Theory and Society (2026) 55:31
https://doi.org/10.1007/s11186-026-09696-w

The "Good bad theory" case in emotion analytics: AI's potential and limits for social theory

Andrey V. Rezaev[1,2] · **Natalia D. Tregubova**[3]

Received: 1 October 2025 / Accepted: 3 March 2026

Abstract
The paper introduces "good bad theory" as a conceptual framework for understanding artificial intelligence's potential in social theory development, focusing specifically on the formalization of human emotions through mathematical models and algorithms. It defines 'good bad theories' as theoretical constructions that are sound ('good') in mathematical accuracy and computational performance, while failing ('bad') to capture essential phenomenological and contextual aspects of human behavior in society. These theories are 'good' from a technical standpoint - precise, testable, and algorithmically implementable, yet 'bad' from a social sciences perspective because they reduce the complexity of human social life to simplistic mathematical relationships. The paper exercises both quantitative and qualitative analysis. Through a quantitative review of publications – papers in top journals and conference proceedings – in sociology and computer science, the authors identify patterns of interest in emotion analysis across both disciplines. The paper employs comparative case study methodology to examine two prominent examples of "good bad theories" from different disciplinary domains. The first is Randall Collins' interaction ritual theory, which formalizes the emotional dynamics of social interaction in the 'emotional utilitarianism' framework. The second is Stuart Russell's game-theoretic approach to human-AI interaction. The authors argue that the "good bad theory" problem in emotion modeling represents a structural feature of AI applications in social science, where mathematical formalization requirements inherently conflict with the phenomenological, contextual, and interpretive dimensions of human emotional experience that social theory recognizes as central to understanding social reality.

Keywords Artificial intelligence · Social theory · Sociology of emotions · Computational social science · Formalization · Phenomenology · Emotion analytics

Extended author information available on the last page of the article

Published online: 25 March 2026

Introduction

The integration of artificial intelligence (AI) technologies into the everyday life of society has created both unprecedented opportunities and challenges for social theory development, especially when it comes to the study of human emotions. As AI systems become increasingly sophisticated in modeling human behavior, fundamental questions emerge about the adequacy of mathematical formalization for understanding the complexity of human emotional experience. The research focus of the paper is the emergence of what we call "good bad theories", that is, theoretical constructions that achieve computational sophistication and empirical use but at the same time sacrifice sociological adequacy. This is not just about theories that are partially accurate or have compromises. It shows a pattern, that is, to make theories easier to compute and make calculations, often takes away what is meaningful in social and personal life.

Our study focuses on emotion analytics as a particularly informative case of this phenomenon. Human emotions occupy a central place in social theory, representing phenomena that are simultaneously personal and social, cognitive and embodied, universal and culturally specific. The attempt to formalize emotional dynamics through mathematical models thus provides an ideal lens to examine broader questions about AI's role in social theory development.

Our analysis operates at two temporal levels. First, we examine current trends in emotion research and show how formal models and methods are already shaping theories. Second, we offer a preliminary critical analysis of a trend where AI technologies prioritize formalized approaches to emotion over understanding real emotional experience. By focusing on both current patterns and future risks, we explore how the needs of computers are changing the way we think about emotions.

Here is how this paper is structured. We begin by defining the term "good bad theory" and justifying its usage in relation to emotion analysis in the age of AI expansion. We then briefly characterize traditions of emotion research in sociology and computer science, highlighting key trends and challenges. We then proceed to develop a thesis on "good bad theories" based on an analysis of the current professional literature. The analysis will involve two stages. The initial phase, a quantitative study, provides a content analysis of articles from top sociology and computer science journals and conference proceedings, highlighting and comparing recent trends in emotion analysis. The second, qualitative stage presents a comparative case study of two "good bad theories" - Randall Collins' interaction ritual theory for sociology and Stuart Russell's theory of human-compatible AI for computer science. We analyze their conceptual frameworks and implementations, then formulate a critique of their arguments.

After that, we present an argument about "good bad theories" of emotions in contemporary society, which, as we argue, take the form of "good bad theory" of a specific kind - "allegedly good/misleading theories". We conclude the paper with a summary of its findings in a series of arguments and counterarguments and prospects for future research.

Theoretical framework: Defining "Good bad theory"

Before we begin to outline our conceptual understanding of what is 'good bad theory' and how it affects the development of theoretical creation in the social sciences in the era of AI, a word needs to be said on what is the 'theory' per se.

It won't be big news to say that disagreements among sociologists exist over the very meaning of 'theory' as a theoretical construct[1]. In this study, we use "theory" as a concept in two basic and one auxiliary meaning. On the one hand, for sociological research, theory' is a concept aimed to explain a particular social phenomenon. On the other hand, 'theory' in sociological theoretical reconstruction is a logically connected system of general propositions, which establishes a relationship between variables. Moreover, our analysis is grounded in a broader understanding of theory as "an overall perspective from which one sees and interprets the world" (Abend, 2008: 179). The point here is that the investigative properties of a good theory depend not only on the specific research question and research design, but also on its ability to grasp important properties of its basic object. We will see further that "good bad theories" are insufficient, even if they are useful for some research inquiries.

The question of which theory is best suited to empirical sociological research has been debated at least since *Methodenstreit*. Today, as it was a century ago, sociological positions are split between positivist-minded and interpretivist-minded thinkers. The former argue that a good theory should be able to make predictions and be consistent with other scholarly disciplines (Whitmeyer & Hopcroft, 2025), as well as aim at generalized explanations instead of focusing on specific cases (Brint, 2025; see also: Healy, 2017; Ermakoff, 2017). The opposite position is that sociological theorizing should start with the particular and the provincial (Sweet, 2023; see also: Mears, 2017). These discussions on the essence of good theories for empirical sociological investigations demonstrate that the understanding of what is 'a good theory' depends not only on paradigmatic preferences, not only on possibilities and limitations of data collection and analysis, but also on the idea of what sociology *is* as a science.

New generative instruments and new technological advances started to determine new debates on the reality of theoretical construction in social research.

In fact, the use of AI tools in sociological theorizing and social research has already been under scrutiny for several decades now (Collins, 1994; Bainbridge et al., 1994). Today, it receives special attention from both computer scientists and social scientists. For instance, Gert Jan Hofstede and colleagues argue for a new interdisciplinary research area of 'artificial sociality' for the creation and testing of computational models of the essentials of human social behavior (Hofstede et al., 2021). However, the issue of how to use AI in sociological theorizing is problematic (Rezaev & Tregubova, 2025). Some scholars support it as a way to develop and test theories to gain conceptual clarity, logical consistency, and testability of sociological theory (Shults, 2025). Others, while finding it promising, point to the problems of semanticisation, transferability, and generativity (Mökander & Schroeder, 2022).

[1] Abend (2008) identifies seven approaches to defining "theory". Some of them can be combined, and some refute each other.

In this article, we aim to contribute to this discussion, highlighting the advantages and limitations of using AI tools for theory-building and theory-testing in sociology. Our focus is on a specific area of research in social sciences and computer sciences, which is the analysis of human emotions. To reach our goal, we introduce, substantiate, and illustrate the concept of a "good bad theory".[2]

Stephen Turner presented, in a quite different sense, the term "good bad theory" in his book *Explaining the Normative* (2010). He uses the term to characterize common sense ideas for explaining human behavior in a particular culture: "When we live in a society, we use a common set of ideas that enables coordination, assessing blame, and all sorts of other activities… Call these Good Bad Theories: they are good for the myriad purposes of coordination they serve, bad as science or explanation" (Turner, 2013: 193).

Our characteristic of "good bad theory" resembles Turner's in outlining a theory that is beneficial for practical purposes and application, but not theoretically sound. However, there are two distinctions in our definition.

First, we conceptualize 'theories' as scientific statements, but not general societal ideas and premises. Second, while for Turner theories are 'bad' because they are pre-scientific (in a sense), for us they are 'bad' because they are one-sidedly scientific or 'too scientific'. In other words, they do best in formalization and calculability while ignoring the full picture of what is going on in societal practices.

Our idea of "good bad theory" stems from a problem of how to apply pure hard sciences methods to social phenomena. Basically, this concept is an ideal type (in the Weberian sense): a theory that combines a) mathematical precision, b) computational tractability, c) predictive capacity, d) empirical application, and e) systematic exclusion of meaning, interpretation, contextual sensitivity, and the complexities of human experience. Adding to these characteristics another one: f) theories that become misleading through self-fulfilling dynamics, we get a subcategory of "good bad theory" - allegedly good/misleading theory." Various theories diverge on their proximity to this ideal type based on these characteristics.

The following three questions will direct our subsequent analysis:

1. *Is there a clearly defined disciplinary distinction in the theoretical approaches to emotions between sociology and computer science?* The idea of "good bad theory" is based on the inconsistency between computational and humanity-oriented perspectives, so the first question is whether there are any differences between them.
2. *To what extent do publications in both fields exemplify "good bad theory"?* We're investigating the extent to which this type of theorizing is present in both areas.
3. *What are the key advantages and limitations of applying formalized models in human emotion analytics?* This question explores the main strengths and weaknesses of "good bad theories," detailing their critical features.

[2] The term is inspired by George Orwell's essay on Good Bad Books: URL: https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/good-bad-books/.

Sociology of emotion: Current debates and prospects for further research in the age of AI

The sociology of emotions emerged as a distinct field in the 1970s, but its origins can be traced back to sociological classics (Barbalet, 1998; Turner, 2009). Today, it is a dynamic field that comprises a set of competing and complementary theoretical approaches. It "emphasizes that virtually every encounter, every corporate unit, every institutional domain, every categoric unit in every system of stratification is driven by emotions affecting commitments to social structures and their cultures." (Turner, 2025: 542).

What are the basic tactics of theorizing emotions in sociology? There are various classifications in specialized literature. Jonathan Turner (2009) outlines seven approaches: evolutionary, symbolic interactionist, symbolic interactionist with psychoanalytic elements, interaction ritual, power and status, stratification, and exchange theories. *Springer Handbook of the Sociology of Emotions* contains an even more detailed classification of theories of emotions, including affect control theory and identity theory (Stets & Turner, 2006).

For our analysis, we take a simple, less detailed typology of approaches to studying emotions. Two questions have been determined as most productive for our purposes: What is the fundamental nature of emotions? At what level of social reality should we conduct emotional analytics? The answers to these questions will help us identify the strategic problems in understanding and modeling emotions with the use of AI instruments.

So, what is the essence of emotion? Current literature provides basically three answers to this question: emotion is (a) a bodily sense, (b) a social construct, (c) a value judgment. The division 'bodily sense vs. social construct' is outlined in *The Managed Heart* by Arlie Hochschild (1983). The scholar distinguished between two models of emotions in social theory: the 'organismic model,' which defines emotion primarily as a biological process, and the 'interactionist model,' which suggests that while emotion involves some biological component, the primary concern is the meaning constructed in interaction.

The dichotomy between understanding emotions as non-reasoning (bodily) movements and value judgments is formulated and discussed in Martha Nussbaum's *Upheavals of Thought* (2001). The author characterizes these two approaches as traditions deeply rooted in Western philosophical discussions.

However, in social research (as opposed to philosophical debates), these 'pure' positions are quite rare. Typically, scholars consider two or three aspects of emotion, with one being dominant.

Assuming that the analysis of emotions in sociology requires consideration of all three aspects, three problems arise for examining/modelling emotions with AI instruments.

The first is the modeling of bodily effects, which vary individually and depend on the characteristics of a particular organism. This is problematic to quantify and imitate[3].

[3] Here, another open question arises: are emotional expressions culturally universal, or not? There is an ongoing debate in neuropsychology regarding this topic (Ekman, 2003, Feldman Barrett, 2017). If the

The second is the difficulty of accounting for cultural context, especially in a specific situation. LLMs address this problem by training on vast amounts of data representing emotional expressions in different cultures, but errors are possible and occur from time to time.

