact with donors persuasively. Beyond, hospitality at virtual or real-world occasions including checking guests at hotels in and out, question answering, and recommendation giving, could be realized more charismatic by enabling future AIs accordingly.

### 5.4.6. Embodied or virtual assistants and companions

More generally, any form of embodied or virtual AI—e.g., assistive or companion agents—could largely benefit from charismatic skills in the interaction with their users. This could help them in their communication and motivation as well as companionship, especially with lonely individuals. Across use cases, charismatic AI may be better positioned to personalize services to users by gaining access to their individual preferences, context, and history.

## 6. Ethics of computational charisma

Of course, we do not aim at dark charisma for appealing to the baser human instinct; moreover, we do not want to harness “deceiving charisma,” i. e., bright charisma for achieving goals that are *per se* unethical. As for spoken language as modality, a cover term might be “emotional speech” in the sense of adding more credibility and user attachment to human–computer interaction. Charisma is thus not a goal in itself, but a means to better achieve its goal. Bright charismatic speech in itself cannot be unethical or ethical—it always depends on the application we are envisioning and the ideology it can be used for. Thus, out of all the (ethical) cornerstones relevant for applications defined in Batliner et al. (2023), most might be “secondary” for charisma, i. e., depend on the (type of) applications that use charismatic speech to pursue its goals. Yet, by providing charisma as a tool, we have to account for the possibility that this tool can be used for “dark goals” or, simply, that the outcome is not favorable. For instance, radical organizations may implement a charismatic communication style to persuade followers and even manipulate them for their goals. Thus, ethical requirements can be higher. In the same way, dealing with vulnerable groups of course puts higher requirements on ethics, as for instance, *privacy* and *avoiding harm* are concerned (Batliner et al., 2022). If neutral information is aimed at or had to be preferred, charisma should not be turned on.

A self-learning system can adapt to templates and users the same way as the chatbot Tay learnt racist language from its users (Wolf et al., 2017). This is a problem for every empathic virtual agent (Pamungkas, 2019). Both dark charisma and bright charisma employed for dark goals can be created unwillingly or on purpose. In the first case, not only do algorithmic measures have to be taken, and in both cases, society has to define red lines against them. Guerini and Stock (2005) reason about different capabilities of persuasive agents in case of ethical dilemmas (conflicting goals): (i) detect them and pass them on to a human; (ii) compute a possible conduct and pass this on to a human for the final decision; (iii) make own decisions. So far, we cannot envision artificial agents capable of doing this kind of reasoning in a reliable way. Thus, two rules should be followed: first, ethical dilemmas should be avoided by design, and if they are encountered, the decision has to be passed on to a human. Principles of persuasive technology design are stated in Berdichevsky and Neuenschwander (1999).

The most prominent specific ethical requirement might be *disclosure of automation* (Mohammad, 2022), belonging to the ethical cornerstones *transparency* and *accountability*: A charismatic AI application has to make clear that the user is not interacting with some human being but with a computer. This must not be done in the small print but in a way that is really visible and understandable to the user, and it has to be done even if it had been doubted that such a better understanding “can protect [users] better against unconscious social reactions” (Krämer and Manzeschke, 2021).

Transparency and accountability seem to be the primary cornerstones that impact *autonomy*, i. e., provide the possibility for the user to be aware of the artificial charisma the application / the agent is equipped with. Then comes *intrusive*: Ethical requirements are higher the more intrusive the application is. Montemayor et al. (2022) claim that genuine empathy in healthcare is not possible for

AI because it cannot be really emotional. Yet, a charismatic agent might at least act *as-if* but we have to make it clear toward the patients that the AI (robots, avatars) is artificial and does not have emotions or empathy itself, in order to prevent this erroneous and dangerous attribution that even can lead the user to fall in love with such a charismatic artificial agent. The illusion of humanness can create the “uncanny valley” effect (Mori et al., 2012) when emotional/charismatic agents are close to humans but still not close enough, by irritating the human interaction partner. Both this uncanny valley and a too-perfect humanness might be avoided by explicitly mentioning the artificial character of the agent, or by creating it in such a way that its non-human character is evident. Note that some authors argue, under certain premises, in favor of anthropomorphic robots (Darling, 2017) or deception-capable robots (Isaac and Bridewell, 2017). Although possible benefits might be evident, it is not clear at all how any dishonest use of such robots—or, in our case, charismatic agents—could be prevented if not banned from the beginning.

## 7. Conclusion and outlook

We discussed a “brick by brick blueprint” for a charisma-savvy AI—able to analyse human charisma and generate charismatic behavior itself. We started by introducing the concepts of charisma from a rather broad perspective. We then discussed functional aspects of charisma in psychological models, mainly introducing two concepts based on the factors of influence and affability as well as the theoretical concept of power, presence, and warmth as pillars of charisma. We argued that charismatic behavior can be acquired and presented a brief summary of the literature on its acquisition. Building up on this reflection, we then moved to formal aspects giving specific details on charisma as portrayed in spoken language. The motivation to concentrate on this modality had already been given in the introduction. We further outlined the computational aspects of modeling charisma. Here, we summarized the small body of literature on the automatic recognition and generation of charisma for audio, language, but also other modalities. As to the generation of charisma, we highlighted two avenues: first, based on the findings in the literature summarized up to that point in this article, one could design a charisma-empowered AI based on expert knowledge. Alternatively, weakly supervised ML could be exploited by either active learning methods questioning users about the charisma skills of an AI or even by learning reinforced. In the latter, an AI would gain charismatic skills, potentially even unknown to date to humans. This would help better accomplish its goal by interacting with users in real-world tasks, ideally at scale. We then moved toward the plethora of potential use cases of charisma-enabled AI, before introducing major ethical concepts to be considered at all times.

