Position paper presented at CONVERSATIONS 2023 – the 7th International Workshop on Chatbot Research and Design, hosted by the University of Oslo, Norway, November 22-23, 2023.

More than Big Five? Towards Modelling and Defining Artificial Personality for Conversational Agents

Alexander Dregger[1]

[1] FZI Research Center for Information Technology,
Haid-und-Neustraße 10-14, 76131 Karlsruhe, Germany
dregger@fzi.de

Abstract. Research on artificial personality in conversational agents (CA) is expanding, but there remains a lack of a clear definition, a valid model, and measurement instruments, such as questionnaires, which hinders implementation and evaluation. Therefore, this paper introduces a more comprehensive definition of artificial personality, drawing from insights in personality theories from psychology. Additionally, in a mixed-method study, experts (N = 16) with interdisciplinary backgrounds in conversational design and robotics evaluated current theoretical approaches for conceptualizing artificial personality by completing a questionnaire. This position paper focusses exclusively on the results of the questionnaire of the mixed-method study. The findings indicate that models initially designed for humans, like the Big Five Model (BFM), can form the foundation for an artificial personality model. However, adjustments are necessary for the BFM in terms of its dimensions, facets, and items. Alternative personality models, such as the Business-focused Inventory of Personality (BIP), should also be considered to enhance the model's validity. Based on these results, the paper presents an initial model for artificial personality, providing a basis for further development and evaluation. The objective is to refine the model, create a questionnaire, and test it with users.

Keywords: Artificial personality, CA, Chatbots, Robotics, Expert study, Big Five Theory, Personalisation, User experience.

1 Introduction

When people interact with CA, they regard them as social agents [14] and attribute personality to them [7]. Such systems are language based and language is a form of communication that humans expect other humans to use [7]. Developers of CA can systematically design the personality of such systems by using modalities such as language, voice pitch, or the avatar of the agent, to change users' perception of personality [13]. The personality is either designed similar to that of the user, as the similarity attraction hypothesis [14] assumes, or it can be complementary to the user personality [9]. The latter means, for example, that a person with little conscientiousness would prefer to interact with a highly conscientious chatbot giving precise advice concerning taxes to ensure a correct tax declaration. The design influences different aspects of the

user experience such as the amount of interaction, acceptance, trust or purchase decisions [19, 22, 24].

Research dealing with both CA and robotics focusses independently on modelling the concept of artificial personality, thus naming it by different terms, such as "chatbot personality" [20] or "robot personality" [13]. This usage of different terms has led to research results that are not aware of each other, causing scattered results [1]. For example, a review on verbal cues to express personality in the field of robotics did not consider at all or to a very limited extent the results for chatbots [13] and a similar result can be found vice versa in a review in the field of chatbots [1]. In order to create a unifying concept, this paper assumes that behind the constructs "robot personality" and "chatbot personality" lies the same construct which can be more generally named artificial personality. There is no clear consensus on how to define and conceptualize artificial personality, which impedes the implementation of personality into CA and leads to a lack of valid measurement instruments to evaluate such implementations. Hence, this paper seeks to clarify the concept and focusses on two research questions:

- **RQ1.** How can artificial personality be defined?
- **RQ2.** How can artificial personality be modelled?

The paper will analyse psychological insights to define artificial personality, present expert study findings to evaluate current conceptual approaches, and propose an initial model.

2 Background and Related Work

2.1 Defining Artificial Personality

There are very few definitions of the concept artificial personality in the literature, and they remain rather superficial. For instance, Endres [4] describes the perception of "personality characteristics in nonhumans" by users as the core of artificial personality. This definition raises several unanswered questions, such as: What are personality characteristics, and how can this definition be distinguished from definitions of animal personality?

Psychological research offers insights to clarify this concept. Personality psychology relies on various paradigms, including the psychodynamic, learning theory-based, humanistic, cognitive, dispositional, biologic, and transactional perspectives [18]. Those paradigms vary in aptitude for the development of a model and definition of artificial personality. Research in robotics and in CA has mainly used the psychodynamic and the dispositional approach for the implementation of personality into machines [1, 13]. This paper focusses on the dispositional approach as it is the most popular one concerning its usage in CA and robotics [1, 13].