The third, the most serious problem, is the inherent subjectivity and intentionality of emotions. An individual's worldview, his/her ambitions and needs influence the reality of emotions. Current AI tools can model this as a system of ordered preferences, but it is only an approximate representation of how real humans behave.

The level of social analysis creates another basis for the classification of theoretical approaches in the sociology of emotion. It typically results in three main groups: structural, cultural, and interactional theories.

Structural theorists, for example, Jack Barbalet (1998), theorize on how emotions and social structure influence each other. Cultural theorists, such as Eva Illouz (2007), examine the cultural patterns of emotion dynamics across different epochs and regions. Randall Collins (2004), Jeffrey Alexander (2006), and Anne Rawls (1987, 2000) are prominent interactional theorists. According to them, emotions are both a condition and an outcome of social interaction. There are scholars, such as Jonathan Turner (2002, 2025) and Arlie Hochschild (1983), who combine these approaches within a single theoretical framework.

Each group of these theories initiates additional complications in the analysis of emotions with the help of AI tools. They facilitate the formulation of specific questions regarding how social structure, culture, and the situation of interaction influence emotional dynamics.

Emotion analysis in the field of AI: Developments in affective computing

The study of emotions has been a research subject for computer scientists for several decades. Aaron Sloman and Monica Croucher (1981) were among the first to suggest that modeling emotion is an important task for computer science. The authors argued that in order to simulate human ability to make decisions in real-time in a complex environment, it is necessary to create an analog of human emotions in a robot. Marvin Minsky, one of the founding fathers of the concept of 'artificial intelligence', claimed: "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions" (Minsky, 1986: 163).

Affective Computing by Rosalind Picard (1997) was a fundamental book that helped to shape the field. Today, affective computing is a rapidly growing research field, with a dominance of computer science and, to a lesser degree, cognitive sciences. Guanxiong Pei and colleagues (Pei et al., 2024) outlined the following key topics of this field: (a) natural language processing techniques used for affective computing and opinion mining; (b) facial expression and micro-expression recognition and analysis; (c) affective computing studies in human–computer interaction;

answer to this question is *no*, then the problem of "reading" bodily characteristics for AI becomes even more complex.

(d) applied research of affective computing in affective disorder analysis; (e) multimodal sentiment analysis based on deep learning. Gustavo Assunção and colleagues (Assunção et al., 2022) characterize two areas of computer science besides affective computing that deal with the problem of emotions: emotions in reinforcement learning, and implementation of emotions through replication of human brain circuitry.

In computer science, when studying and designing emotions, the two types of research are clearly distinct. The first type deals with emotional recognition and imitation. It includes tasks such as recognizing emotions based on data from users' facial expressions, voice, text, and other information. It also tries to find effective and accurate ways to display emotions with the help of AI instruments. It deals with the uncanny valley problem (Mori et al., 2012) and correlates the user's emotional state with AI's response.

The second research approach builds on the ideas developed by Sloman, Croucher, and Minsky. Basically, the expectation is that AI agents, when interacting with complex environments, develop operational equivalents of human emotions. In this context, researchers can consider both the intentional development and the unexpected arising of 'emotions' within the realm of artificial intelligence. The appearance of "artificial liars" who mislead users for multiple purposes can serve as an example of the latter (Castelfranchi, 2000).

There are various computational models of emotions and their implementations (Marsella et al., 2010), yet existing models are very limited, particularly in their ability to accurately assess the intensity of emotions, as well as in their capacity for ethical emotion regulation (Ojha et al., 2021).

Another conceptual problem is the computational modeling of social dynamics of emotions. The mainstream approach in computer science is still to regard emotions as properties of an individual. However, there are interesting developments in the field of modeling collective and crowd emotions (Aydt et al., 2011; Bosse et al., 2013; Garcia et al., 2016; He et al., 2024), including the analysis of emotions in human-AI collaboration (Ferrada & Camarinha-Matos, 2024).

The problems of emotion analysis in computer science raise not only academic concerns, but also increasingly pragmatic concerns and fears. One of the most pressing worries in the field today is *explainable* affective computing (Cortiñas-Lorenzo & Lacey, 2023). Despite the growing body of research and applications, many affective computing systems are designed as 'black boxes' with a non-transparent link between input and output data. These systems, thus, are not entirely reliable in real-time interactions. It is obvious, especially for such domains as healthcare, education, and law.

Quantitative content analysis of publications on emotions in sociology and computer science

To compare the organization of emotion analysis within contemporary sociology and computer science, we examined scholarly publications, including journal articles (for both disciplines) and conference proceedings (exclusively for computer science).

We chose our samples to include both established trends from journals and new trends from conferences:

Table 1 The sample of journals

Sociological journals	Computer science journals
Annual Review of Sociology (ARS)	Foundations and Trends in Machine Learning (FTML)
American Sociological Review (ASR)	International Journal of Information Management (IJIM)
Sociological Methods and Research (SMR)	Science Robotics (SR)
American Journal of Sociology (AJS)	Nature Machine Intelligence (NMI)
Sociological Theory (ST)	Computers and Education: Artificial Intelligence (CEAI)

– Five top-cited journals per discipline (2020–2025) represent established scholarly discourse.
– Conference proceedings of the leading computer science association indicate emerging trends. We examined the following: (a) 100 recent ACM papers (2025 one-year retrospective sample) for category development; (b) 341 papers from September-November 2025 (three-month intensive sample) for pattern confirmation.

We shall commence with an analysis of the journal articles. We have selected the five most cited journals for each discipline according to Scimago Journal & Country Rank (by September 2025)[4]. The journals in the sample are listed in Table 1.

This study covers the period from January 2020 to September 2025. We determined the total number of articles in each journal during this timeframe, reviewed the titles and abstracts of the publications, and identified articles related to emotion analysis.[5] The sampling results journal is presented in Table 2.

The primary focus of our content analysis was on (a) the proportion of articles devoted to emotions in the total number of articles; (b) the subject and research design of the publications.

Below, the results are analyzed and presented by their respective disciplines.

Sociology

A review of sociological journals between 2020 and 2025 revealed 963 publications, 45 of which included emotion analysis (see Table 2). This is about 5% of the total number of publications.

[4] For sociology, we searched in *Social Sciences* (subject area), *Sociology and Political Science* (subject category), and then selected the five most cited journals with the word "sociology" or "sociological" in the title. For computer science, we selected the five most cited journals in *Computer Science* (subject area), *Artificial Intelligence* (subject category).

[5] We have searched for the following terms and their derivatives in title or abstract: 'emotion', 'sentiment', 'feeling', 'affect', 'empathy', 'trust'. We supplemented the manual search with an automated keyword search to ensure we had not missed any relevant papers. Only papers whose central topic was emotion analysis were included in our survey. We did not consider book reviews, errata, or other technical pieces.

Table 2 Sampling results

Journal	Total number of publications (01/2020-09/2025)	Number of publications of emotions (01/2020-09/2025)
ARS	167	6
ASR	215	13
SMR	303	4
AJS	185	14
ST	93	8
For all sociological journals	**963**	**45**
FTML	30	0
IJIM	719	78
SR	625	9
NMI	940	10
CEAI	463	38
For all computer science journals	**2777**	**135**
For all journals	**3740**	**180**

Figure 1 illustrates that sociological journals vary widely in how much space they give to examining emotions: *Sociological Methods and Research* has the lowest share, *Sociological Theory* has the highest share. Thus, we see an interest in emotions from theorists rather than from methodologists. However, considering the time dynamics (see Fig. 2), the share of publications in our sample remains small, but relatively stable, ranging from 4 to 6%.

Analyzing the content of the articles, we have identified six categories (see Table 5)[6]. The first three correspond to the basic division of the sociology of emotions into cultural, structural, and interactional approaches we have outlined. As Table 3 shows, the relationship between emotions and social structure (class, gender, race, etc.) is the dominant topic. Researchers pay less attention to the interactional and cultural aspects of emotions.

The category "emotions and decision-making" was added during the coding process. It reflects the interest shared by different disciplines in the interrelations between emotions, reasoning and human actions[7]. The categories "theory" and "method" characterize publications that are devoted to the actual development of theoretical or methodological foundations for the analysis of emotions in sociology. Compared to the "emotions and social structure" category, these three topics have three times fewer researchers.

Only two of the six theoretical papers in our sample try to use mathematical formalization to make their points. One paper examines the relationship between status and sentiments using the Expectation States Tradition from social psychology (Bianchi & Shelly, 2020). Another article correlates Durkheim's ideas on culture and emotions with the formalization in network analysis (Puetz, 2024).

[6] An article may be labelled by one or several categories.

[7] Another interdisciplinary turn should be noted, the interest in mental health, which was presented in few publications included in the sample.

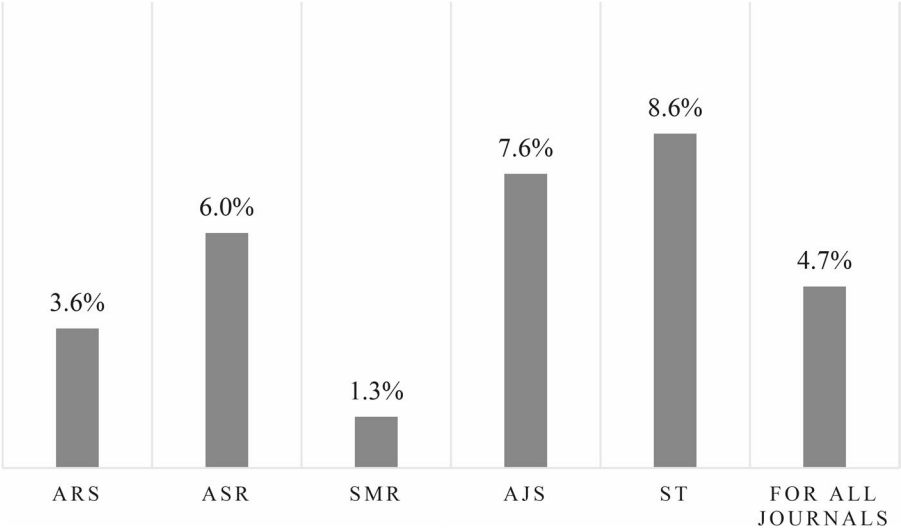

Fig. 1 Share of sociological publications on emotions (by journal)

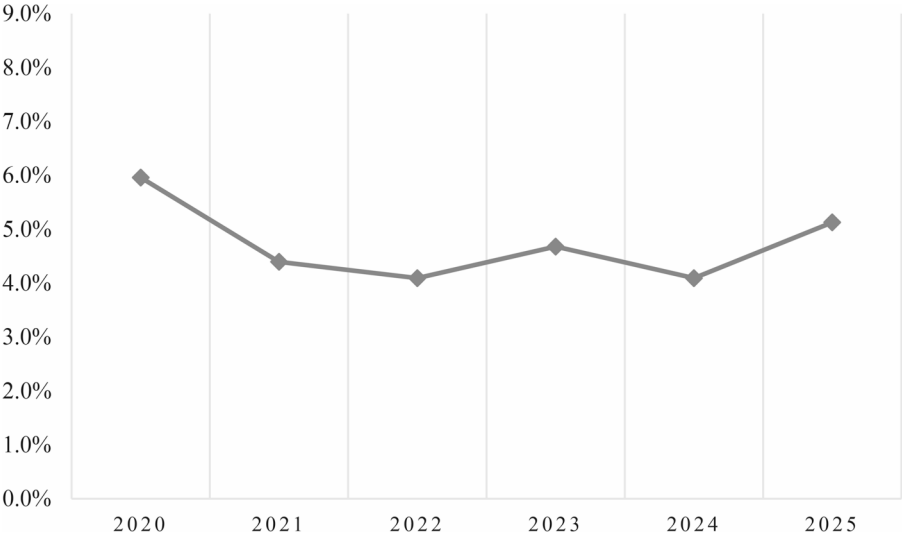

Fig. 2 Share of sociological publications on emotions (by year)

Table 3 Categorization of sociological journal papers on emotions

Category	ARS	ASR	SMR	AJS	ST	Total
Emotions and social structure	2	5	1	9	2	19
Emotions and culture	0	3	1	3	3	10
Emotions and social interactions and relations	1	4	0	4	4	13
Emotions and decision-making	1	2	0	1	1	5
Theory	1	0	0	0	5	6
Methods	1	0	4	0	0	5

The relationship between emotion analysis in sociology and the development of computer algorithms is not explored in theoretical texts, but rather in those dealing with methods that utilize algorithms to analyze the emotional content of texts (Edelmann et al., 2020; Voyer et al., 2022).