Given the state-of-knowledge on charisma and the state-of-play in today’s AI, it seems perfectly possible to endow AI with charisma skills. Currently, the literature on using ML for the recognition or generation of charisma or traits thereof largely focuses on the individual in isolation related to fields such as Affective Computing. However, disciplines such as Social Signal Processing moved also the consideration of the interplay between

communicating parties into the foreground, which can be crucial for modeling charisma, e. g., when it comes to entrainment. Audio, text, and video have so far been mostly considered, but touch, and more general haptics, have been addressed as well. In the future, other modalities including smell and other biological signals could be included. Further, the loop between recognition and generation of charismatic behavior might be fully closed by learning the charismatic input/output of an AI “end-to-end.”

As first-of-its-kind, this article sheds light on computational charisma with broadly selected bricks of its concepts, computational analysis and generation, particularly in speech. However, more research was conducted in more specific fields like leadership research and non-verbal behavior and may be further interesting for follow-up in depth-research.

Overall, we envision a plethora of use cases with great value of charisma-savvy AI. Weakly supervised AI learning from large data may easily lead to new charismatic behaviors found by AI potentially reflecting back on human-to-human charismatic behavior. Charismatic AI may also empower “dark” purposes or lead to negative effects such as AI influencing voters, e-shoppers, getting users addicted to or falling in love with the AI, and many more. As a community, we have to always contribute our best to assure positive usage and the protection of users—including in particular from a technical end. Thus, let us be best prepared for the rapid advent of charismatic AI.

## References

- Abel, D., Salvatier, J., Stuhlmüller, A., and Evans, O. (2016). “Agent-agnostic human-in-the-loop reinforcement learning,” in *Proc. of NIPS Workshop FILM* (Barcelona), 13.
- Amiriparian, S., Gerczuk, M., Ottl, S., Cummins, N., Freitag, M., Pugachevskiy, S., et al. (2017). “Snore sound classification using image-based deep spectrum features,” in *Proc. of INTERSPEECH* (Stockholm: ISCA), 3512–3516. doi: 10.21437/Interspeech.2017-434
- Amiriparian, S., Han, J., Schmitt, M., Baird, A., Mallol-Ragolta, A., Milling, M., et al. (2019). Synchronization in interpersonal speech. *Front. Robot. AI* 6, 116. doi: 10.3389/frobt.2019.00116
- Amiriparian, S., and Schuller, B. (2021). “AI hears your health: computer audition for health monitoring,” in *IC-IHAW*, Vol. 1538, eds E. Pissaloux, G. A. Papadopoulos, A. Achilleos, and R. Velázquez (Cham: Springer), 227–233. doi: 10.1007/978-3-030-94209-0\_20
- Amiriparian, S., Schuller, B. W., Asghar, N., Zen, H., and Burkhardt, F. (2023). Guest editorial: special issue on affective speech and language synthesis, generation, and conversion. *IEEE Trans. Affect. Comput.* 14, 3–5. doi: 10.1109/TAFFC.2022.3233120
- Anderson, R. C., and Klofstad, C. A. (2012). Preference for leaders with masculine voices holds in the case of feminine leadership roles. *PLoS ONE* 7, e51216. doi: 10.1371/journal.pone.0051216
- Anderson, R. C., Klofstad, C. A., Mayew, W. J., and Venkatchalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS ONE* 9, e97506. doi: 10.1371/journal.pone.0097506
- Ando, A., Asakawa, R., Masumura, R., Kamiyama, H., Kobashikawa, S., Aono, Y., et al. (2018). “Automatic question detection from acoustic and phonetic features using feature-wise pre-training,” in *Proc. of INTERSPEECH* (Hyderabad: ISCA), 1731–1735. doi: 10.21437/Interspeech.2018-1755
- Antonakis, J. (2012). “Transformational and charismatic leadership,” in *The Nature of Leadership*, 2nd ed., eds D. V. Day, and J. Antonakis (Thousand Oaks, CA: Sage Publications), 256–288. ISBN: 978-1-4129-8020-3.
- Antonakis, J., and Dalgas, O. (2009). Predicting elections: Child’s play! *Science* 323, 1183–1183. doi: 10.1126/science.1167748
- Antonakis, J., Fenley, M., and Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Acad. Manag. Learn. Educ.* 10, 374–396. doi: 10.5465/amle.2010.0012
- Antonakis, J., Fenley, M., and Liechti, S. (2012). *Learning Charisma Transform Yourself into the Person Others Want to Follow*. Technical Report 3, Academy of Management Learning and Education.
- Awamleh, R. (2011). Towards a model of charismatic non-verbal impression management. *Int. Bus. Econ. Res. J.* 2. doi: 10.19030/iber.v2i7.3818
- Awamleh, R., and Gardner, W. L. (1999). Perceptions of leader charisma and effectiveness: the effects of vision content, delivery, and organizational performance. *Leadersh. Q.* 10, 345–373. doi: 10.1016/S1048-9843(99)00022-3
- Baird, A., Amiriparian, S., and Schuller, B. (2019). “Can deep generative audio be emotional? Towards an approach for personalised emotional audio generation,” in *Proc. of IEEE MMSP* (Kuala Lumpur: IEEE), 1–5. doi: 10.1109/MMSP.2019.8901785
- Batliner, A., Hantke, S., and Schuller, B. (2022). Ethics and good practice in computational paralinguistics. *IEEE Trans. Affect. Comput.* 13, 1236–1253. doi: 10.1109/TAFFC.2020.3021015
- Batliner, A., Kießling, A., Burger, S., and Nöth, E. (1995). “Filled pauses in spontaneous speech,” in *In Proc. of ICPhS*, Vol. 3 (Stockholm), 472–475.
- Batliner, A., and Möbius, B. (2020). “Prosody in automatic speech processing,” in *The Oxford Handbook of Language Prosody*, eds C. Gussenhoven, and A. Chen (Oxford: Oxford University Press), 633–645. doi: 10.1093/oxfordhb/9780198832232.013.42
- Batliner, A., Neumann, M., Burkhardt, F., Baird, A., Meyer, S., Vu, N. T., et al. (2023). Ethical awareness in paralinguistics: a taxonomy of applications. *Int. J. Hum. Comput. Interact.* 39, 1904–1921. doi: 10.1080/10447318.2022.2140385
- Berdichevsky, D., and Neuenschwander, E. (1999). Toward an ethics of persuasive technology. *Commun. ACM* 42, 51–58. doi: 10.1145/301353.301410
- Beyan, C., Katsageorgiou, V.-M., and Murino, V. (2019). A sequential data analysis approach to detect emergent leaders in small groups. *IEEE Trans. Multimedia* 21, 2107–2116. doi: 10.1109/TMM.2019.2895505
- Biancardi, B., Cafaro, A., and Pelachaud, C. (2017). “Analyzing first impressions of warmth and competence from observable nonverbal cues in expert-novice interactions,” in *Proc. of ICMI* (Glasgow: ACM), 341–349. doi: 10.1145/3136755.3136779