The dispositional approach is very intuitive for non-psychologists, as it assumes that people have stable characteristics or dispositions. Other approaches, such as the transactional paradigm, are far more complex [18] and thus more difficult to apply for non-psychologists. In contrast to the psychodynamic approach and to folk psychology, these characteristics are precisely described and they are based on empirical data [18]. The Big Five Model (BFM) is very popular in this context. Nevertheless, it has not been

specifically developed for this context or has been adapted at all, and thus, the validity of this approach remains unclear. It rests on the psycholexical method [18], assuming that important individual differences have been encoded into language and that questionnaires were established on adjective lists from dictionaries. Using factor analysis, five factors (also called dimensions) were deduced from the questionnaire data to describe individual differences in human personality: Openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (nowadays called emotional stability). Many correlations between personality traits or dimensions and behaviour, such as verbal and non-verbal language, have been found [15]. Such correlations were also found for personality and user behaviour in CA and robots as it impacts trust, purchase decisions, and acceptance [19, 22, 24]. Ahmad et al. [1] have created a framework describing the relationship of linguistic cues and the Big Five dimensions based on empirical results from research. Also, in the field of robotics, the BFM is very popular, as 24 of the 40 studies reviewed by Mou et al. [13] applied the BFM in order to design the personality of a robot.

The definition of artificial personality in this paper follows the dispositional paradigm and adapts a definition of personality developed for humans [17]: Personality for a human is the "whole and organisation of relatively stable tendencies and patterns of experience and behaviour within one person (distinguishing it from other persons)". This definition is modified for artificial personality:
Artificial personality describes the relatively stable tendencies and patterns of behaviour of an AI-based machine that can be designed by developers and designers via different modalities, such as language, creating the impression of individuality of a humanized social agent when users interact with the machine.

Three aspects were adapted comparing human and artificial personality definition:

First, artificial personality only refers to the behaviour and not to the inside experience a human has. A precondition for experience is having consciousness and research is intensively debating whether AI-based machines have consciousness [10]. Several criteria have been mentioned for consciousness, such as self-awareness, or an autobiographic self [10]. This paper questions whether AI-based machines have abilities such as an autobiographic self, and thus, it is not assumed that they can actually experience like a human does.

Second, machine personality is not an inherent trait possessed by the machine itself. The perception of personality only forms in the user's mind during interactions with the machine. This attribution of personality appears to depend on expressive modalities, such as appearance or language. Designers can systematically craft personality through these modalities, deliberately shaping the user's perception of personality. This systematic design typically aims to enhance interaction by humanizing the communication process, potentially boosting trust in the machine [19]. Nevertheless, it remains unclear at which point the usage of personality leads to an uncanny valley effect [12].

Third, the design of personality helps to create the impression of individuality of the humanized social agent. Thus, a chatbot, for example, although fulfilling the same tasks as another chatbot, is perceived differently due to the way it uses modalities such as language based on its artificial personality. This is important, for example, for companies to distinguish themselves from their competitors.

Apart from those adaptations, the aspect of relatively stable tendencies and patterns is identical. Relatively stable refers to the relative consistency of personality in different situations (trans-situational consistency) and over time (temporal stability) [18].

2.2 Modelling Artificial Personality

Based on the definition of artificial personality a model of artificial personality has to be developed. Two approaches have been used to develop models of artificial personality so far: The first one is the *human-centred approach* transferring personality models developed for humans to machines. Ahmad et al. [1] used the BFM in an identical manner as it is used for humans to describe which language cues could be used to express which personality dimension in CA. The second one is the *machine-centred approach* which questions whether human personality models can be transferred to machines [13, 23] and tries to develop machine-adapted models. Völkel et al. [23], for example, analysed with NLU-techniques reviews of CA regarding adjectives describing personality, conducted surveys, experiments and collected BFM adjectives to deduce ten dimensions describing personality in CA: Confrontational, dysfunctional, serviceable, unstable, approachable, social-entertaining, social-inclined, social-assisting, self-conscious and artificial.

The two approaches contain several weaknesses: The human-centred approach using the BFM is questionable as it is unclear whether personality dimensions developed for humans apply to machines [23]. Furthermore, research dealing with BFM has criticised the concept on grounds of personality psychology. Several studies have found that there are more than just five dimensions or factors of personality. Paunonen and Jackson [16] and Rauthmann [18] describe ten additional dimensions that have been deduced in several studies: Religiousness, manipulation, honesty, attraction, thrift, traditionality/conservatism, gender, arrogance, humour and risk. Thus, such additional dimensions of personality could be especially relevant in machines as they are an emergent entity. Furthermore, the validity can be questioned as the BFM is a general personality model. CA usual fulfil specific tasks [5] and thus, personality models specifically developed for the job context such as the Business-focused Inventory of Personality (BIP) [8] could be more suitable. Finally, the dimensions described by the BFM are very abstract. Hence, it is difficult for a designer to implement abstract dimensions such as conscientiousness. Consequently, the granularity of the dimensions needs to be increased. Research in personality psychology introduced the concept of facets for the dimensions to describe the personality dimensions more specifically. There are two approaches for the BFM to describe facets. For each dimension, either six or three facets are attributed. Figure 1 shows the three-facets version.