Randall Collins' interaction ritual theory has been examined in two different publications in our sample. Iddo Tavory and Nicholas Hoynes (2025) propose the development of Collins' theoretical ideas. However, they do not apply to formalizations. On the contrary, they correlate Collins' theory with more interpretative traditions in sociology. And John N. Parker, with colleagues (Parker et al., 2020) propose to use 'sociometers' (wearables that collect data about human interactions) in order to test the theory. Thus, there is clearly a trend towards quantification here.

To summarize, our analysis has demonstrated that the sociology of emotions is a quite narrow but stable research area. The dominant problem in the sociology of emotions today is the interrelations between emotions and social structure. The degree of formalization of the theory is relatively low, and computer technologies are employed as one of the methods of data analysis, rather than as the instruments of theory-building.

Computer science

A review of computer science journals between 2020 and 2025 discovered 2,777 publications, of which 135 were related to emotion analysis (see Table 2). So, the volume of publications in computer science is about three times larger than in sociological journals. However, the proportion of publications on emotions is similar for the two disciplines, at around 5%.

Yet, the distribution of publications in computer science differs from that in sociology. Looking at time dynamics, we observe here more pronounced fluctuations, with a decline in interest in emotions in the middle of the period and a rise in interest in 2025 (see Fig. 3).

The proportion of publications on emotions among computer science journals varies greatly (see Fig. 4). We found no publications addressing emotion in the *Foundations and Trends in Machine Learning* journal. *Science Robotics* and *Nature Machine Intelligence*, which address the 'hardcore' computer science community, have a share of publications on emotion of slightly more than 1%. At the same time, the *International Journal of Information Management* and *Computers and Education: Artificial Intelligence*, which discuss applications of computer science in management and education, have a relatively high share of publications on emotion.

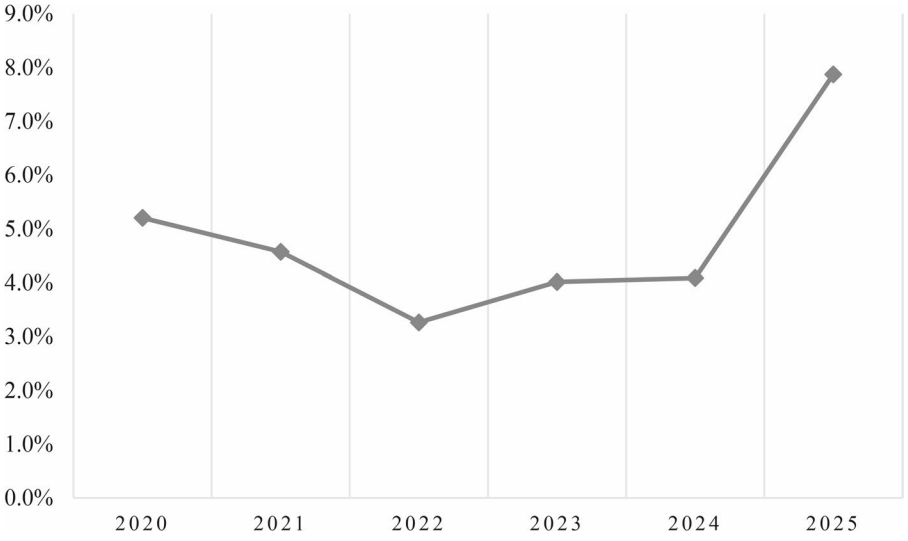

Fig. 3 Share of computer science publications on emotions (by year)

There are significant differences in the themes of articles published in different journals. We identified three categories corresponding to the three main tasks in computer emotion analysis: emotion modelling, analyzing emotions in human-computer interactions (HCI)[8], and emotion recognition. Table 4 illustrates how the articles are categorized.

The most popular category is 'Emotions in HCI', which is divided into two subcategories. The first one, 'Technology as an Agent', embraces analysis of the emotional aspects of human interaction with computer systems, including both human emotional experiences and imitation of emotional reactions by algorithms. This topic dominates all journals in which we found publications on emotions. The second subcategory, 'Technology as a Milieu', contains publications that analyze the emotional effects of the computer-based setting of human interactions (such as social networks, online shopping platforms, MOOCs, etc.). This category is prevalent in the *International Journal of Information Management*.

The second most popular category is 'Emotion Recognition', which includes three subcategories: 'Sentiment Analysis', 'Facial Recognition', and 'Biomarkers'. While sentiment analysis is already widely used in various studies, facial recognition is just entering the applied research arena, and biomarkers are mentioned only once.

The third, the least popular category, 'Emotion modeling' in AI, is of most interest to us. It includes only five publications out of 145 – three in *Science Robotics* and two in *Nature Machine Intelligence*. None of them is a research article. Four are essays on whether AI may/may not/needs to have emotions (Empathic AI…, 2024; Shteynberg

[8] Publications about animal-computer and human-animal-computer interactions are categorized under 'human-computer interaction'. We decided not to create a special category for them because, in the AI field, interest in animal intelligence and emotions is closely connected to interest in human intelligence and emotions.

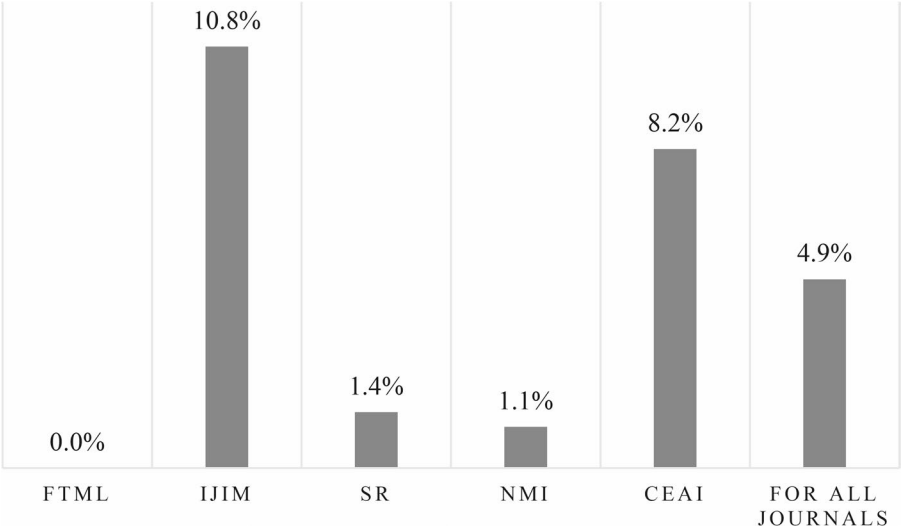

Fig. 4 Share of computer science publications on emotions (by journal)

Table 4 Categorization of computer science journal papers on emotions

Category		FTML	IJIM	SR	NMI	CEAI	Total
Emotion modelling		0	0	3	2		6
Emotions in HCI	**Technology as an agent**	0	30	7	8	33	78
	Technology as a milieu	0	32	0	0	1	33
Emotion recognition	**Sentiment analysis**	0	24	0	0	4	28
	Face recognition	0	0	1	1	5	7
	Biomarkers	0	0	0	0	1	1

et al., 2024; Christov-Moore et al., 2023; Murphy, 2025). One paper contains a short report on the design of a robot whose emotional states are modeled by simulating changes in hormonal regulation (Yashinsky, 2024). Essentially, this is the only publication in our sample that addresses the issue of modeling emotions in AI (rather than imitating emotional responses in interactions with humans). However, this paper is concerned with modeling neurophysiological conditions of emotions.

Thus, the review of journal articles provides evidence of a specific interest in emotion analysis within the field of computer science. However, this interest is primarily driven by applied problems of emotion recognition and tracking user emotional dynamics, rather than by the challenges of computer science itself.

Computer science: Proceedings

Unlike in sociology, computer science scholars rely heavily on conference proceedings to share their academic findings. Moreover, in recent years, AI technologies (especially generative AI models) have shown stable advances every few months. Due to these considerations, we chose to incorporate an analysis of conference pro-

ceedings from the previous year, in addition to our content analysis of computer science journals.

The research design for conference proceedings analysis is driven by the combination of our research interests with the specifics of the data (a large number of proceedings compared to journal articles).

Our analysis of conference papers is not focused on identifying the overall distribution of topics, as it was for journal papers. Here, we aim to: (1) assess the validity of the categories we have found, and (2) determine current and emerging trends.

As an empirical material, we choose conferences of the Association for Computing Machinery (ACM), the largest and most recognized association for computer science, founded in 1947[9].

The search has two steps. First, we analyzed 100 papers published over the past year, sorted by recency. Our goal was to identify new categories and test the resilience of the old ones. After that, we analyzed all ACM conference proceedings from the last three months (September-November 2025), which included 341 additional papers. Here, we aimed to achieve empirical saturation of the sample of publications[10]. Out of 441 papers, we excluded 54 (off-topic).

Overall, we examined 387 papers from conference proceedings. The majority of these papers come from various computer science areas, but some social scientists also contributed.

Table 5 presents the category distribution for the proceedings papers[11]. We can notice that the coding scheme has proved its resilience. Also, several new sub-categories within the 'emotion recognition' category were identified (new sub-categories are indicated in *italics*).

Compared to journal articles, conference proceedings highlight several new trends.

The first one is the growth of various data types used in emotion recognition, such as text, video, sound (voice, music, and animal vocalizations), facial expressions, body language, and biological data (electroencephalograms, electrodermal activity, heart rate, and other metrics).

The second one is multimodal emotion recognition that is based on analysis and synthesis of emotion recognition for different modalities. For this purpose, researchers now often use multimodal LLMs, which have made significant progress in the last few years.

[9] It is also the major association for computer science per se, as compared, for instance, to IEEE, where computer science is just one of the fields. The search was conducted in November 2025 on the ACM Digital Library's website (The ACM Full-Text Collection). In most cases, the paper was available in PDF; if it was not, we analyzed the abstract and the highlights from the paper available through the search engine. The term 'emotion' was used in the abstract search. The choice of the most universal term was due by to our interest in emerging trends rather than precise distributions of topics. As an argument in favor of comparability of with journals' sample, we note that many proceedings included terms that we used for the journal search (sentiment, empathy, etc.).

[10] We admit that focus on last three month may introduce seasonal bias, and that spring conferences might show different trends. The one-year sample gives a broader perspective and helps to address this limitation. We noticed no remarkable difference between the two stages of research in terms of specific topics and methods.

[11] As for journals, one paper here could be labelled as one or several categories.