## Author contributions

All authors contributed equally to writing, revising, and editing the manuscript and approved the final version of the manuscript.

## Funding

This work was supported by the DFG’s Reinhart Koselleck project No. 442218748 (AUDIONOMOUS).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Biancardi, B., Wang, C., Mancini, M., Cafaro, A., Chanel, G., Pelachaud, C., et al. (2019). "A computational model for managing impressions of an embodied conversational agent in real-time," in *Proc. of ACHI* (Cambridge), 1–7. doi: 10.1109/ACII.2019.8925495
- Bietti, E. (2020). "From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy," in *Proc. of FAccT* (Barcelona), 210–219. doi: 10.1145/3351095.3372860
- Bilal, M., Chamoux, M., and Lim, A. (2017). "Gaze and filled pause detection for smooth human-robot conversations," in *Proc. of IEEE-RAS Humanoids* (Birmingham: IEEE), 297–304. doi: 10.1109/HUMANOIDS.2017.8246889
- Bono, J. E., and Judge, T. A. (2004). Personality and transformational and transactional leadership: a meta-analysis. *J. Appl. Psychol.* 89, 901–910. doi: 10.1037/0021-9010.89.5.901
- Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., et al. (2023). Audiolm: a language modeling approach to audio generation. *IEEE/ACM Trans. Audio Speech Lang. Process.* 31, 2523–2533. doi: 10.1109/TASLP.2023.3288409
- Burgoon, J. K. (1991). Relational message interpretations of touch, conversational distance, and posture. *J. Nonverbal Behav.* 15, 233–259. doi: 10.1007/BF00986924
- Burgoon, J. K., Guerrero, L. K., and Floyd, K. (2016). *Nonverbal Communication*. New York, NY: Routledge. doi: 10.4324/9781315663425
- Burgoon, J. K., Johnson, M. L., and Koch, P. T. (1998). The nature and measurement of interpersonal dominance. *Commun. Monogr.* 65, 308–335. doi: 10.1080/03637759809376456
- Burgoon, J. K., Manusov, V., and Guerrero, L. K. (2022). *Nonverbal Communication*. New York, NY: Routledge. doi: 10.4324/9781003095552
- Burke, K. L., and Brinkerhoff, M. B. (1981). Capturing charisma: Notes on an elusive concept. *J. Sci. Study Relig.* 20, 274. doi: 10.2307/1385549
- Charfuelan, M., and Schröder, M. (2011). "Investigating the prosody and voice quality of social signals in scenario meetings," in *Proc. of ACHI* (Memphis, TN: Springer), 46–56. doi: 10.1007/978-3-642-24600-5\_8
- Cherulnik, P. D., Donley, K. A., Wiewel, T. S. R., and Miller, S. R. (2001). Charisma is contagious: the effect of leaders' charisma on observers' affect. *J. Appl. Soc. Psychol.* 31, 2149–2159. doi: 10.1111/j.1559-1816.2001.tb00167.x
- Christ, L., Amiriparian, S., Baird, A., Tzirakis, P., Kathan, A., Müller, N., et al. (2022a). "The muse 2022 multimodal sentiment analysis challenge: humor, emotional reactions, and stress," in *Proc. of MuSe* (Lisbon), 5–14. doi: 10.1145/3551876.3554817
- Christ, L., Amiriparian, S., Milling, M., Aslan, I., and Schuller, B. W. (2022b). Automatic emotion modelling in written stories. *arXiv*. [preprint]. doi: 10.48550/arXiv.2212.11382
- Colman, A. M. (1982). "Game theory and experimental games: the study of strategic interaction," in *International Series in Experimental Social Psychology*, Vol. 4, 1st ed., ed. M. Argyle (Pergamon: Legacy), 301. ISBN: 9780080260709.
- Conger, J. A., Kanungo, R. N., Menon, S. T., and Mathur, P. (2009). Measuring charisma: dimensionality and validity of the conger-kanungo scale of charismatic leadership. *Can. J. Adm. Sci.* 14, 290–301. doi: 10.1111/j.1936-4490.1997.tb00136.x
- Cuddy, A. J., Fiske, S. T., and Glick, P. (2008). Warmth and competence as universal dimensions of social perception: the stereotype content model and the BIAS map. *Adv. Exp. Soc. Psychol.* 40, 61–149. doi: 10.1016/S0065-2601(07)00002-0
- Darling, K. (2017). "Who's Johnny?" anthropomorphic framing in human-robot interaction, integration, and policy," in *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, eds P. Lin, K. Abney, and R. Jenkins (Oxford: Oxford University Press), 1–22. doi: 10.1093/oso/9780190652951.003.0012
- Davidson, L. (2021). The versatility of creaky phonation: segmental, prosodic, and sociolinguistic uses in the world's languages. *WIREs Cogn. Sci.* 12, e1547. doi: 10.1002/wcs.1547
- D'Errico, F., Signorello, R., Demolin, D., and Poggi, I. (2013). "The perception of charisma from voice: a cross-cultural study," in *Proc. of ACHI* (Geneva), 552–557. doi: 10.1109/ACII.2013.97
- Erez, A., Misangyi, V. F., Johnson, D. E., LePine, M. A., and Halverson, K. C. (2008). Stirring the hearts of followers: charismatic leadership as the transfer of affect. *J. Appl. Psychol.* 93, 602–616. doi: 10.1037/0021-9010.93.3.602
- Ernst, B. A., Banks, G. C., Loignon, A. C., Frear, K. A., Williams, C. E., Arciniega, L. M., et al. (2022). Virtual charismatic leadership and signaling theory: a prospective meta-analysis in five countries. *Leadersh. Q.* 33, 101541. doi: 10.1016/j.leaqua.2021.101541
- Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., et al. (2016). The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans. Affect. Comput.* 7, 190–202. doi: 10.1109/TAFFC.2015.2457417
- Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., et al. (2005). Development and validation of the political skill inventory. *J. Manage.* 31, 126–152. doi: 10.1177/0149206304271386
- Fiske, S. T., Cuddy, A. J., and Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005
- Fitzgerald, H. (2002). *How Different Are We? Spoken Discourse in Intercultural Communication. Multilingual Matters*. Clevedon. doi: 10.21832/9781853596216
- Fox Cabane, O. (2013). *The Charisma Myth: How Anyone Can Master the Art and Science of Personal Magnetism*, Vol. 1. Edmonton, AB: Portfolio Trade. ISBN: 9781591845942.
- Fragouli, E. (2018). "The dark-side of charisma and charismatic leadership," in *Proc. of ABRM ROGE*, Vol. 9 (Oxford), 298–307.
- Fraser, K. C., Kiritchenko, S., and Nejadgholi, I. (2022). Computational modeling of stereotype content in text. *Front. Artif. Intell.* 5, 826207. doi: 10.3389/frai.2022.826207
- Fraser, K. C., Nejadgholi, I., and Kiritchenko, S. (2021). Understanding and countering stereotypes: a computational approach to the stereotype content model. *arXiv*. [preprint]. doi: 10.48550/arXiv.2106.02596
- Freeth, M., and Bugembe, P. (2019). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism* 23, 503–513. doi: 10.1177/1362361318756786
- Ga, W. (2016). The making of a charismatic killer. *Int. J. Forensic Sci.* 1, 1–21. doi: 10.23880/IJFSC-16000108
- Ghorbani, S., Ferstl, Y., Holden, D., Troje, N. F., and Carbonneau, M.-A. (2023). ZeroEGGS: zero-shot example-based gesture generation from speech. *Comput. Graphics Forum* 42, 206–216. doi: 10.1111/cgf.14734
- Greer, S. D. F., and Winters, S. J. (2015). "The perception of coolness: differences in evaluating voice quality in male and female speakers," in *Proc. of ICPhS* (Glasgow), 1–102.
- Grèzes, F., Richards, J., and Rosenberg, A. (2013). "Let me finish: automatic conflict detection using speaker overlap," in *Proc. of INTERSPEECH* (Lyon: ISCA), 200–204.
- Guerini, M., and Stock, O. (2005). "Toward ethical persuasive agents," in *Proc. of CMNA* (Edinburgh), 1–5.
- Guerini, M., Stock, O., and Zancanaro, M. (2003). "Persuasion models for intelligent interfaces," in *Proc. of CMNA* (Acapulco), 1–6.
- Hamilton, M. A., and Stewart, B. L. (1993). Extending an information processing model of language intensity effects. *Commun. Q.* 41, 231–246. doi: 10.1080/01463379309369882
- Harrigan, J. A., Oxman, T. E., and Rosenthal, R. (1985). Rapport expressed through nonverbal behavior. *J. Nonverbal Behav.* 9, 95–110. doi: 10.1007/BF00987141
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X
- Hertenstein, M. J., Holmes, R., McCullough, M., and Keltner, D. (2009). The communication of emotion via touch. *Emotion* 9, 566–573. doi: 10.1037/a0016108
- Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., and Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion* 6, 528–533. doi: 10.1037/1528-3542.6.3.528
- Hilton, K. (2016). "The perception of overlapping speech: effects of speaker prosody and listener attitudes," in *Proc. of INTERSPEECH* (San Francisco, CA: ISCA), 1260–1264. doi: 10.21437/Interspeech.2016-1456
- Holladay, S. J., and Coombs, W. T. (1993). Communicating visions: an exploration of the role of delivery in the creation of leader charisma. *Manag. Commun. Q.* 6, 405–427. doi: 10.1177/0893318993006004003
- Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., et al. (2018). The lancet psychiatry commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry* 5, 237–286. doi: 10.1016/S2215-0366(17)30513-8
- Holz, N., Larrouy-Maestri, P., and Poeppel, D. (2021). The paradoxical role of emotional intensity in the perception of vocal affect. *Sci. Rep.* 11, 9663. doi: 10.1038/s41598-021-88431-0
- Horstmann, A. C., and Krämer, N. C. (2022). The fundamental attribution error in human-robot interaction: an experimental investigation on attributing responsibility to a social robot for its pre-programmed behavior. *Int. J. Soc. Robot.* 14, 1137–1153. doi: 10.1007/s12369-021-00856-9
- House, R. J. (1976). *A 1976 Theory of Charismatic Leadership*. Working paper series 76–06. Technical report. ERIC.
- Huzaiyah Bin Md Shahrin, M., and Wyse, L. (2020). Applying visual domain style transfer and texture synthesis techniques to audio: insights and challenges. *Neural Comput. Appl.* 32, 1051–1065. doi: 10.1007/s00521-019-04053-8
- Iqbal, T., Kong, Q., Plumbley, M. D., and Wang, W. (2018). "General-purpose audio tagging from noisy labels using convolutional neural networks," in *Proc. of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE)* (Surrey: Tampere University of Technology), 212–216.
- Isaac, A. M. C., and Bridewell, W. (2017). "White lies on silver tongues: why robots need to deceive (and how)," in *Robot Ethics 2.0: From Autonomous Cars to Artificial*