The machine-centred approach also comes along with several shortcomings: As Völkel et al. [23] report, the dimensions identified by them which are supposed to be unique for CA overlap in many cases with the BFM originally developed for human personality. The dimension agreeableness of the BFM, for example, is similar to the dimensions approachable, social-inclined and social-assisting in the model of Völkel et al. [23]. Also, the dimensions such as confrontational, dysfunctional, and unstable are intercorrelated. Furthermore, ten dimensions are quite complex, for the BFM such as

describing its relationship with language cues in CA and in robots. Transferring these results to a completely new model is challenging, especially when considering dimensions such as extraversion. Extraversion is a component of the BFM and has been extensively studied in terms of its language cues. However, it is unclear to which dimensions in Völkel et al.'s [23] model extraversion should be attributed, as it could potentially relate to dimensions like self-conscious, social-entertaining, or approachable.

Facets \ Dimensions	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
	Sociability	Compassion	Organisation	Anxiety	Aesthetic Sensing
	Assertiveness	Respectfulness	Productiveness	Depression	Intellectual Curiosity
	Energy Level	Trust	Responsibility	Emotional Volatility	Creative Imagination

Fig. 1. BFM and its facets according to the BFI questionnaire [3]

As both approaches, on their own, have several weaknesses, this paper assumes that they should be combined: The studies of Völkel et al. [23] show that the personality of a machine does not have to be conceptualised in a completely other way than a human personality. A basis for such a model could be a BFM, but this model needs to be evaluated concerning its external validity. This evaluation of validity is necessary, as otherwise it remains unclear how to measure artificial personality in a valid manner, as the questionnaires used to analyse artificial personality in CA are based on the BFM. Besides the BFM, some dimensions of the model Völkel et al. [23] proposed might be integrated into a model of artificial personality.

3 Method

This paper uses an expert study approach, as experts have an in-depth perspective on theoretical aspects that are difficult to evaluate for non-experts, such as normal users. Hence, experts from the field of conversational design and robotics filled out a questionnaire on the validity of dimensions, facets, and items originating from human personality models for CA. A standardized questionnaire was used for two reasons: First of all, the experts should evaluate all the same dimensions, the same facets, and items to make their evaluations comparable. If a more qualitative approach had been used, the results would have been less comparable. Second, research uses questionnaires to evaluate the personality of CA. Hence, the experts can evaluate the validity of the items of the exact same questionnaires, facilitating their improvement, and this would not be possible with another method in a standardized way.

The sample consisted of N = 16 experts in CA and robotics originating from practice (N = 13) and from research institutes (N = 3). Among the practitioners, five worked for small companies with < 200 employees, and eight for larger companies (> 5,000 employees). Nine were consultants or conversational designers, one was a consulting manager, three worked in companies, and three in research. Experts were recruited via the personal network of the authors, via contacting profiles from LinkedIn, or via snowball sampling asking the experts which already participated in the study about relevant

contacts. Three participants filled out only partially the questionnaire. The sample had a mean age of 32.31 years (SD = 4.34 years). 75 % of the participants were female, and all participants had at least a bachelor's degree. The current position the participants held for M = 2.94 years (SD = 1.61 years), and they had experience in the field of conversational design for M = 4.91 years (SD = 2.35 years). They worked in several sectors with the service industry (12 participants), finance (10), energy (9), and ICT (9) being the top four mentioned sectors.

The data originated from a mixed-method study, including interviews covering topics beyond the questionnaire related to personality in CA. This paper focusses, due to limited space, solely on the data from the questionnaire encompassing four parts:

The first part asked questions about demographics and job experience.

The second part asked about the relevance or validity of several personality dimensions to describe technology such as CA. Relevance in this context describes whether the experts regard, e.g., personality dimensions which are considered to be important for humans, as valid for CA. In total, the participants were asked about 14 dimensions originating from several theoretical backgrounds: The first five dimensions were the Big Five dimensions: openness, conscientiousness, agreeableness, emotional stability, and extraversion. Furthermore, from the ten dimensions found by Paunonen and Jackson [16] and Rauthmann [18], the three dimensions: honesty, gender, and humour were chosen, as they were regarded as possibly relevant for CA in general. Finally, the users were asked about dimensions originating from the BIP [8]. Two versions of the BIP exist, a six-and a four-factor version. The six-factor version was used: Commitment, social competence, discipline, psychological stability, dominance, and cooperation. The experts evaluated the dimensions using a Likert scale ranging from 1 "not relevant at all" to 5 "fully relevant," with an additional category of "inadequate description".