Table 5 Categorization of computer science proceedings papers on emotions

Category		ACM Proceedings
Emotion modelling		9
Emotions in HCI	**Technology as an agent**	112
	Technology as a milieu	114
Emotion recognition	**Sentiment analysis*(text & image)***	44
	Face recognition*(including gaze)*	25
	Biomarkers	14
	Body movements	*7*
	Vocal	*13*
	Multimodal emotion recognition	*103*
	Emotional reasoning/ understanding	*40*

The third is "emotional reasoning/understanding" that has become a specific task. This task also typically employs LLMs (see: Lian et al., 2025; Zhu et al., 2025). 'Emotional reasoning/understanding' category lies actually in-between 'emotion recognition' (that is, a very popular field) and 'emotional modeling' (which is presented sporadically in proceedings, as it is in journals). Besides just evoking 'reasoning' about emotions from LLMs, there are several attempts to integrate emotion models frameworks from cognitive and computer sciences, such as the dual-process model (Sethi et al., 2025), attachment theory (Wang et al., 2025), and game theory (Mozikov et al., 2025).

Finally, emotion analysis in HCI equally considers the technology as both an active agent and a medium for interaction. The overarching aim of HCI research often revolves around the refinement of emotional interactions, with the intention of making them both smooth and accessible to users. The established research methodology involves the application of behavioral science, combined with empirical induction, to ascertain viable solutions.

Comparison

By comparing the results of the content analysis for sociology and computer science, we can answer two of the three initial questions we formulated in this paper.

Does a clear disciplinary boundary exist between theorizing about emotions in sociology and computer science? - Yes, a clear disciplinary boundary exists, separating fields through their research topics, approaches, and methodologies. The zone of convergence between the two is sentiment analysis, a method of automatic emotion recognition. With recent developments in multimodal emotion recognition, we can expect that social science will likely incorporate more research tools from computer science.

To what extent do publications in both fields exemplify "good bad theory"? - Neither sociology nor computer science considers the theoretical study of emotions as a dominant research topic. Sociologists tend to avoid over-formalizing human emotions, while computer scientists tend to avoid theorizing emotions at all. Currently, we cannot see the prevalence of "good bad theorizing" about emotions in mainstream

sociology and computer science. Moreover, based on our analysis, we propose that in the coming years, the primary approach to modeling emotions in computer science will shift to training extensive models on various data types and developing their capacity for "emotional reasoning".

What could these findings suggest?

At first, we thought that "good bad theory" was a problem. However, it now seems more likely that the real issue is that it is missing.

The way we define "good bad theory" is more abstract than what we found in our sample. In sociology, emotions are seen as socially important, but there has been little effort to develop theories about them, even though the field is rich in different concepts. In computer science, emotion research is mostly about practical problems and new technology, not about building theories.

Thus, 'good bad theory" is not a diagnosis of an already prevalent mode of theorizing, but an anticipatory critique of an emerging trajectory enabled by AI developments.

However, there *are* examples of "good bad theories" that are acknowledged by the researchers today. In the next section we move to a qualitative analysis of two cases of 'good bad theories' in order to answer the third, and most important, question: *What are the key advantages and limitations of applying formalized models in human emotion analytics?*

Comparative case study: Randal Collins vs. Stuart Russell

In what follows, our discussion will address, as the case study, two emotion theories: Randall Collins' interaction ritual theory (2004) and Stuart Russell's theory of human-compatible AI (2019; 2022). The authors of these theories are distinguished experts and authorities in their fields. Both scholars are recognized for their attempts to find common ground between social and technical issues and to bridge the gap between the social and hard sciences.

Randall Collins was among the first sociologists to write about AI technology and its potential to advance sociological knowledge (Collins, 1992; 1994). Also, as we will explore further in more detail, his theory relies on quantification in the analysis of human emotions and interactions.

Stuart Russell, for his part, is one of the most humane-oriented computer scientists of our time. He criticizes the technology-first approach in modern AI research and advocates for technology development that prioritizes social considerations. He aims to ensure that human-AI interactions are socially constructive and safe.

What led to the selection of these theories for examination? We believe that these theories are paradigmatic examples of "good bad theory" for each discipline and consider them to be 'extreme cases'.

Our content analysis demonstrates that in sociology and computer science, collaboration is taking place basically at the level of methods (the former borrows from the latter) rather than theories. The reason for this is different conceptual orientations. Sociologists typically do not seek to formalize the analysis of emotions, while computer scientists are interested in emotion modeling mainly through analyzing patterns

in data and obtaining reproducible results. Stuart Russell and Randall Collins, as we will see, offer contrasting perspectives on these trends.

In our examination of these cases, we will pursue the following logic: an overview of the theorist's core concepts, an analysis of studies that employ the theory, and finally, a critical evaluation.

Case 1: Collins' emotional utilitarianism

Randall Collins' interaction rituals theory (IRT) represents a paradigmatic example of how to formalize social theory's insights and make them friendly to computational analysis, and at the same time, to ensure their sociological adequacy.

The key idea of IRT is that the reality of interaction with other people shapes human consciousness and behavior. As Collins put it, "At the center of an interaction ritual is the process in which participants develop a mutual focus of attention and become entrained in each other's bodily micro-rhythms and emotions… [Rituals] result in the ritual outcomes of solidarity, symbolism, and individual emotional energy… The key process is participants' mutual entrainment of emotion and attention, producing a shared emotional/cognitive experience" (Collins, 2004: 47–49).

From the individual's perspective, the balance of successful and unsuccessful interactions is expressed in the level of emotional energy, which represents the general emotional background of short-term emotions in specific interactions. Emotional energy can be empirically measured in various ways: as physical comfort at the start of an interaction, as the emotional attractiveness of certain types of interactions, as the vividness of memories and experiences of past interactions, and as the ease of synchronization between participants in their speech and movements.

Collins employs a form of utilitarian anthropology, which we refer to as *emotional utilitarianism*[12]. The basic idea is that people want to maximize emotional energy through successful interaction rituals. Collins argues: "Humans are not very good at calculating costs and benefits, but they feel their way toward goals because they can judge everything subconsciously by its contribution to a fundamental motive: seeking maximal emotional energy in interaction rituals" (Collins, 2004: xii-xiv). This formula obviously orients the algorithmic representation of complex social processes while maintaining claims about human sociality.

Interaction ritual theory is a subject of considerable interest among researchers. There are several major trends associated with IRT in social science production.

First, IRT is further developed by Collins himself (Collins, 2012, 2020) and by his colleagues (Boyns & Luery, 2015; Weininger et al., 2018; McCaffree & Shults, 2022). However, some differences exist here. There are social analysts who follow the path of applying the IRT theory within the logic of 'emotional utilitarianism' (Berikat, 2014; Baker, 2019), and there are those who combine Collins' ideas with more interpretative and phenomenological approaches (Summers Effler, 2010; Grønnestad et al., 2020; Tavory & Hoynes, 2025).

Second, researchers use IRT as a basis for organizing and conducting applied research in commercially oriented areas, such as management and marketing (Methot et

[12] This idea was discussed in detail in (Rezaev & Tregubova, 2022).

al., 2021; Xiang et al., 2022; Joo et al., 2023; Cayla & Brigitte, 2025; Yim et al., 2025). IRT assists them in achieving optimal operationalization and in delivering predictions. Within the framework of applied research, IRT is also used in fields such as educational studies (Droppe, 2022) and sport studies (Cottingham, 2012; Zhang et al., 2024).

Third, there has been a growing use of IRT in analyzing online interactions (DiMaggio et al., 2018; Tregubova & Ni, 2020; Chen & Skey, 2024; Mizrahi-Werne et al., 2025; Bazan Royuela, 2025). Contemporary research suggests that the physical presence of others is not always necessary for generating emotional energy, despite Collins' initial hypothesis. The internet and online technologies facilitate the manifestation of relatively intense emotions, and the online milieu is suitable for the measurement of the constituent elements of interaction rituals quite well in various research designs. Moreover, the mediation of interaction rituals through technologies needs systematic research on the types of technologies used, by whom, in what cultural settings, and with what effect (Johannessen, 2023).

We have not found any attempts to formalize IRT to make it suitable for AI analysis, although it has been done with another Collins' theory, the state breakdown model (Mökander & Schroeder, 2022). However, looking at IRT from a technical standpoint, it is evident that its structure might be attractive to researchers who develop AI models to study social processes. It provides clear operational definitions, measurable variables, and testable hypotheses about social interaction. Also, this approach shows how to develop algorithms that can model emotional dynamics within social systems.

Now, regarding criticism, the formalization of Collins' theory has substantial drawbacks. The shortcomings are especially evident when we examine IRT in terms of traditional sociological criteria.

Anne Rawls' critique (1989) illuminates the most fundamental problem with Collins' emotional utilitarianism. Rawls argues that Collins reduces social interaction to instrumental pursuit of emotional energy, missing the constitutive role of communication in creating and maintaining human selfhood. Referring to sociological classics, she argues that sociality is intrinsically valuable because it serves the creation and maintenance of Self and meaning, not because it produces emotional satisfaction. Human sociality shapes the very existence of Self, which depends on the existence of other Selves and interaction with them. Other people are not simply part of the environment, even the most important part, and Self simply a designation of a mechanism for (un)consciousness calculation of gains and losses.

This critique reveals how Collins' approach embodies the "good bad theory" problem. While the theory allows computational modeling and algorithmic implementation, it is blind to the constitutive role of interaction in creating selfhood, the qualitative irreducibility of emotional states, and the interpretive nature of meaning-making. And these are the characteristics that social theory has identified as central to explain the reality of human social life.

Case 2: Russell's game-theoretic approach to human-AI interaction

Stuart Russell's influential work on human-compatible AI (2019; 2022) provides a compelling example of "good bad theory" in computer science approaches to human emotion modeling.

Russell opens his thesis with a critique of what he calls 'the standard model of AI'. His central argument is that AI creators provide machines with human attributes – humans have goals and they try to find how to achieve them, so algorithms need to have the same qualities. However, as technology advances, the ability of machines to achieve goals raises risks and dangers.

What is the reason for this? Russell argues that humans and AI agents differ to such an extent that correctly "to translate" human goals into machine language is impossible.

Now, what does Russell suggest as a solution? He proposes the following criterion: AI tools are beneficial to the extent that their actions can be expected to achieve *human* goals. Russell's approach treats human-AI interaction as a strategic game where AI systems must learn to understand and optimize human preferences through observation of human behavior. The theory is operationalized through three core principles. First, the machine's only objective is to maximize the realization of human preferences. Second, initially, the machine is uncertain about what these preferences might be. Third, the most important source of information about human preferences is nothing but human behavior (Russell, 2019).

From a computational perspective, Russell's approach represents a significant advance in AI safety and alignment research. This framework translates emotional responses and value judgments into utility functions that can be mathematically modeled and computationally processed. The elegance of this formulation lies in its ability to handle uncertainty about human preferences and maintaining at the same time mathematical accuracy and algorithmic tolerance.

Russell and his colleagues develop and implement these ideas in various research directions: algorithmic prediction of human decisions (Bourgin et al., 2019; Plonsky et al., 2025), accounting for various and changing human preferences (Conitzer et al., 2024; Carrol et al. 2024), design of AI agents 'motivated' to ask people for help/information (Emmons et al., 2024; Plaut et al., 2025), and design of 'internal motivation' in AI agents in general (Lidayan et al., 2024). This work provides a solid base to researchers who reflect upon similar problems, such as design of AI agents 'motivated' to adapt to a specific user (Wan et al., 2025), AI alignment with cultural values and social preferences (Tay et al., 2020), and formalizing mechanisms of human decision-making (Ho & Griffiths, 2022; Virca, 2024).

Concerning emotional aspects of human-AI interaction, several scientists employ Russell's ideas to argue for 'sociaffective alignment' between human and AI (Kirk et al., 2025), as well as for the use of emotional human feedback in AI learning (Pollak et al., 2022). Others, on the contrary, use Russell's arguments in their critique of AI limitations in replication (Montemayor et al., 2022; Nerantzi, 2025) and recognition (Latif et al., 2022) of human emotions.