- Intelligence, eds P. Lin, K. Abney, and R. Jenkins (Oxford: Oxford University Press), 157–172. doi: 10.1093/oso/9780190652951.003.0011
- Jones, S. E., and LeBaron, C. D. (2002). Research on the relationship between verbal and nonverbal communication: emerging integrations. *J. Commun.* 52, 499–521. doi: 10.1111/j.1460-2466.2002.tb02559.x
- Kathan, A., Amiriparian, S., Christ, L., Triantafyllopoulos, A., Müller, N., König, A., et al. (2022). “A personalised approach to audiovisual humour recognition and its individual-level fairness,” in *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge* (Lisbon), 29–36. doi: 10.1145/3551876.3554800
- Keating, C. F. (2011). Channelling charisma through face and body status cues. in *Social Psychological Dynamics*, eds D. Chadee and A. Kostic (Jamaica: University of the West Indies Press), 93–111.
- Keating, C. F., Adjei Boateng, F., Loiacono, H., Sherwood, W., Atwater, K., Hutchison, J., et al. (2022). Charismatic nonverbal displays by leaders signal receptivity and formidability, and tap approach and avoidance motivational systems. *Front. Psychol.* 11, 526288. doi: 10.3389/fpsyg.2020.526288
- Keller, C., Siegrist, M., Earle, T. C., and Gutscher, H. (2011). The general confidence scale: coping with environmental uncertainty and threat. *J. Appl. Soc. Psychol.* 41, 2200–2229. doi: 10.1111/j.1559-1816.2011.00811.x
- Kelly, M., Svrcek, C., King, N., Scherpbier, A., and Dornan, T. (2020). Embodying empathy: a phenomenological study of physician touch. *Med. Educ.* 54, 400–407. doi: 10.1111/medu.14040
- Kirkland, A., Gustafson, J., and Székely, É. (2023). “Pardon my disfluency: the impact of disfluency effects on the perception of speaker competence and confidence,” in *Proc. of INTERSPEECH* (Dublin: ISCA), 5217–5221. doi: 10.21437/Interspeech.2023-887
- Klofstad, C. A., Anderson, R. C., and Nowicki, S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. *PLoS ONE* 10, e0133779. doi: 10.1371/journal.pone.0133779
- Klofstad, C. A., Nowicki, S., and Anderson, R. C. (2016). How voice pitch influences our choice of leaders. *Am. Sci.* 104, 282. doi: 10.1511/2016.122.282
- Knight, H., and Simmons, R. (2013). “Estimating human interest and attention via gaze analysis,” in *Proc. of the IEEE International Conference on Robotics and Automation* (Karlsruhe: IEEE), 4350–4355. doi: 10.1109/ICRA.2013.6631193
- Koppensteiner, M., Stephan, P., and Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness, and competence in body motion. *Pers. Individ. Dif.* 94, 101–106. doi: 10.1016/j.paid.2016.01.013
- Korta, K., and Perry, J. (2020). “Pragmatics,” in *The Stanford Encyclopedia of Philosophy*, ed. E. N. (Stanford, CA: Zalta Metaphysics Research Lab, Stanford University).
- Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., et al. (2016). “Eye tracking for everyone,” in *Proc. of CVPR* (Las Vegas, NV: IEEE), 2176–2184. doi: 10.1109/CVPR.2016.239
- Krämer, N., and Manzeschke, A. (2021). “Social reactions to socially interactive agents and their ethical implications,” in *Handbook on Socially Interactive Agents – 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 1: Methods, Behavior, Cognition*, eds B. Lugrin, C. Pelachaud, and D. Traum (New York, NY: ACM books), 77–103. doi: 10.1145/3477322.3477324
- Latif, S., Cuayahuitl, H., Pervez, F., Shamshad, F., Ali, H. S., and Cambria, E. (2023). A survey on deep reinforcement learning for audio-based applications. *Artif. Intell. Rev.* 56, 2193–2240. doi: 10.1007/s10462-022-10224-2
- Lennon, S. J. (1990). Effects of clothing attractiveness on perceptions. *Home Econ. Res. J.* 18, 303–310. doi: 10.1177/1077727X9001800403
- Levy, C. (2010). Charismatic leadership in resistance to change. *Leadersh. Q.* 21, 127–143. doi: 10.1016/j.leaqua.2009.10.010
- Levinson, S. C. (1983). *Pragmatics*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511813313
- Li, S., and Deng, W. (2022). Deep facial expression recognition: a survey. *IEEE Trans. Affect. Comput.* 13, 1195–1215. doi: 10.1109/TAFFC.2020.2981446
- Li, Y., Zeng, J., Shan, S., and Chen, X. (2019). Occlusion aware facial expression recognition using CNN with attention mechanism. *IEEE Trans. Image Process.* 28, 2439–2450. doi: 10.1109/TIP.2018.2886767
- Liang, H., Yang, L., Cheng, H., Tu, W., and Xu, M. (2017). “Human-in-the-loop reinforcement learning,” in *Proc. of CAC* (Jinan: IEEE), 4511–4518. doi: 10.1109/CAC.2017.8243575
- Liu, R., Sisman, B., and Li, H. (2021). “Reinforcement learning for emotional text-to-speech synthesis with improved emotion discriminability,” in *Proc. of INTERSPEECH* (Brno), 4648–4652. doi: 10.21437/Interspeech.2021-1236
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). Roberta: a robustly optimized bert pretraining approach. *arXiv*. [preprint]. doi: 10.48550/arXiv.1907.11692