The third part assessed facets of dimensions that need to be introduced into a model of artificial personality to facilitate the implementation. Using the BFI-2 questionnaire [3], the participants evaluated three facets for each BFM dimension. For the three additional dimensions - honesty, gender, and humour – facets were not available, so they were excluded. As for the BIP, facets are only available for the four-factor version, and the included facets were: openness to contact, sensitivity, sociability, team orientation, assertiveness, performance motivation, design motivation, leadership motivation, flexibility, action orientation, working under pressure and self-confidence.

The fourth part of the questionnaire analysed the relevance of the items as valid items are necessary to develop a questionnaire measuring artificial personality. The items originated from the BFI-2 questionnaire for the BFM, and the items for the BIP were based on the four-factor model in the informant observation version. There is no informant observation version available for the BFI questionnaire, which is why the self-report questionnaire was adapted. The items were reformulated from a first-person perspective into a third person perspective by replacing the "I" personal pronoun with the formulation "the system". Facilitating the imagination of the relevance of so many items, the experts should imagine building the following system: "*Imagine a client requests the development of a chatbot to aid hospital patients during admission. This chatbot should be accessible both remotely and on-site. It gathers anamnesis information and offers guidance on scheduled procedures. The client emphasizes the*

importance of the system conveying trust and professionalism to encourage active patient engagement in discussing health information. Simultaneously, the system should ensure data collection and efficient storage within the hospital's information management system." The experts assessed each item for the scenario described, but also with a generalist perspective using a Likert scale from 1 "not relevant at all" to 5 "fully relevant" along with an "inadequate description" category. Reliability analysis of the items yielded Cronbach's alpha values ranging from .64 to .80 for the BFM dimensions. For six facets the measurement was not reliable for the BFM: Sociability, compassion, respectfulness, responsibility, emotional volatility, and intellectual curiosity as the level was below .60. For the BIP the reliability was only calculated for the facets and for two facets it was below .60: Action orientation and working under pressure.

The questionnaire contained 151 closed items, and participants provided general feedback and suggested new dimensions, facets, and items via two open-ended questions. Data analysis involved using SPSS for descriptive statistics and MAXQDA for qualitative content analysis of open-ended questionnaire responses. The latter encompassed a classification of the answers into categories that were inductively generated using a content-structuring approach [11].

4 Results

The results of the expert study are described for four aspects: Comparison of the BFM and the BIP model, the relevance of the dimensions of both models and the three additional dimensions of Paunonen & Jackson [16], the facets, and the items of both models.

4.1 Relevance of Models

The BIP was compared with the BFM regarding its validity as many CA and robots usually fulfil job roles and the BIP was specifically designed to measure personality in the context of work. Thus, the BIP could be a better inspiration for the development of a model of artificial personality in contrast to the generalist BFM. The mean relevance of both models for the dimensions, facets and items was compared. On the one hand, regarding the dimensions, the mean approval for the BFM is higher (dimensions: mean approval dimensions BFM M = 3.33, SD = 1.14; mean approval dimensions BIP M = 3.04, SD = 1.23) and on the other hand, the facets and items of the BIP are regarded as more valid (facets: mean approval items BFM M = 2.88, SD = 0.86; mean approval items BIP = 3.21, SD = 0.97; items: mean approval items BFM M = 2.02, SD = 0.60; mean approval items BIP M = 2.94, SD = 1.00).

4.2 Relevance of Dimensions

To define relevance, the cut-off was set at the middle of the scale at the value of ≥ 3. Based on this criterion only three dimensions of the BFM are regarded as relevant (in descending order): Extraversion (M = 4.20, SD = 1.56), agreeableness (M = 4.20, SD = 1.42) and conscientiousness (M = 3.80, SD = 2.04). The dimensions openness (M

=2.60, SD = 1.77) and emotional stability (M =1.87, SD = 2.23) were not BFM dimensions (green) and the BIP model (yellow) as rated by the experts (N = 16) seen as relevant, but the dimension emotional stability had the highest standard deviation of all dimensions meaning there was a higher disagreement between the experts.

From the three dimensions that were debated as additional dimensions of the Big Five, the dimensions humour (M = 4.07, SD = 0.8) and honesty (M = 3.93, SD = 2.05) were regarded as relevant with humour having the lowest standard deviation. The dimension gender was not regarded as relevant (M = 2.29, SD = 1.68).