This controversy highlights the real problem that the usage of human emotional reactions and connections for AI alignment with its users seems very promising, yet its realization is complicated and problematic.

Russell's approach offers an apparently simple way to solve this problem when an AI agent explicitly asks a human, and these answers, combined with human actions (rather than preceding and accompanying emotional processes), are the primary source of information for the AI. Russell's solution, however, has its weaknesses.

The assumption that humans possess consistent, discoverable preferences that AI systems can optimize reflects a utilitarian anthropology that has been extensively critiqued in social theory[13]. For humans, and even for higher animals, the system of preferences is shaped by basic needs that are irreducible to each other and potentially conflicting (Midgley, 2002). Meeting various needs, their combination within the framework of one's own life, makes a human being rational, not an implementation of a certain algorithm based on ranked preferences. Emotions are important for a human being precisely because they help to be rational, to make decisions without an ordered framework of conscious preferences.

Another counter-argument against Russell's framework concerns the narrow and artificial situations in which it is tested. In fact, this argument is relevant for the majority of research and design in computer science. In the history of AI development, it has often been the case that technologies that worked well in laboratory conditions have ceased to be effective in a complicated and unpredictable environment. The famous philosopher and AI critic Hubert Dreyfus called this situation 'a first step fallacy'. As he puts it, "Climbing a hill should not give one any assurance that if he keeps going he will reach the sky" (Dreyfus, 2012: 87).

In fact, Collins' and Russell's frameworks deal with similar theoretical limitations that, when implemented in AI systems, become practical problems. AI systems designed according to their principles may successfully optimize revealed behavioral and emotional preferences while missing entirely the interpretive processes through which humans make sense of their emotional experiences and the social contexts that give those experiences meaning. Here we are facing the phenomenological challenge to AI development, which we will consider in more detail in the next section.

The phenomenological challenge to computational emotion analytics

The phenomenological dimensions of human emotional experience present perhaps the most fundamental challenge to computational approaches in emotion analytics. Phenomenology, understood as the study of consciousness and experience from the first-person perspective, reveals qualitative aspects of human emotional life that appear categorically resistant to mathematical formalization.

Human emotional experience involves qualitative distinctions that cannot be captured through quantitative measurement. The difference between joy and despair involves not merely degree along a single dimension but qualitatively distinct modes of being-in-the-world. Martha Nussbaum's analysis of emotions as 'upheavals of thought' illuminates this qualitative irreducibility: emotions are cognitive evaluations that attribute significance to objects in relation to personal flourishing, involving an ineliminable reference to Self that resists computational representation (Nussbaum, 2001). Nussbaum also emphasizes emotional ambivalence in human life. She identi-

[13] Compared to Collins' emotional utilitarianism, Russell's 'preference utilitarianism' is more straightforward but less developed. Russell relies on a simpler model of ordered conscious human preferences while Collins considers 'subconscious' calculations of emotional energy.

fies a core conflict: the things that cause emotions are essential for well-being but are not completely controllable.

Comparison between Collins' and Nussbaum's theories of emotions demonstrates that Collins' framework does not help in understanding the vital qualities of human emotions (Rezaev & Tregubova, 2022). While his concept of emotional energy explains the ritual dynamics of how humans develop affections and solidarities, it cannot grasp emotions' unique qualities and distinctiveness, their inherent ambivalence, and their connection with the human image of personal well-being, as Nussbaum's theory does. In this sense, Nussbaum provides a more comprehensive account of human emotions.

Accordingly, Russell's emotion analysis relies on the ability to measure and rank human preferences based on the gathering of behavioral data, including human responses to AI prompts. Nussbaum's argument suggests that this approach may be practical in instrumental interactions, but not in situations of high personal relevance.

Several arguments from the phenomenological tradition should be added to this basic criticism.

First, the temporal structure of human emotional experience presents additional challenges for computational formalization. Human emotional life involves what Edmund Husserl (1991) termed "retention" and "protention", the ongoing synthesis of past experience and future anticipation that constitutes present awareness. This temporal synthesis creates meaning contexts that inform emotional responses but cannot be decomposed into discrete computational elements without losing their essential character.

Second, the intersubjective dimension of human emotional life reveals another aspect of the phenomenological challenge. Human emotions emerge through what Maurice Merleau-Ponty described as embodied interactions that create shared meaning prior to explicit communication or reflection (1945). These pre-reflective forms of social connection involve synchrony, rhythm, and mutual attunement that operate below the threshold of conscious awareness while fundamentally shaping emotional experience.

Finally, Martin Buber's distinction between "I-Thou" and "I-It" relationships illuminates the moral dimension of this phenomenological challenge (1970). Human emotional life involves encounters with others as autonomous subjects rather than merely complex objects within an environment. This recognition of others' subjectivity creates moral and emotional obligations that cannot be reduced to utility maximization or strategic calculation while remaining central to human social experience.

These phenomenological considerations suggest that certain aspects of human emotional life may be inaccessible to formalization and, as a result, to computational analysis. This limitation is not merely technical but reflects the categorical difference between information processing and conscious experience.

However, for many researchers, the phenomenological challenge to computational emotion analysis is not immediately evident today. The following section will explore the reasons behind it.

The historical challenge: From 'Good bad theories' to 'Misleading theories' in emotion analytics

Our discussion so far has focused on the theoretical and methodological analysis of emotions. However, the way people perceive and handle emotions within a specific culture and historical time also plays a crucial role. And here we can see that in the current situation (at least in countries with a high level of digital transformation and developed capitalism), it is apparently easy to study emotions as quantifiable, because that is how humans perceive them. For this, we have two explanations.

The first is the development of "emotional capitalism," when emotions are viewed as a subject of calculation and investment (Illouz, 2007; Cabanas & Illouz, 2019). In the economic sphere, this corresponds to the emergence of an "experience economy" in which profit depends on the experience that a product or service evokes in consumers (Pine & Gilmore, 2011; Illouz, 2017).

The second is the rapid development of the online milieu, which itself presupposes the quantification of emotions (in the form of likes, dislikes, stars, etc.) and standardized expressions of emotions (emojis, memes, and so on). Current research demonstrates that emotions spread through social networks in ways that can be modeled mathematically, suggesting possibilities for computational approaches to collective emotional dynamics influence (Centola, 2018). These transformations involve not only changes in communication channels but also fundamental alterations in emotional experience itself (Haidt, 2024).

When we examined the application of IRT in contemporary social research, we noticed both of these trends. Collins' theory is in demand both for the assessment of effects of social connections and experiences in the service sector and for analyzing emotional dynamics on the Internet.

These trends are intertwined. The online environment is favorable for the development of capitalist logic in the perception and management of emotions. Users' emotional expressions become a source of data that is used to stimulate new emotions (Illouz & Kotliar, 2022). "Quantified self" emerges when a person evaluates his/her own well-being perception with constant use of data from digital devices and software (Lupton, 2016).

The digitization of social interaction has direct impact on AI development. AI tools online become a part of the social settings acting as various bots, algorithms, and apps. This raises a number of conceptual questions about emotions: How do AI systems shape emotional experience and expression? What are the social consequences of algorithmic emotion detection and management? How do people adapt their emotional behavior to technological systems designed to monitor and respond to emotions? These questions require new theoretical frameworks that can address the co-evolution of human emotional experience and artificial intelligence systems.

Regarding the issue of "good bad theories," the conclusion is quite simple. These theories could become a self-fulfilling prophecy because social interventions and information technologies, created on their basis, would make them seem even more plausible for designers, users, and even for social researchers.

Thus, the danger is that we will be dealing not simply with "good bad theory" but with a special kind of it, a "misleading/allegedly good" theory that is essentially false.

Discussion

The analysis developed in this paper has demonstrated that the "good bad theory" problem represents a fundamental challenge to emotion analytics. Through comparative analysis of prominent approaches in both social science and computer science, we have shown how theories that achieve mathematical precision and algorithmic welcome often do so by eliminating precisely those aspects of emotional life that make it human and social. We analyze the topic at two levels, and each level offers its own ways that support the main argument. Quantitative content analysis reveals current trends in the field and indicates the different ways researchers study emotion. The comparative case study focuses on extreme examples of formalization and shows that technical success can sometimes fail to understand real experiences. As a result, we identified systematic tensions between computational requirements and the phenomenological complexity of human emotional experience.

Our empirical examination of publication patterns shows that today's focus on emotional issues in computer science stems more from applied problems. LLMs have become increasingly capable of fulfilling this demand, as they are used for data analysis and for the comprehension of emotional processes.

Of course, the phenomenological critique outlined in this article is also valid for these tools. However, they are an easy target. Studies in this area often either avoid theory altogether or depend on simple basic ideas, such as evaluating emotions with only two dimensions: pleasantness ('valence') and intensity ('arousal'). Now, computer scientists are exploring the limits of LLMs' capabilities. Once the limits are identified, the leading researchers will move towards more detailed theoretical frameworks, like Russell's and Collins'.

In this regard, the evolution of "emotional capitalism" and "quantified self" may awaken a more substantive interest in emotions, precisely the kind that will be satisfied by "good bad theories." This is both good and bad news for sociology. The good news is that, along with social psychology, it will be in demand as a source of knowledge about emotions. The bad news is that, out of the rich tradition of social analytics of emotions, only a few approaches will be in demand, and their critique has been extensively presented in this paper.

Considering this issue, we would like to address two questions:

- Taking into account the limitations of computational techniques for emotion analysis, should we conclude that traditional qualitative methods in social sciences also do not face these limitations? Are *they* susceptible to the phenomenological critique?
- Are there any significant strengths and areas with high potential for AI to analyze and simulate emotions?

The answer to the first question can be summarized in two propositions.

1) A human can understand a human (although not necessarily correctly); a computer cannot. Here we agree with Searle's classical argument about syntax and semantics: AI acts according to syntax and has no access to semantics. This is

> true not only for symbolic AI but also for approaches based on searching for patterns in data. A simple observation illustrates this idea: LLMs recognize not words, but tokens, which could be meaningless from a human point of view, but are more convenient for AI models.[14].

Searle's thesis on syntax and semantics is a subject of intense debate. For the purposes of this article, it is relevant to consider one of the most significant counterarguments, which states that we can view understanding, in Wittgensteinian terms, as an ability to participate in 'language games'. If an AI agent can use language in a way that seems meaningful within specific situations and contexts, why not to assume that it is capable of understanding? The recent development of AI models makes this argument quite compelling. Modern transformer models learn to predict words by finding patterns in large amounts of human language. Although this is still technically pattern matching, these models are getting better at responding to context in ways that are closer to how humans communicate.

If we follow this idea, we must ask: when does an AI's ability to take part in emotional language games count as real understanding?

We suggest that understanding emotion involves at least three main abilities: (1) pattern recognition, which means spotting emotional expressions and their patterns (here AI increasingly excels); (2) contextual responsiveness, or changing interpretations based on the situation (here AI shows growing competence); (3) experiential grounding, which is linking emotional ideas to real-life experience (where AI remains definitely limited). Current AI systems are getting better and better at the first two abilities, but they are still very limited when it comes to the third. This is not just a technical problem. AI systems do not have the physical, time-based, and shared experiences that form the basis of human emotional understanding.

To see, why it is so, let us address the arguments from philosophy and social theory. One was developed by phenomenologists Hubert Dreyfus in his polemics with Harry Collins, who has formulated a version of wittgensteinian argument for AI "understanding" (Dreyfus, 1992, 1996; Collins, 1996; Selinger et al., 2007). Dreyfus points out that the embodied and intentional character of human experience is the basis of human forms of life that determine the use of language and the process of understanding. Therefore any imitation of understanding in a computer (which is a fundamentally differently entity than a human being) will always be imperfect. Another argument, from the perspective of symbolic interactionist theory, was developed by Alan Wolfe (1991). He observes that artificial intelligence is capable of following a rule or even breaking a rule when the patterns in learning data suggest it. The human mind, however, is also capable of creating rules (that is, meaning-making) in specific, novel, unique situations.