- Lugrin, B., and Rehm, M. (2021). “Culture for socially interactive agents,” in *Handbook on Socially Interactive Agents – 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 1: Methods, Behavior, Cognition*, eds I. B. Lugrin, C. Pelachaud, and D. Traum (New York, NY: ACM), 463–493. doi: 10.1145/3477322
- Lukes, S. (2019). “Trump’s charisma,” in *Antidemocracy in America: Truth, Power, and the Republic at Risk*, eds E. Klinenberg, S. Marcus, and C. Zalom (New York, NY: Columbia University Press), 199–206. doi: 10.7312/klin19010-024
- Luzardo, G., Guamán, B., Chiluiza, K., Castells, J., and Ochoa, X. (2014). “Estimation of presentations skills based on slides and audio features,” in *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (Istanbul), 37–44. doi: 10.1145/2666633.2666639
- Manzelli, R., Thakkar, V., Siahkamari, A., and Kulis, B. (2018). “Conditioning deep generative raw audio models for structured automatic music,” in *Proc. of the 19th International Society for Music Information Retrieval Conference* (Paris), 1–8.
- Maran, T., Furtner, M., Liegl, S., Kraus, S., and Sachse, P. (2019). In the eye of a leader: eye-directed gazing shapes perceptions of leaders’ charisma. *Leadersh. Q.* 30, 101337. doi: 10.1016/j.leaqua.2019.101337
- Maran, T., Liegl, S., Moder, S., Kraus, S., and Furtner, M. (2021). Clothes make the leader! How leaders can use attire to impact followers’ perceptions of charisma and approval. *J. Bus. Res.* 124, 86–99. doi: 10.1016/j.jbusres.2020.11.026
- McCrae, R. R., Costa, P. T., John, O. P., Robins, R. W., and Pervin, L. A. (1999). *Handbook of Personality: Theory and Research*. New Haven County, CT: Guilford.
- Mehrabian, A., and Wiener, M. (1967). Decoding of inconsistent communications. *J. Pers. Soc. Psychol.* 6, 109–114. doi: 10.1037/h0024532
- Meyer, E. (2014). *The Culture Map: Breaking through the Invisible Boundaries of Global Business*. New York, NY: PublicAffairs.
- Mhatre, K. H., and Riggio, R. E. (2014). *Charismatic and Transformational Leadership*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199755615.013.012
- Michalsky, J., Niebuhr, O., and Penke, L. (2020). “Do charismatic people produce charismatic speech? On the relationship between the big five personality traits and prosodic features of speaker charisma in female speakers,” in *Proc. of Speech Prosody* (Tokyo), 700–704. doi: 10.21437/SpeechProsody.2020-143
- Minaei, S., Minaei, M., and Abdolrashidi, A. (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors* 21, 3046. doi: 10.3390/s21093046
- Mohammad, S. M. (2022). Ethics sheet for automatic emotion ecognition and sentiment analysis. *Comput. Linguist.* 48, 239–278. doi: 10.1162/coli\_a\_00433
- Montemayor, C., Halpern, J., and Fairweather, A. (2022). In principle obstacles for empathic AI: why we can’t replace human empathy in healthcare. *AI Soc.* 37, 1353–1359. doi: 10.1007/s00146-021-01230-z
- Mori, M., MacDorman, K., and Kageki, N. (2012). The uncanny valley. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811
- Moskowitz, G. B., and Gill, M. J. (2013). “Person perception,” in *The Oxford Handbook of Cognitive Psychology*, ed. D. Reisberg (Oxford: Oxford University Press), 918–942. doi: 10.1093/oxfordhb/9780195376746.013.0058
- Nasir, M., Baucom, B., Bryan, C., Narayanan, S., and Georgiou, P. (2022). Modeling vocal entrainment in conversational speech using deep unsupervised learning. *IEEE Trans. Affect. Comput.* 13, 1651–1663. doi: 10.1109/TAFFC.2020.3024972
- Niebuhr, O., Neitsch, J., and Michalsky, J. (2020). Akustisches charisma profiling – Auf dem Weg zur digitalen Rhetorik. *Akustik J.* 02, 7–22.
- Niebuhr, O., Voße, J., and Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of steve jobs tone of voice. *Comput. Human Behav.* 64, 366–382. doi: 10.1016/j.chb.2016.06.059
- Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., and Castells, J. (2018). “The RAP system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors,” in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (Sydney, NSW: ACM), 360–364. doi: 10.1145/3170358.3170406
- Okada, S., Nguyen, L. S., Aran, O., and Gatica-Perez, D. (2019). Modeling dyadic and group impressions with intermodal and interperson features. *ACM Trans. Multimedia Comput. Commun. Appl.* 15(1s), 1–30. doi: 10.1145/3265754
- Pamungkas, E. W. (2019). Emotionally-aware chatbots: a survey. *arXiv*. [preprint]. doi: 10.48550/arXiv.1906.09774
- Pasquale, C. D., Cullen, C., and Vaughan, B. (2019). “An investigation of therapeutic rapport through prosody in brief psychodynamic psychotherapy,” in *Proc. of INTERSPEECH* (Graz: ISCA), 3043–3047. doi: 10.21437/Interspeech.2019-2551
- Pauser, S., Wagner, U., and Ebster, C. (2018). An investigation of salespeople’s nonverbal behaviors and their effect on charismatic appearance