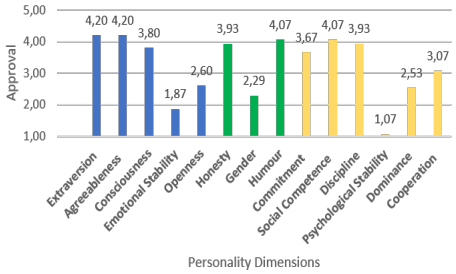


Fig. 2. Relevance of the dimensions originating from the BFM (blue), the additional BFM dimensions (green) and the BIP model (yellow) as rated by the experts (N = 16)

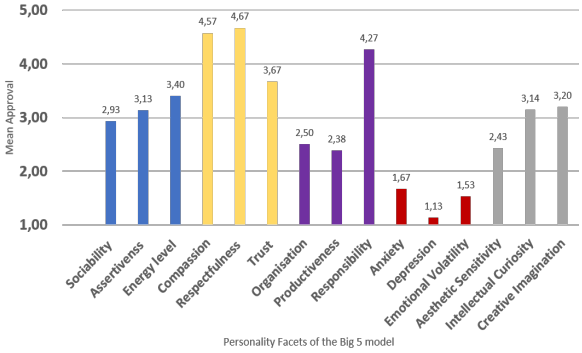
Concerning the six dimensions of the BIP model, four dimensions were regarded as relevant as seen in Figure 2: social competence (M = 4.07, SD = 1.28), discipline (M = 3.93, SD = 2.05), commitment (M = 3.67, SD = 1.68) and cooperation (M = 3.07, SD = 2.31), while the latter being very close to the cut-off criterion. Dominance (M = 2.52, SD = 2.00) and psychological stability (M = 1.07, SD = 1.86) were not seen as relevant. Participants could answer two open questions: "Do you have any comments regarding the personality dimensions?" and "Are you missing specific personality dimensions?". Seven experts commented on the first question. Experts 4 and 9 criticised that the dimensions despite describing important questions regarding the development of CA, the dimensions are too much human-specific concerning their formulation. Expert 8 adds that the relevance of the dimensions depends on the task, the user group and the context of the use. Seven experts gave feedback on the second question: Expert 6 is aware that dimensions are missing, but she did not mention any new dimension. Others named concrete dimensions: Impulsivity (expert 13), empathy (expert 18), honesty (expert 5). Expert 5 described what honesty means: transparency, comprehensibility and aims and limits of the system. Cultural aspects are mentioned by two experts (experts 4 and 13) and expert 13 also describes aspects such as background and job.

4.3 Relevance of Facets

Concerning the facets of the BFM all facets of the dimension emotional stability, namely anxiety (M = 1.67, SD = 1.35), depression (M = 1.13, SD = 0.99) and emotional volatility (M = 1.53, SD = 1.51) are not regarded as relevant as seen in Figure 3. On the contrary, all facets of the dimension agreeableness are seen as relevant, namely compassion (M = 4.57, SD = 0.65), respectfulness (M = 4.67, SD = 1.29) and trust (M

= 3.67, SD = 2.50). Concerning the other dimensions, the results are more differentiated: Regarding the dimension extraversion, only the facets assertiveness and energy level are seen as relevant. Nevertheless, all three facets are rather close to the cut-off of 3. Only the facet responsibility (M = 4.27, SD = 1.49) is seen regarding the dimension conscientiousness as relevant, the facets organisation (M = 2.50, SD = 1.74) and productiveness (M = 2.38, SD = 1.98) are not seen as relevant. Concerning the dimension openness, the facets intellectual curiosity (M = 3.14, SD = 1.75) and creative imagination (M = 3.20, SD = 1.52) are seen as relevant, but the values are very close to the cut-off. Only aesthetic sensitivity is not seen as relevant (M = 2.43, SD = 1.91).

Fig. 3. Relevance of the dimensions originating from the BFM (blue), the additional BFM dimensions (green) and the BIP model (yellow) as rated by the experts (N = 16)



For the BIP model, two facets, namely conscientiousness and emotional stability, were not included as their name is in German and in English identical with dimensions of the BFM. Of the remaining 14 facets, the following results were found for the four-dimensional version of the BIP model: Factor one named social competencies encompasses five facets openness to contact, social sensitivity, sociability, team orientation and assertiveness. Openness to contact (M = 4.14, SD = 1.41), social sensitivity (M = 4.13, SD = 1.30) and sociability (M = 3.60, SD = 1.68) are seen as relevant, while team orientation (M = 2.67, SD = 1.87) and assertiveness (M = 2.73, SD = 1.58) are not regarded as relevant as indicated in Figure 4. All facets of the second factor occupational orientation, namely achievement motivation (M = 2.47, SD = 2.00), power motivation (M = 1.87, SD = 1.64) and leadership motivation (M = 2.27, SD = 1.87) are not seen as relevant. All facets asked concerning the third factor occupational behaviour, namely flexibility (M = 4.67, SD = 0.62) and action orientation (M = 3.93, SD = 1.58) are regarded by the experts as relevant. Finally, regarding the dimension psychological constitution, the facet self-confidence (M = 2.87, SD = 1.81) is not seen as relevant, but the facet working under pressure (M = 3.07, SD = 2.05) is seen as relevant. Nevertheless, both facets are close to the cut-off value of 3.