So, it is true, that current AI tools can execute "emotional understanding" (in a sense of imitating human verbal understanding) in a wide range of contexts. However, in new, atypical and personal contexts LLMs are quite unreliable. Even Harry Collins, a proponent of incorporating machines into 'language games', is skeptical in his assessment of the achievements of modern AI tools (Collins, 2018, 2021).

[14] See: URL: https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them.

2) Arguments about the specific strengths of qualitative methods in emotion analysis could be formulated only if we take a moderate phenomenological stance, not an extreme one. Radical phenomenological criticism calls into question the possibility of scientific knowledge about a human being, as opposed to philosophical, humanitarian and artistic comprehension. From this point of view, any scientific study of emotions, whether qualitative or quantitative, leads to a basic distortion. A more appropriate position is the moderate phenomenological position developed by phenomenologists who avoid both scientistic reduction and more radical ontological reconfigurations. It accepts that full objectivity is not possible. However, people can still reach shared understanding through common experiences and culture. In the framework of moderate stance, qualitative methods are regarded as having a significant advantage over quantitative ones.

The question about qualitative methods leads us to the origins of sociology. There have been discussions about the irreducibility of social science to natural science since its origins. Sociological classics, Weber and Durkheim, proposed their 'programs' for social science as different from (though, in some sense, inspired by) natural sciences (Smelser, 1976). In the middle of the 20th century, this issue was again brought to discussion by Peter Winch's seminal book *The Idea of a Social Science and its Relation to Philosophy* (1958). Our interpretation of this problem does not rely on formal analysis of individualistic and holistic properties, as, for instance, in (Hansson Wahlber, 2019) and (Di Orio, 2024). We focus on a broader argument formulated from the standpoint of existential and phenomenological critique. It is broader because it transcends social sciences disciplinary issues and argues for the irreducibility of human existence to the system of objectified categories.

What does this argument mean for sociological research? Reflecting on essentially the same question, Max Weber introduced the concepts of 'Verstehen' and an ideal type. His position is that we can grasp not necessarily the unique nuances of the situation, but the typical meaning of action, which is also comprehensible to the members of society. Qualitative methods have advantages over computational approaches because they let researchers access the lived experience of human beings. By using methods such as participant observation, in-depth interviews, and interpretive analysis, qualitative researchers can get close to 'Verstehen', even though they know that full transparency is not possible. This demands 'qualitative literacy' (Small, & Calarco, 2022) which includes cognitive empathy – an ability to look at the situation through the lens of another person considering his/her position in social structure, as well as biases that depend on a researcher's own position and values.

This moderate view lets qualitative methods partly avoid phenomenological critique in three ways: (1) reflexivity, where researchers reflect on their own position; (2) dialogue, where meaning comes from interaction between researcher and participant instead of just taking information; and (3) thick description, which keeps the rich context connected to lived experience. Although these methods do not close the gap between experience and representation, they help keep connections that computational methods often disrupt.

Thus, universal human ability to recognize and interpret emotions of other people in combination with social science professional training helps to mitigate the exis-

tential and phenomenological critique of limitation of scientific cognition. In the age of AI it means that, though LLMs could provide interpretations of data, it is a human researcher who assesses the plausibility of these interpretations.

Regarding the second question, it may be rephrased: In what situations will the move from semantics to syntax be the least challenging and most advantageous for emotion analysis? Or in other words, when can we trust AI to assist in collecting, analyzing, and interpreting data?

Here is the most plausible correct response - when human emotional expressions are simple and/or routinized. For example, sentiment analysis is growing in popularity in the analysis of online settings, where expressions of emotion are both typical and recurring.

One more area should be pointed out here – highly institutionalized contexts and scenarios that AI algorithms can be trained to recognize. Examples include genres of TV shows, movies, public speaking, and so on[15]. Here, we can expect promising uses of multivariate emotion recognition and reasoning (see, e.g., Huang et al., 2025). However, there are at least two limitations: (a) if typical patterns and scenarios change (and they change quickly in contemporary societies), AI tools need to be retrained, and (b) in high-context cultures, these tools likely will not be so effective.

Another type of AI technology to be mentioned is the simulation of emotions by AI agents. It is promising in cases of simple, typical and numerous manifestations of emotions (for example, panic on the stock market or the spread of rumors).

So, with social scientists' high 'qualitative literacy' and 'computational thinking' (Wing, 2006), prospects of AI in emotion analysis are substantial. However, we should keep in mind that in human-AI tandem understanding is carried out by a human, and only a human (Esposito, 2022).

Conclusion

The position of "good bad theory" problem in contemporary emotion analytics in sociology and computer science seems to be quite ambivalent.

On the one hand, our analysis has revealed the limitations and perils of "good bad theories". They *do* systematically ignore the problem of meaning, seeking to universally formalize what formalizable only in certain contexts and under specific conditions. Unlike simple critiques of reductionism or complaints about oversimplification, the 'good bad theory' framework identifies a specific mechanism. That is, the requirements of mathematical formalization create selection pressures that systematically favor certain aspects of phenomena (those open to quantification) while excluding others (those calling for interpretation and understanding). We have also shown that, under the contemporary capitalism, there is a danger that "good bad theories" could become a self-fulfilling prophecy.

[15] A common place illustration here is AI models that have 'learned' to write scripts at the level of a mediocre Hollywood worker. Another example is special programs for academia, such as *NotebookLM*, that can create a popular science presentation or blog following very recognizable, very regular scenarios that include repeatable intonation and lexical emotional markers.

On the other hand, the "good bad theory" examples that were discussed are really not so bad. Despite their limitations, Collins' and Russell's frameworks are far superior to most conceptualizations revealed during quantitative analysis of scholarly publications. In this regard, the more pressing problem today is not "good bad theory", but the general lack of interest in conceptual understanding of emotions from both sociologists and computer scientists. In particular, it is urgent to reflect on limitations of modern AI tools that are widely used for recognizing and interpreting emotions.

Regarding this, where can we put a line between a good theory and a "good bad theory"? When there is "too much" formalization? We believe that the answer to this question depends on the context in which the theory is applied. For instance, in many cases, the Collins' theory of interaction rituals is simply a good theory. But, say, in a study of emotional conflict between qualitatively different demands other approaches, like Nussbaum's neo-Stoic framework, will be needed. Thus, the problem of "good bad theory" arises when the theory is used inappropriately, and more suitable alternatives are ignored.

In an attempt to generalize our analysis, let us formulate several fundamental arguments and counterarguments of the "good bad theory" problem.

The first argument we call the **'Accuracy Argument'**. In a nutshell, it contends that mathematical formalization helps to establish logical and mathematical accuracy that those who study emotion should embrace. The point is that formal models enable precise hypothesis testing, systematic empirical validation, and aggregate knowledge development that interpretive approaches cannot achieve. Computational emotion analytics extends these advantages by processing large datasets, identifying indirect patterns in emotional behavior, and enabling real time analysis impossible through traditional methods.

This argument creates the **'Phenomenological Counterargument'.** It maintains that mathematical formalization inevitably rejects the empirical dimensions that constitute emotional reality. Computational approaches may achieve predictive accuracy, but at the same time, they miss entirely what makes passions emotionally significant for humans in their social activities. The qualitative distinctions between different emotional states, the temporal flow of emotional experience, and the moral dimensions of sensitive evaluations cannot be captured through quantitative measures without fundamental distortion.

The second argument might be called the **'Pragmatic Argument'.** It suggests that theoretical adequacy should be evaluated relative to specific research purposes rather than being oriented towards abstract philosophical criteria. If the goal is behavioral prediction, computational approaches may prove superior even if they sacrifice experimental complexity. Emotion analytics for AI systems requires operational definitions and algorithmic implementation that may be more important than phenomenological accuracy.

The argument that appeals to pragmatism causes the **'Constitutive Counterargument'.** It claims that methodological choices condition what counts as legitimate knowledge and valid research questions. Computational approaches tend to privilege phenomena that can be quantified and modeled. This selection bias may systematically distort understanding of human emotional experience by focusing attention on measurable behaviors while ignoring interpretive meanings.

The third is the **'Innovation Argument'.** It basically suggests that new AI technologies represent natural extensions of scientific methodology that enable new forms of emotional analysis. Just as statistical methods transformed social research in the twentieth century, machine learning and computational modeling may enable insights impossible through traditional approaches. Resistance to computational methods in a way reflects disciplinary conservatism rather than principled theoretical commitment.

This argument disputes the **'Ontological Counterargument'.** It maintains that the emotional reality of human beings possesses in its essence properties that cannot be decomposed into computational primitives without fundamental transformation. Emotional meanings, cultural values, and intersubjective understandings of human behavior exist at levels of analysis that resist algorithmic processing. The attempt to formalize these phenomena necessarily transforms them into something categorically different from what they are in human experience. This ontological limitation cannot be overcome through technical advancement but reflects fundamental differences between computational processing and conscious practices.

Future directions

The "good bad theory" problem has direct implications for the future development of emotion analytics and computational social science more broadly. Rather than viewing the limitations identified in this analysis as temporary obstacles to be overcome through technical advancement, we argue they represent structural features of the relationship between computational formalization and human emotional experience.

The first suggestion concerns the scope and appropriate applications of computational emotion analytics. Our analysis suggests that AI systems may be most effective when applied to specific, well-defined emotional behaviors rather than attempting to model the full complexity of human emotional experience.

The second implication involves the need for hybrid methodologies that can combine computational analysis with interpretive insight. Such approaches might use AI systems to identify patterns in emotional behavior that interpretive analysis could then investigate for meaning and significance. Conversely, ethnographic and phenomenological insights might inform the design and validation of computational models.

The third suggestion concerns the social and ethical dimensions of emotion analytics applications. As AI systems increasingly mediate human emotional experience through recommendation algorithms and social media platforms, the theoretical frameworks used to design these systems become consequential for emotional life itself. The "good bad theory" problem thus extends beyond academic debate to questions about how emotional experience should be monitored, understood, and managed in the age of online culture and AI penetration into the everyday life of society.

Authors contribution All authors contributed to the conception and design of this study.
Material preparation, data collection, and analysis were performed by both authors, Andrey V. Rezaev and Natalia D. Tregubova.

Funding No funding received for this research from any agency.

Data availability The data that support the findings of this study are derived from secondary sources publicly available online.

Declarations

Ethical approval N/A.

Conflict of interest Author Andrey V. Rezaev declares that he has no conflicts of interest. Author Natalia D.
Tregubova declares that she has no conflicts of interest.

References

Abend, G. (2008). The meaning of 'theory.' *Sociological Theory, 26*(2), 173–199. https://doi.org/10.1111/j.1467-9558.2008.00324

Alexander, J. C. (2006). Cultural Pragmatics: Social Performance between Ritual and Strategy. In: Alexander, J. C., Giesen, B., & Mast, J. L. (Eds.). (2006). *Social performance: Symbolic action, cultural pragmatics, and ritual*. Cambridge University Press. P. 29–90.