- and favorable consumer responses. *J. Pers. Sell. Sales Manag.* 38, 344–369. doi: 10.1080/08853134.2018.1480383
- Pearsell, S., and Pape, D. (2023). The effects of different voice qualities on the perceived personality of a speaker. *Front. Commun.* 7, 909427. doi: 10.3389/fcomm.2022.909427
- Pentland, A. (2010). To signal is human: real-time data mining unmasks the power of imitation, kith and charisma in our face-to-face social networks. *Am. Sci.*, 98, 204–211. doi: 10.1511/2010.84.203
- Prehn-Kristensen, A., Wiesner, C., Bergmann, T. O., Wolff, S., Jansen, O., Mehdorn, H. M., et al. (2009). Induction of empathy by the smell of anxiety. *PLoS ONE* 4, 1–9. doi: 10.1371/journal.pone.0005987
- Raven, J. C., and Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales*, Vol. 759. Oxford: Oxford Psychologists Press.
- Rességuier, A., and Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data Soc.* 7, 205395172094254. doi: 10.1177/2053951720942541
- Revina, I., and Emmanuel, W. S. (2021). A survey on human face expression recognition techniques. *J. King Saud Univ. Comput. Inform. Sci.* 33, 619–628. doi: 10.1016/j.jksuci.2018.09.002
- Rogers, S. L., Speelman, C. P., Guidetti, O., and Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-22726-7
- Rosenberg, A., and Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Commun.* 51, 640–655. doi: 10.1016/j.specom.2008.11.001
- Rothenpieler, D., and Amiriparian, S. (2023). "METEOR guided divergence for video captioning," in *Proc. of IJCNN* (Goldcoast: IEEE), 1–7. doi: 10.1109/IJCNN54540.2023.10191075
- Saleh, A., Jaques, N., Ghandeharioun, A., Shen, J., and Picard, R. (2020). "Hierarchical reinforcement learning for open-domain dialog," in *Proc. of the AAAI Conference on Artificial Intelligence*, Vol. 34 (New York, NY), 8741–8748. doi: 10.1609/aaai.v34i05.6400
- Sapru, A., and Bourlard, H. (2015). Automatic recognition of emergent social roles in small group interactions. *IEEE Trans. Multimedia* 17, 746–760. doi: 10.1109/TMM.2015.2408437
- Schuller, B., and Batliner, A. (2014). *Computational Paralinguistics – Emotion, Affect, and Personality in Speech and Language Processing*. Chichester: Wiley. doi: 10.1002/9781118706664
- Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., et al. (2015). A survey on perceived speaker traits: personality, likability, pathology, and the first challenge. *Comput. Speech Lang.* 29, 100–131. doi: 10.1016/j.csl.2014.08.003
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., et al. (1998). Development and validation of a measure of emotional intelligence. *Pers. Individ. Dif.* 25, 167–177. doi: 10.1016/S0191-8869(98)00001-4
- Sharma, A., Miner, A. S., Atkins, D. C., and Althoff, T. (2020). A computational approach to understanding empathy expressed in text-based mental health support. *arXiv*. [preprint]. doi: 10.48550/arXiv.2009.08441
- Tang, Y., Huang, Y., Wu, Z., Meng, H., Xu, M., Cai, L., et al. (2016). "Question detection from acoustic features using recurrent neural network with gated recurrent unit," in *Proc. of ICASSP* (Shanghai: IEEE), 6125–6129. doi: 10.1109/ICASSP.2016.7472854
- Tickle-Degnen, L., and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychol. Inq.* 1, 285–293. doi: 10.1207/s15327965pli0104\_1
- Towheed, S. (2022). "An examination of bookshelves in the age of the COVID-19 pandemic as a "liminal space,"" in *New Directions in Book History*, eds. C. Norrick-Rühl and S. Towheed (Cham: Springer International Publishing), 31–47. doi: 10.1007/978-3-031-05292-7\_2
- Towler, A. J. (2003). Effects of charismatic influence training on attitudes, behavior, and performance. *Pers. Psychol.* 56, 363–381. doi: 10.1111/j.1744-6570.2003.tb00154.x
- Trompenaars, F., and Hampden-Turner, C. (1998). *Riding the Waves of Culture: Understanding Diversity in Global Business*, 2nd ed. New York, NY: McGraw-Hill Companies, Incorporated.
- Trouvain, J., Weiss, B., and Barkat-Defradas, M. (2021). "Voice attractiveness: concepts, methods, and data," in *Voice Attractiveness – Studies on Sexy, Likable, and Charismatic Speakers*, eds B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Singapore: Springer), 3–16. doi: 10.1007/978-981-15-6627-1\_1
- Trouvain, J. and Werner, R. (2022). "A phonetic view on annotating speech pauses and pause-internal phonetic particles," in *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion*, eds. C. Schwarze and S. Grawunder (Tübingen: Narr), 55–73.
- Tskhay, K. O., Zhu, R., Zou, C., and Rule, N. O. (2018). Charisma in everyday life: conceptualization and validation of the general charisma inventory. *J. Pers. Soc. Psychol.* 114, 131–152. doi: 10.1037/pspp0000159
- Tubbs, R. S. (2019). "Learning charisma," in *A Guide to the Scientific Career: Virtues, Communication, Research and Academic Writing*, eds. M. Shoja, A. Arynchyna, M. Loukas, A. V. D'Antoni, S. M. Buerger, M. Karl, et al. (Wiley Online Library), 103–107. doi: 10.1002/9781118907283.ch13