The participants commented on the facets indicated and they should name additional facets: Regarding the first question, experts 1 and 6 criticise that the facets are formulated for humans and the relevance of the facets depends on the context (expert 6). Furthermore, the facets would still be vague and thus leaving a big “scope of interpretation” (expert 5). Humour (expert 4) and trust (expert 5) are mentioned as additional

facets. Expert 5 considers the tone of voice as an extra facet, and experts 6 and 13 emphasize that the role of the machine influences the relevance attached to the facets.

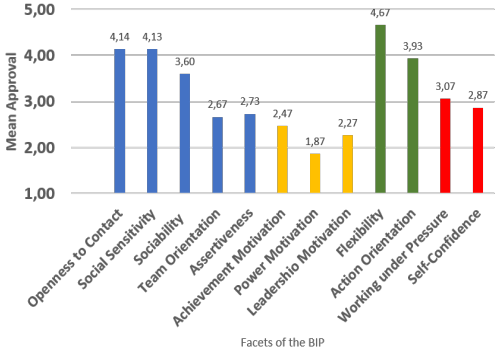


Fig. 4. Relevance of the facets of BIP as rated by the experts (N = 16), The four dimensions are represented in colours: social competencies (blue), occupational orientation (yellow), occupational behaviour (green) and psychological constitution (red)

4.4 Relevance of Items

The experts rated the items as relevant or not on a Likert-scale reaching from 1 “not relevant at all” to 5 “fully relevant” with an additional category of “inadequate item”. Especially the latter is regarded as important and if five or more experts regard an item as inadequate, the relevance respectively the validity of the item will be questioned. Each dimension of the Big Five questionnaire is measured with twelve items. Those twelve items can be divided into three facets with four items per facet. The items are available for free in German [3] and in English [21].

Regarding the dimension extraversion, the items 26 (8 experts regarded it as inadequate), 31 (5), 36 (5), 41 (5) and 51 (7) were regarded as inadequate by at least five experts. The same items achieved also low scores concerning their mean relevance. The items originate from all three facets. For the dimension agreeableness only, item 47 (6) was not regarded as adequate. It originates from the facet compassion. For the dimension conscientiousness items 3 (5), 33 (5) and 48 (8) were not regarded as adequate. They originate from the facet organisation. For the dimension emotional stability, the items 39 (5) and 54 (5) are regarded as inadequate. They originate from the facet depression. For the dimension openness all items except items 10 and 15 are regarded as inadequate: 5 (8), 20 (7), 25 (6), 30 (6), 35(9), 40 (8), 45 (8), 50 (11), 55 (9) and 60 (6).

Concerning the BIP model an informant report version of the questionnaire exists for the 4-factor version of the model. This questionnaire was used and the formulation “the person I assessed” was replaced by “the system” for each item. The items of the questionnaire were used with permission from the publisher. For the dimension social competencies eight of the 15 items were regarded as inadequate, 23 (6), 24 (6), 28 (5), 29 (5), 30 (9), 31 (6), 32 (8) and 33 (7). The facet openness to contact (two items), the complete facet team orientation (28, 29, 30), and the complete facet assertiveness are regarded as inadequate. For the dimension occupational orientation, the items 1 (5), 3 (6), 6 (5), 7 (5) and 8 (5) are regarded as inadequate. All facets of this dimension

encompass inadequate items. For the dimension occupational behaviour, the items 11 (5), 15 (5) and 18 (8) are regarded as inadequate. These items concern the dimensions of flexibility and action orientation. Regarding the final dimension of psychological constitution, three items are evaluated by the experts as inadequate: 34 (6), 39 (5) and 42 (6). One item is inadequate for each facet.

Regarding the items, the experts could provide in general comments on them. Three participants criticise that it is difficult to transfer the items from a human to a machine context as they are very human-centred regarding their formulation (experts 5, 6, 7). Many items were found to diverge from actually important relevant aspects (expert 7).