Assunção, G., Patrão, B., Castelo-Branco, M., & Menezes, P. (2022). An overview of emotion in artificial intelligence. *IEEE Transactions on Artificial Intelligence*, *3*(6), 867–886. https://doi.org/10.1109/TAI.2022.3159614

Aydt, H., Lees, M., Luo, L., Cai, W., Low, M. Y. H., & Kadirvelen, S. K. (2011, August). A computational model of emotions for agent-based crowds in serious games. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 2, pp. 72–80). IEEE. https://doi.org/10.1109/WI-IAT.2011.154

Bainbridge, W. S., Brent, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J. (1994). Artificial social intelligence. *Annual Review of Sociology, 20*(1), 407–436. https://doi.org/10.1146/annurev.so.20.080194.002203

Baker, W. E. (2019). Emotional energy, relational energy, and organizational energy: Toward a multilevel model. *Annual Review of Organizational Psychology and Organizational Behavior, 6*(1), 373–395. https://doi.org/10.1146/annurev-orgpsych-012218-015047

Barbalet, J. M. (1998). *Emotion, social theory, and social structure: A macrosociological approach*. Cambridge University Press.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Bazan Royuela, D. (2025). *Algorithmic Interaction Ritual Chains on TikTok: Scrolling through Feedback Loops* (Doctoral dissertation, Lund University).

Bericat, E. (2014). The socioemotional well-being index (SEWBI): Theoretical framework and empirical operationalisation. *Social Indicators Research, 119*(2), 599–626. https://doi.org/10.1007/s11205-013-0528-z

Bianchi, A. J., & Shelly, R. K. (2020). Sentiments as status processes? A theoretical reformulation from the expectation states tradition. *Sociological Theory, 38*(3), 217–235. https://doi.org/10.1177/0735275120941176

Bosse, T., Hoogendoorn, M., Klein, M. C., Treur, J., Van Der Wal, C. N., & Van Wissen, A. (2013). Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions. *Autonomous Agents and Multi-Agent Systems, 27*(1), 52–84. https://doi.org/10.1007/s10458-012-9201-1

Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., & Griffiths, T. L. (2019, May). Cognitive model priors for predicting human decisions. In *International conference on machine learning* (pp. 5133–5141). PMLR. https://doi.org/10.48550/arXiv.1905.09397

Boyns, D., & Luery, S. (2015). Negative emotional energy: A theory of the dark-side of interaction ritual chains. *Social Sciences*, *4*(1), 148–170. https://doi.org/10.3390/socsci4010148

Brint, S. (2025). High-leverage sociological concepts and the progress of theory, part one. *Theory and Society*. https://doi.org/10.1007/s11186-025-09607-5
Buber, M. (1970). *I and Thou*. T&T Clark.
Cabanas, E., & Illouz, E. (2019). *Manufacturing happy citizens: How the science and industry of happiness control our lives*. Wiley.
Carroll, M., Foote, D., Siththaranjan, A., Russell, S., & Dragan, A. (2024). AI alignment with changing and influenceable reward functions. *arXiv preprint arXiv: 2405 17713*. https://doi.org/10.48550/arXiv.2405.17713
Castelfranchi, C. (2000). Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics and Information Technology*, *2*(2), 113–119. https://doi.org/10.1023/A:1010025403776
Cayla, J., & Auriacombe, B. (2025). Emotional energy: When customer interactions energize service employees. *Journal of Marketing*, *89*(1), 1–18. https://doi.org/10.1177/00222429241260637
Centola, D. (2018). *How behavior spreads: The science of complex contagions* (Vol. 3). Princeton University Press.
Chen, Z., & Skey, M. (2024). I produce songs for her… In this way, I gradually know her more. The more I know her, the more I like her': Using Collins' model of interactive ritual chains to study the case of virtual idol fandom in China. *Convergence, 30*(2), 841–859. https://doi.org/10.1177/13548565241246045
Christov-Moore, L., Reggente, N., Vaccaro, A., Schoeller, F., Pluimer, B., Douglas, P. K., Iacoboni, M., Man, K., Damasio, A., & Kaplan, J. T. (2023). Preventing antisocial robots: A pathway to artificial empathy. *Science Robotics, 8*(80), Article eabq3658. https://doi.org/10.1126/scirobotics.abq3658
Collins, H. (2018). *Artifictional intelligence: against humanity's surrender to computers*. Wiley.
Collins, H. (2021). The science of artificial intelligence and its critics. *Interdisciplinary Science Reviews*, *46*(1–2), 53–70.
Collins, H. M. (1996). Embedded or embodied? A review of Hubert Dreyfus' what computers still can't do. *Artificial Intelligence, 80*(1), 99–118.
Collins, R. (1992). Can sociology create an artificial intelligence. *Sociological Insight: An Introduction to Non-Obvious Sociology* (pp. 155–184). Oxford University Press.
Collins, R. (1994, June). Why the social sciences won't become high-consensus, rapid-discovery science. *Sociological forum* (Vol. 9, pp. 155–177). Kluwer Academic Publishers-Plenum. 2https://doi.org/10.1007/BF01476360
Collins, R. (2004). *Interaction ritual chains*. Princeton University Press.
Collins, R. (2012). C-escalation and D-escalation: A theory of the time-dynamics of conflict. *American Sociological Review, 77*(1), 1–20. https://doi.org/10.1177/0003122411428221
Collins, R. (2020). Social distancing as a critical test of the micro-sociology of solidarity. *American Journal of Cultural Sociology, 8*(3), 477–497. https://doi.org/10.1057/s41290-020-00120-z
Conitzer, V., Freedman, R., Heitzig, J., Holliday, W. H., Jacobs, B. M., Lambert, N., & Zwicker, W. S. (2024). Social choice should guide ai alignment in dealing with diverse human feedback. *arXiv preprint arXiv: 2404 10271*. https://doi.org/10.48550/arXiv.2404.10271
Cortiñas-Lorenzo, K., & Lacey, G. (2023). Toward explainable affective computing: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 13101–13121. https://doi.org/10.1109/TNNLS.2023.3270027
Cottingham, M. D. (2012). Interaction ritual theory and sports fans: Emotion, symbols, and solidarity. *Sociology of Sport Journal, 29*(2), 168–185. https://doi.org/10.1123/ssj.29.2.168
Di Iorio, F. (2024). Methodological individualism and agent-based computational simulation: A reply to Kincaid and Zahle. *Social Science Information, 63*(2), 155–167. https://doi.org/10.1177/05390184241258370
DiMaggio, P., Bernier, C., Heckscher, C., & Mimno, D. (2018). Interaction Ritual Threads: Does IRC Theory Apply Online? *Ritual, emotion, violence* (pp. 81–124). Routledge.
Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.
Dreyfus, H. L. (1996). Response to my critics. *Artificial Intelligence*, *80*(1), 171–191.
Dreyfus, H. L. (2012). A history of first step fallacies. *Minds and Machines, 22*(2), 87–99. https://doi.org/10.1007/s11023-012-9276-0
Droppe, A. (2022, July). Emotional Ambience in Classroom Interaction Rituals. In *Conference Proceedings. The Future of Education 2022*.
Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology, 46*(1), 61–81. https://doi.org/10.1146/annurev-soc-121919-054621
Effler, E. S. (2010). *Laughing saints and righteous heroes: Emotional rhythms in social movement groups*. University of Chicago Press.

Ekman, P. (2003). *Emotions revealed: Understanding faces and feelings*. Times books.
Emmons, S., Oesterheld, C., Conitzer, V., & Russell, S. (2024). Observation interference in partially observable assistance games. *arXiv preprint arXiv: 2412 17797*. https://doi.org/10.48550/arXiv.2412.17797
Empathic, A. I., & can't get under the skin. (2024). *Nature Machine Intelligence*, 6(5), 495. https://doi.org/10.1038/s42256-024-00850-6.
Ermakoff, I. (2017). Shadow plays: Theory's perennial challenges. *Sociological Theory, 35*(2), 128–137. https://doi.org/10.1177/0735275117709774
Esposito, E. (2022). *Artificial communication: How algorithms produce social intelligence*. MIT Press.
Ferrada, F., & Camarinha-Matos, L. M. (2024, September). Emotions in human-ai collaboration. In *Working Conference on Virtual Enterprises* (pp. 101–117). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71739-0_7
Garcia, D., Garas, A., & Schweitzer, F. (2016). An agent-based modeling framework for online collective emotions. *Cyberemotions: Collective emotions in cyberspace* (pp. 187–206). Springer International Publishing. https://doi.org/10.1007/978-3-319-43639-5_10
Grønnestad, T. E., Sagvaag, H., & Lalander, P. (2020). Interaction rituals in an open drug scene. *Nordic Studies on Alcohol and Drugs, 37*(1), 86–98. https://doi.org/10.1177/1455072519882784
Haidt, J. (2024). *The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness*. Penguin.
Hansson Wahlberg, T. (2019). Why the social sciences are irreducible. *Synthese*, *196*(12), 4961–4987. https://doi.org/10.1007/s11229-017-1472-2
Healy, K. (2017). Fuck nuance. *Sociological Theory, 35*(2), 118–127. https://doi.org/10.1177/0735275117709046
He, J. K., Wallis, F. P., Gvirtz, A., & Rathje, S. (2024). Artificial intelligence chatbots mimic human collective behaviour. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12764
Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling*. University of California Press.
Hofstede, G. J., Frantz, C., Hoey, J., Scholz, G., & Schröder, T. (2021). Artificial sociality manifesto. *Review of Artificial Societies and Social Simulation*.
Ho, M. K., & Griffiths, T. L. (2022). Cognitive science as a source of forward and inverse models of human decisions for robotics and control. *Annual Review of Control, Robotics, and Autonomous Systems, 5*(1), 33–53. https://doi.org/10.1146/annurev-control-042920-015547
Huang, X., Zhou, Y., Li, J., Lu, S., & Wang, S. (2025, October). EmoDETective: Detecting, Exploring, and Thinking Emotional Cause in Videos. In *Proceedings of the 33rd ACM International Conference on Multimedia* (pp. 5735–5744). https://doi.org/10.1145/3746027.3755522
Husserl, E. (1991). *On the phenomenology of the consciousness of internal time*. Kluwer Academic.
Illouz, E. (2007). *Cold intimacies: The making of emotional capitalism*. Polity.
Illouz, E. (Ed.). (2017). *Emotions as commodities: Capitalism, consumption and authenticity*. Routledge.
Illouz, E., & Kotliar, D. M. (2022). Capitalist subjectivity, Tinder, and the emotionalization of the web. *The Routledge handbook of digital consumption* (pp. 229–240). Routledge.
Johannessen, L. E. (2023). Interaction rituals and technology: A review essay. *Poetics, 98*, Article 101765. https://doi.org/10.1016/j.poetic.2023.101765
Joo, D., Cho, H., Woosnam, K. M., & Suess, C. (2023). Re-theorizing social emotions in tourism: Applying the theory of interaction ritual in tourism research. *Theoretical Advancement in Social Impacts Assessment of Tourism Research* (pp. 138–153). Routledge.
Kirk, H. R., Gabriel, I., Summerfield, C., Vidgen, B., & Hale, S. A. (2025). Why human–AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications, 12*(1), 1–9. https://doi.org/10.1057/s41599-025-04532-5
Latif, S., Ali, H. S., Usama, M., Rana, R., Schuller, B., & Qadir, J. (2022). AI-based emotion recognition: Promise, peril, and prescriptions for prosocial path. *arXiv preprint arXiv*, 221107290. https://doi.org/10.48550/arXiv.2211.07290
Lian, Z., Liu, R., Xu, K., Liu, B., Liu, X., Zhang, Y., & Tao, J. (2025, October). Mer 2025: When affective computing meets large language models. In *Proceedings of the 33rd ACM International Conference on Multimedia* (pp. 13837–13842). https://doi.org/10.1145/3746027.3762007
Lidayan, A., Dennis, M., & Russell, S. (2024). BAMDP shaping: a unified theoretical framework for intrinsic motivation and reward shaping. *arXiv e-prints*, *arXiv-2409*. https://doi.org/10.48550/arXiv.2409.05358
Lupton, D. (2016). *The quantified self*. Wiley.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, *11*(1), 21–46.
McCaffree, K., & Shults, F. L. (2022). Distributive effervescence: Emotional energy and social cohesion in secularizing societies. *Theory and Society, 51*(2), 233–268. https://doi.org/10.1007/s11186-021-09470-0
Mears, A. (2017). Puzzling in sociology: On doing and undoing theoretical puzzles. *Sociological Theory*, *35*(2), 138–146. https://doi.org/10.1177/0735275117709775
Merleau-Ponty, M. (1945). *Phenomenology of perception*. Routledge.
Methot, J. R., Rosado-Solomon, E. H., Downes, P. E., & Gabriel, A. S. (2021). Office chitchat as a social ritual: The uplifting yet distracting effects of daily small talk at work. *Academy of Management Journal, 64*(5), 1445–1471. https://doi.org/10.5465/amj.2018.1474
Midgley, M. (2002). *Beast and man: The roots of human nature*. Routledge.
Minsky, M. (1986). *Society of mind*. Simon and Schuster.
Mizrahi Werner, J., Liebst, L. S., & Demant, J. (2025). Beyond bodily co presence: A micro sociological study of online interaction rituals. *Symbolic Interaction, 48*(1), 46–68. https://doi.org/10.1002/symb.1206
Mökander, J., & Schroeder, R. (2022). AI and social theory. *AI & Society, 37*(4), 1337–1351. https://doi.org/10.1007/s00146-021-01222-z
Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & Society, 37*(4), 1353–1359. https://doi.org/10.1002/symb.1206
Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811
Mozikov, M., Orekhov, D., Nasonov, I., Baltsat, K., Pedashenko, V., Abramov, D., & Makarov, I. (2025, October). HL-EAI: A Multimodal Framework Enabling Emotional Reciprocity in Human-AI Strategic Decision-Making. In *Proceedings of the 33rd ACM International Conference on Multimedia* (pp. 13501–13503). https://doi.org/10.1145/3746027.3754468
Murphy, R. R. (2025). Would a robot ever get angry enough to attack a person? *Science Robotics, 10*(98), Article eadv3128. https://doi.org/10.1126/scirobotics.adv3128
Nerantzi, E. (2025). "All AIs are psychopaths"? The scope and impact of a popular analogy. *Philosophy & Technology, 38*(1), Article 27. https://doi.org/10.1007/s13347-025-00856-x
Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
Ojha, S., Vitale, J., & Williams, M. A. (2021). Computational emotion models: A thematic review. *International Journal of Social Robotics, 13*(6), 1253–1279. https://doi.org/10.1007/s12369-020-00713-1
Parker, J. N., Cardenas, E., Dorr, A. N., & Hackett, E. J. (2020). Using sociometers to advance small group research. *Sociological Methods & Research, 49*(4), 1064–1102. https://doi.org/10.1177/0049124118769091
Pei, G., Li, H., Lu, Y., Wang, Y., Hua, S., & Li, T. (2024). Affective computing: Recent advances, challenges, and future trends. *Intelligent Computing, 3*, Article 0076. https://doi.org/10.34133/icomputing.0076
Picard, R. W. (1997). *Affective computing*. MIT Press.
Pine, B. J., & Gilmore, J. H. (2011). *The experience economy*. Harvard Business.
Plaut, B., Lievano-Karim, J., Zhu, H., & Russell, S. (2025). Safe Learning Under Irreversible Dynamics via Asking for Help. *arXiv preprint arXiv:2502.14043*.
Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., & Erev, I. (2025). Predicting human decisions with behavioural theories and machine learning. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-025-02267-6
Pollak, M., Salfinger, A., & Hummel, K. A. (2022). Teaching drones on the fly: Can emotional feedback serve as learning signal for training artificial agents? *arXiv preprint arXiv:2202.09634.* https://doi.org/10.48550/arXiv.2202.09634
Puetz, K. (2024). Relational Durkheim: Homo duplex as the foundation of a formalist cultural sociology. *Sociological Theory, 42*(3), 205–230. https://doi.org/10.1177/07352751241241517
Rawls, A. W. (1987). The interaction order sui generis: Goffman's contribution to social theory. *Sociological Theory*. https://doi.org/10.2307/201935
Rawls, A. W. (1989). Interaction order or interaction ritual: Comment on Collins. *Symbolic Interaction, 12*(1), 103–109. https://doi.org/10.1525/si.1989.12.1.103
Rawls, A. W. (2000). "Race" as an interaction order phenomenon: W.E.B. Du Bois's "double consciousness" thesis revisited. *Sociological Theory, 18*(2), 241–274. https://doi.org/10.1111/0735-2751.00097