- Van Erp, J. B. F., and Toet, A. (2015). Social touch in human–computer interaction. *Front. Digit. Humanit.* 2, 2. doi: 10.3389/fdigh.2015.00002
- Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., and Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Sci. Rep.* 11, 1–16. doi: 10.1038/s41598-021-81987-x
- Vinciarelli, A., and Mohammadi, G. (2014). A survey of personality computing. *IEEE Trans. Affect. Comput.* 5, 273–291. doi: 10.1109/TAFFC.2014.2330816
- Voße, J., Niebuhr, O., and Wagner, P. (2022). How to motivate with speech. findings from acoustic phonetics and pragmatics. *Front. Commun.* 7, 910745. doi: 10.3389/fcomm.2022.910745
- Wagner, B. (2018). "Ethics as an escape from regulation: From ethics-washing to ethics shopping?" in *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, eds E. Bayamlioglu, M. Hildebrandt, I. Baraluic, and L. Janssen (Amsterdam: Amsterdam University Press), 84–89. doi: 10.2307/j.ctvhrd092.18
- Wang, C., and Chanel, G. (2021). "An open dataset for impression recognition from multimodal bodily responses," in *Proc. of ACII* (Nara), 1–8. doi: 10.1109/ACII52823.2021.9597421
- Wang, K., Peng, X., Yang, J., Lu, S., and Qiao, Y. (2020). "Suppressing uncertainties for large-scale facial expression recognition," in *Proc. of CVPR* (Seattle, WA), 6896–6905. doi: 10.1109/CVPR42600.2020.00693
- Weber, M. (1922). *Grundriss Der Sozialökonomie. III. Abteilung. Wirtschaft Und Gesellschaft. Bearbeitet von Max Weber*. Tübingen: J.C.B. Mohr.
- Weiss, B., and Burkhardt, F. (2010). "Voice attributes affecting likability perception," in *Proc. of INTERSPEECH* (Makuhari: ISCA), 2014–2017. doi: 10.21437/Interspeech.2010-570
- Weiss, B., Trouvain, J., and Burkhardt, F. (2021). "Acoustic correlates of likable speakers in the NSC database," in *Voice Attractiveness – Studies on Sexy, Likable, and Charismatic Speakers*, eds B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Singapore: Springer Singapore), 245–262. doi: 10.1007/978-981-15-6627-1\_13
- Weiss, S. J., Wilson, P., Hertenstein, M. J., and Campos, R. (2000). The tactile context of a mother's caregiving: Implications for attachment of low birth weight infants. *Infant Behav. Dev.* 23, 91–111. doi: 10.1016/S0163-6383(00)00030-8
- Weiste, E., and Peräkylä, A. (2014). Prosody and empathic communication in psychotherapy interaction. *Psychother. Res.* 24, 687–701. doi: 10.1080/10503307.2013.879619
- Welsh, E.-C. O., and Lenzenweger, M. F. (2021). Psychopathy, charisma, and success: a moderation modeling approach to successful psychopathy. *J. Res. Pers.* 95, 104146. doi: 10.1016/j.jrp.2021.104146
- Weninger, F., Krajewski, J., Batliner, A., and Schuller, B. (2012). The voice of leadership: Models and performances of automatic analysis in on-line speeches. *IEEE Trans. Affect. Comput.* 3, 496–508. doi: 10.1109/T-AFFC.2012.15
- Wolf, M. J., Miller, K., and Grodzinsky, F. S. (2017). Why we should have seen that coming: Comments on microsoft's 'tay' experiment, and wider implications. *ACM SIGCAS Comput. Soc.* 47, 54–64. doi: 10.1145/3144592.3144598
- Won, J., Gopinath, D., and Hodgins, J. (2021). Control strategies for physically simulated characters performing two-player competitive sports. *ACM Trans. Graph.* 40, 1–11. doi: 10.1145/3450626.3459761
- Wu, S., Zhou, L., Hu, Z., and Liu, J. (2022). Hierarchical context-based emotion recognition with scene graphs. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15. doi: 10.1109/TNNLS.2022.3196831
- Xiao, B., Georgiou, P. G., Imel, Z. E., Atkins, D. C., and Narayanan, S. (2013). "Modeling therapist empathy and vocal entrainment in drug addiction counseling," in *Proc. of INTERSPEECH* (Lyon: ISCA), 2861–2865. doi: 10.21437/Interspeech.2013-251
- Xu, R., Wu, R., Ishiwaka, Y., Vondrick, C., and Zheng, C. (2020). Listening to sounds of silence for speech denoising. *Adv. Neural Inf. Process. Syst.* 33, 9633–9648.
- Yang, Z., Huynh, J., Tabata, R., Cestero, N., Aharoni, T., Hirschberg, J., et al. (2020). "What makes a speaker charismatic? Producing and perceiving charismatic speech," in *Proc. of Speech Prosody* (Tokyo), 685–689. doi: 10.21437/SpeechProsody.2020-140
- Zdarsky, N., Treue, S., and Esghaei, M. (2021). A deep learning-based approach to video-based eye tracking for human psychophysics. *Front. Hum. Neurosci.* 15, 685830. doi: 10.3389/fnhum.2021.685830
- Zhang, Y.-J., Pan, S., He, L., and Ling, Z.-H. (2019). Learning latent representations for style control and transfer in end-to-end speech synthesis. In *Proc. of ICASSP*, pages 6945–6949, Brighton, UK. IEEE. doi: 10.1109/ICASSP.2019.8683623
- Zoghaib, A. (2019). Persuasion of voices: the effects of a speaker's voice characteristics and gender on consumers' responses. *Rech. Appl. Mark.* 34, 83–110. doi: 10.1177/2051570719828687