5 Discussion

5.1 An Initial Model of Artificial Personality Based on the Results of the Expert Study

Summarising the expert study, a 1:1 transfer of human models of personality to machines is not valid as the experts' feedback on the questionnaire shows that artificial personality deviates from human personality in several aspects: While the BFM is a good foundation for a model of artificial personality, not all dimensions of the model are relevant for machines as the dimensions emotional stability and openness to experience are questioned by the experts. For openness, the experts doubt the validity of the dimension itself, some of the facets and most of the items. Concerning the dimension of emotional stability, the dimension and the facets are regarded as non-valid in the context of artificial personality. Furthermore, the paper introduces the concept of facets to describe personality to give a more differentiated picture of the dimensions: Only for the dimension of agreeableness are all facets regarded as valid. This is not the case for extraversion and conscientiousness. Apart from that, research on the Big Five theory debates additional dimensions such as humour, gender, or honesty. The expert study supports the need for additional dimensions. At least, the constructs of honesty and humour seem to be relevant to conceptualise artificial personality. Furthermore, CA and robots usually fulfil job-related tasks and thus it was assumed, that a personality model such as the BIP, developed for the business context, could be more valid than a general model such as the BFM. The results show a nuanced picture. The dimensions of the BFM are more relevant or valid for the experts for artificial personality than the dimensions of the BIP (six-factor model). Nevertheless, the items and facets of the BIP are in general more relevant than those of the BFM. Based on this result, the dimensions of the BFM could be connected to the facets and items of the BIP. This is theoretically possible as the BFM and the BIP are intercorrelated to some extent [8]. Furthermore, Völkel et al. [23] developed a personality model with a special focus on CA. This model overlaps with the BFM [23]. Nevertheless, it has unique dimensions that are challenging to integrate into the BFM. Such dimensions are "artificial" and "serviceable". These dimensions could be added to the BFM. Given these results, a first draft of a model of artificial personality can be designed, as presented in Figure 5.

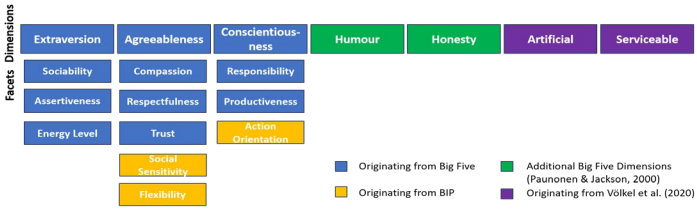


Fig. 5. First draft of a model of artificial personality based on the expert study

5.2 Theoretical Contributions and Practical Implications

The paper makes four theoretical contributions: First, it provides a systematic definition of the concept of artificial personality integrating technological perspectives from robotics and CA. This definition helps in developing a research model.

Second, it introduces an initial artificial personality model that combines the human-centred and machine-centred approach, based on an empirical validation. This model is crucial for assessing personality in CA accurately. Besides, a more valid model facilitates the implementation for designers, as they can focus on dimensions and facets that actually matter to users in the context of CA.

Third, this paper focuses on experts in conversational design and robotics, which is rarely done in research in those areas. The results show that experts can help to build theories and to specifically improve their ecological validity. Furthermore, they provide insights into the relevance of a topic and best practices.

Fourth, based on the first draft of a model of artificial personality as seen in Figure 5, measurement instruments could be designed. This is necessary as valid measurement instruments are missing [2] and the model could be evaluated with such instruments.

On a practical level, the paper presents an initial model that can serve as the basis for developing practical tools, such as a questionnaire to measure the perception of artificial personality in machines. Questionnaires and workshop methods can be developed to measure the affordances of clients and tools for the engineering of artificial personality such as a toolbox of language cues to implement the theory into a system.

5.3 Limitations and Future Research

Several methodological improvements are possible for this study: the sample could have been enhanced by including more robotics experts, and the scenario described could have been broadened beyond conversational agents to improve results' validity for robotics. Furthermore, achieving greater diversity in the sample, including experts from various cultural backgrounds, could have been valuable as it may impact personality perception [13]. The sample size was sufficient for the context of qualitative research for stimulating the generation of theory, but the results are not representative for experts in conversational design and robotics as the size is too small for quantitative representativeness. The results are part of a mixed method study, and for theoretical saturation and theory generation, sample sizes between N = 9 and 17 suffice [6].

Regarding the methodology of the study, the cut-off criteria for “relevance” of dimensions, facets and items are a bit arbitrary and could be more or less strict.