Rezaev, A. V., & Tregubova, N. D. (2022). Emotional utilitarianism" and the frontiers of artificial intelligence evolvement. *Monitoring Public Opinion: Economy, Society, Change, 2*, 4–23. https://doi.org/10.14515/monitoring.2022.2.2127

Rezaev, A. V., & Tregubova, N. D. (2025). Looking at human-centered artificial intelligence as a problem and prospect for sociology: An analytic review. *Current Sociology, 73*(1), 120–138. https://doi.org/10.1177/00113921231211580

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Russell, S. (2022). Human-Compatible Artificial Intelligence. *Human-like machine intelligence*, *1*, 3–22.

Selinger, E., Dreyfus, H., & Collins, H. (2007). Interactional expertise and embodiment. *Studies in History and Philosophy of Science*, *38*(4), 722–740.

Sethi, R. J., Qiu, H., Courchaine, C., & Iacoboni, J. (2025, November). Do LLMs Dream of Electric Emotions? Towards Quantifying Metacognition and Generalizing the Teacher-Student Model Using Ensembles of LLMs. *In Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (pp. 5223–5227). https://doi.org/10.1145/3746252.3760839

Shteynberg, G., Halpern, J., Sadovnik, A., Garthoff, J., Perry, A., Hay, J., & Fairweather, A. (2024). Does it matter if empathic AI has no empathy? *Nature Machine Intelligence*, *6*(5), 496–497. https://doi.org/10.1038/s42256-024-00841-7

Shults, F. L. (2025). Simulating theory and society: How multi-agent artificial intelligence modeling contributes to renewal and critique in social theory. *Theory and Society*. https://doi.org/10.1007/s11186-025-09606-6

Sloman, A., & Croucher, M. (1981). Why Robots will Have Emotions. *Proceedings IJCAI*, 1–10.

Small, M. L., & Calarco, J. M. (2022). *Qualitative literacy: A guide to evaluating ethnographic and interview research*. University of California Press.

Smelser, N. J. (1976). *Comparative methods in the social sciences*. Prentice-Hall.

Stets, J. E., & Turner, J. H. (Eds.). (2006). *The Handbook of the Sociology of Emotions*. Springer.

Sweet, P. L. (2023). The particular and the provincial: Thinking with Dorothy Smith's phenomenology. *Sociological Theory, 41*(4), 290–300. https://doi.org/10.1177/07352751231197833

Tavory, I., & Hoynes, N. (2025). Order and potentiality in interaction ritual theory. *Sociological Theory, 43*(1), 52–66. https://doi.org/10.1177/07352751251315940

Tay, Y., Ong, D., Fu, J., Chan, A., Chen, N., Tuan, L. A., & Pal, C. (2020, July). Would you rather? a new benchmark for learning machine alignment with cultural values and social preferences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5369–5373).

The Good Bad Theory Case in Emotion Analytics: AI's Potential and Limits for Social Theory.

Tregubova, N., & Nee, M. (2020). Beyond nations and nationalities: Discussing the variety of migrants' identifications in Russian Social Media. *Changing Societies & Personalities, 4*(3), 323–349. https://doi.org/10.15826/csp.2020.4.3.104

Turner, J. H. (2002). *Face to face: Toward a sociological theory of interpersonal behavior*. Stanford University Press.

Turner, J. H. (2009). The sociology of emotions: Basic theoretical arguments. *Emotion Review, 1*(4), 340–354. https://doi.org/10.1177/1754073909338305

Turner, J. H. (2025). The stratification of emotions as a driving force in human societies. *Theory and Society*. https://doi.org/10.1007/s11186-025-09632-4

Turner, S. (2013). The argument of explaining the normative. *Revista Internacional de Sociología, 71*(1), 192–194. https://doi.org/10.3989/ris.2013.i1.503

Turner, S. P. (2010). *Explaining the normative*. Polity.

Virca, T. (2024). *The Moral Machine Experiment: Predicting Moral Decision-making Based on Personal Values* (Doctoral dissertation, Tilburg university).

Voyer, A., Kline, Z. D., Danton, M., & Volkova, T. (2022). From strange to normal: Computational approaches to examining immigrant incorporation through shifts in the mainstream. *Sociological Methods & Research, 51*(4), 1540–1579. https://doi.org/10.1177/00491241221122596

Wang, B., Song, Y., Cao, J., Yu, P., Guo, H., & Li, Z. (2025, November). DinoCompanion: An Attachment-Theory Informed Multimodal Robot for Emotionally Responsive Child-AI Interaction. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (pp. 6084–6092). https://doi.org/10.1145/3746252.3761511

Wan, Y., Wu, J., Abdulhai, M., Shani, L., & Jaques, N. (2025). Enhancing personalized multi-turn dialogue with curiosity reward. *arXiv preprint arXiv: 2504 03206*. https://doi.org/10.48550/arXiv.2504.03206

Weininger, E. B., Lareau, A., & Lizardo, O. (Eds.). (2018). *Ritual, emotion, violence: Studies on the micro-sociology of Randall Collins*. Routledge.

Whitmeyer, J., & Hopcroft, R. L. (2025). What makes a good social science theory, and why the evolutionary model of the actor is one. *Theory and Society*. https://doi.org/10.1007/s11186-025-09620-8

Winch, P. (1958). *The idea of a social science and its relation to philosophy*. Routledge.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wolfe, A. (1991). Mind, self, society, and computer: Artificial intelligence and the sociology of mind. *American Journal of Sociology*, *96*(5), 1073–1096. https://doi.org/10.1086/229649

Xiang, K., Huang, W. J., Gao, F., & Lai, Q. (2022). COVID-19 prevention in hotels: Ritualized host-guest interactions. *Annals of Tourism Research, 93*, Article 103376. https://doi.org/10.1016/j.annals.2022.103376

Yashinski, M. (2024). Robot behavior that can adapt to user interaction. *Science Robotics*. https://doi.org/10.1126/scirobotics.adr9645

Yim, J., Foulk, T. A., Klotz, A. C., & Schilpzand, P. (2025). Is Everyone Having a Good Time? The Effects of Complex Organizational Rituals on Employee Engagement and Behavior. *Journal of Management*, 01492063251355251. https://doi.org/10.1177/01492063251355525

Zhang, J., Zhang, D., & Dai, G. (2024). Mechanisms of emotional experiences of online spectators of E-sports events from the perspective of interactive ritual chain. *Communication and Sport, 12*(6), 1054–1074. https://doi.org/10.1177/21674795241227771

Zhu, Y., Lyu, Y., Yu, Z., Shao, R., Zhou, K., & Nie, L. (2025, October). EmoЬIym: A symbiotic framework for unified emotional understanding and generation via latent reasoning. In *Proceedings of the 33rd ACM International Conference on Multimedia* (pp. 5451–5460). https://doi.org/10.1145/3746027.3754549

Authors and Affiliations

Andrey V. Rezaev[1,2] · **Natalia D. Tregubova**[3]

✉ Andrey V. Rezaev
rezaev@hotmail.com

Natalia D. Tregubova
n.tregubova@spbu.ru

[1] Tashkent State University of Economics, Tashkent, Uzbekistan

[2] Kimyo International University, Tashkent, Uzbekistan

[3] Faculty of Sociology, St. Petersburg State University, St. Petersburg, Russia