Furthermore, the relevance or validity of the items was evaluated by presenting a scenario to facilitate the evaluation. Maybe more than one scenario should have been presented to guarantee a more valid result. Additionally, the experts had problems interpreting the dimensions and facets as the terms such as “conscientiousness” are very abstract. The evaluation of such dimensions could be facilitated by giving definitions for each dimension and facet that are adapted for machines, but exactly such definitions are still missing research. Finally, only three dimensions of the ten additional dimensions for the Big Five suggested by Paunonen and Jackson [16] and Rauthmann [18] were analysed. Maybe even more dimensions could play a role in artificial personality.

This paper represents an initial step in developing a more valid artificial personality model. It defines the concept and outlines potential development paths. The next stages involve defining and adapting dimensions, facets, and items for the machine context. Additionally, the model will be translated into a questionnaire for empirical evaluation.

6 Conclusion

Research on artificial personality gains importance. Nevertheless, a systematic definition and a valid model were missing, but first steps were already made by, e.g., Völkel et al. [23] and Ahmad et al. [1]. This paper introduces a systematic definition and shows possibilities for the development of a model for artificial personality based on experts’ feedback. More research is necessary to fully specify the model and to develop measurement instruments and tools for engineering CA or robots.

References

1. Ahmad, R. et al.: A Framework of Personality Cues for Conversational Agents. Presented at the Hawaii International Conference on System Sciences (2022).
2. Christoforakos, L. et al.: Die Roboterpersönlichkeit - Konzeption, Gestaltung und Evaluation der Persönlichkeit von sozialen Technologien. (2019).
3. Danner, D. et al.: Die deutsche Version des Big Five Inventory 2 (BFI-2). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). (2016).
4. Endres, L.S.: Personality engineering: Applying human personality theory to the design of artificial personalities. In: Symbiosis of Human and Artifact. pp. 477–482.
5. Følstad, A., Brandtzaeg, P.B.: Users’ experiences with chatbots: findings from a questionnaire study. Qual User Exp. 5, 1, 3 (2020).
6. Hennink, M., Kaiser, B.N.: Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Social Science & Medicine. 292, 114523 (2022).
7. Holtgraves, T.M. et al.: Perceiving artificial social agents. Computers in Human Behavior. 23, 5, 2163–2174 (2007).
8. Hossiep, R., Paschen, M.: BIP Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung. Hogrefe, Göttingen (2003).

9. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. International Journal of Human-Computer Studies. 53, 2, 251–267 (2000).
10. Krauss, P., Maier, A.: Will We Ever Have Conscious Machines? Front. Comput. Neurosci. 14, 556544 (2020).
11. Kuckartz, U., Rädiker, S.: Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden. Beltz Juventa, Weinheim Basel (2022).
12. Mori, M. et al.: The Uncanny Valley [From the Field]. IEEE Robot. Automat. Mag. 19, 2, 98–100 (2012).
13. Mou, Y. et al.: A Systematic Review of the Personality of Robot: Mapping Its Conceptualization, Operationalization, Contextualization and Effects. International Journal of Human–Computer Interaction. 36, 6, 591–605 (2020).
14. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. J Social Isssues. 56, 1, 81–103 (2000).
15. Neff, M. et al.: Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents. In: Allbeck, J. et al. (eds.) Intelligent Virtual Agents. pp. 222–235 Springer Berlin Heidelberg, Berlin, Heidelberg (2010).
16. Paunonen, S.V., Jackson, D.N.: What Is Beyond the Big Five? Plenty! J Personality. 68, 5, 821–835 (2000).
17. Rauthmann, J.F.: Grundlagen der Differentiellen und Persönlichkeitspsychologie: Eine Übersicht für Psychologie-Studierende. Springer Fachmedien Wiesbaden, Wiesbaden (2016).
18. Rauthmann, J.F.: Persönlichkeitspsychologie: Paradigmen – Strömungen – Theorien. Springer, Berlin [Heidelberg] (2017).
19. Shumanov, M., Johnson, L.: Making conversations with chatbots more personalized. Computers in Human Behavior. 117, 106627 (2021).
20. Smestad, T.L., Volden, F.: Chatbot Personalities Matters: Improving the User Experience of Chatbot Interfaces. In: Bodrunova, S.S. et al. (eds.) Internet Science. pp. 170–181 Springer International Publishing, Cham (2019).
21. Soto, C.J., John, O.P.: The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. Journal of Personality and Social Psychology. 113, 1, 117–143 (2017).
22. Tapus, A. et al.: User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intel Serv Robotics. 1, 2, 169–183 (2008).
23. Völkel, S.T. et al.: Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–14 ACM, Honolulu HI USA (2020).
24. Zhou, M.X. et al.: Trusting Virtual Agents: The Effect of Personality. ACM Trans. Interact. Intell. Syst. 9, 2–3, 1–36 (2019).