IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 25, 2021, accepted May 10, 2021, date of publication May 12, 2021, date of current version May 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079838

Trait Based Trustworthiness Assessment in Human-Agent Collaboration Using Multi-Layer Fuzzy Inference Approach

SADAF HUSSAIN[1], RIZWAN ALI NAQVI[2], SAGHEER ABBAS[1], MUHAMMAD ADNAN KHAN[3,4], TANWEER SOHAIL[5], AND DILDAR HUSSAIN[6]

[1]School of Computer Science, National College of Business Administration and Economics, Lahore 54660, Pakistan
[2]Department of Unmanned Vehicle Engineering, Sejong University, Seoul 05006, South Korea
[3]Pattern Recognition and Machine Learning Laboratory, Department of Software Engineering, Gachon University, Seongnam 13557, South Korea
[4]Faculty of Computing, Riphah School of Computing and Innovation, Riphah International University, Lahore Campus, Lahore 54000, Pakistan
[5]Department of Mathematics, University of Jhang, Jhang 35200, Pakistan
[6]School of Computational Sciences, Korea Institute for Advanced Study (KIAS), Seoul 02455, Republic of Korea

Corresponding authors: Muhammad Adnan Khan (adnan.khan@riphah.edu.pk) and Dildar Hussain (hussain@kias.re.kr)

This work was supported in part by the Korea Institute for Advanced Study (KIAS) under Grant CG076601, and in part by the Sejong University Faculty Research Fund.

ABSTRACT Trust is an essential requirement for effective Human-Agent interaction as artificial agents are becoming part of human society in a social context. To blend into our society and maximize their acceptability and reliability, artificial agents need to adapt to the complexity of their surroundings, like humans. This adaptation should come through knowing whom to trust by evaluating the trustworthiness of its human mate. It is therefore required to build cognitive agents with trust models that may allow them to trust humans the same way a human trusts other humans keeping under consideration all factors influencing the human agent trust mechanism. Several antecedents within the cognitive system itself and the surroundings dynamically influence the trust mechanism. Personality, as a trusted antecedent has been found to have a substantial impact in predicting human interactor's trustworthiness that critically assists trust decision making. Current research, therefore, aims to infuse characteristics of respective humans as the antecedent of the human agent trust process. This is accomplished by incorporating into the trust model the agent's capability to perceive the personality traits of the human interactor. The current work is focused on introducing a trustworthiness assessment model (TAMFIS) based on fuzzy inference to assess human's trustworthiness towards artificial agents by exploring the human's personality traits that predict trustworthiness. The artificial agent could develop its character towards its human collaborators that will help it in effective interactions. The testing of the proposed architecture is carried out using Dempster Shafer Theory of belief and estimation. It is anticipated that the proposed trust model will effectively evaluate the trustworthiness of human collaborators and develop a more reliable human–agent trust relationship.

INDEX TERMS Big 5 personality traits, multi-agent systems, trustworthiness, artificial agent, Dempster Shafer theory.

I. INTRODUCTION

In human-agent collaborative societies, working together often involves having interdependence and therefore the team members need to depend on each other to accomplish collaborative tasks. Artificial agents are becoming a part of human society rapidly and are expected to work in collaboration with humans. Humans and cognitive agents are required to become teammates like human mates. Collaborative teammates, in the accomplishment of a common goal or a task, require partnerships based on trust. Artificial agents are therefore equally required to establish a trust relationship with their human mates as humans do. This trust-building has a strong association with the trustworthiness perception of the agent for its human collaborator. Perceiving the trustworthiness of a human is governed by various subjective factors

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Da Lin.

including his behaviors. The intensifying degree of autonomy of artificial agents is contributing to the complexity of the nature of a partnership, shared autonomy of collaboration, reasoning, and understanding in social setups. Trust, therefore, plays an important role in deciding when to rely on the teammates.

Previous studies have revealed that the characteristics of a trustee, situational factors, and perception of trustor have a significant impact on trust decision [1]. A trustor's desire of trusting others is also affected by the status of opponent [2], nationality [3], [4], gender [5], [6], fear of social segregation [7], presence of a monitoring system [8], [9], trustor's emotions [10] and dutifulness [11]. Cues and behaviors that govern trust mechanisms have extensively been studied, however, lesser consideration has been put into characteristics and traits that assess the trustworthiness of the collaborator. Literature has also explored the perceptions of trustworthiness. Perception of trustworthiness is found to be understood by understanding Mayer, Davis, and Schoorman's (1995) ability, benevolence, and integrity (ABI) model. Individuals who are perceived as highly intelligent, capable, and competent (high ability), empathic, and caring (high benevolence), and consistent and ethical (high integrity) are more probable to be trustworthy. Trustworthiness assessment can also be performed based on the opponent's characteristics, social setup, and situations; like whether the trustee has fulfilled his promises in past and how the trustee is gullible for breaking promises [12], [13].

Under these findings, the current study aims towards modeling a trustworthiness assessment model that influences human and agent collaborative interaction. This work considers cognitive trust development that captures the human view of perceived trustworthiness and the ability to assess an appropriate level of confidence in humans.

The paper is organized in sections, where section II covers briefly the related work, section III describes the proposed model proceeded by both through Mamdani fuzzy-based approach and Dempster Shafer's theory of belief and estimation. Simulation results of the proposed Fuzzy Inference System are also compared with the results obtained through Dempster Shafer Theory. Section IV concludes the work and possible future work.

II. LITERATURE REVIEW

In human-agent social setups, relatively few scholars have investigated the correlates of trustworthiness and examined the relationship between the Big Five personality traits (extraversion, openness, agreeableness, neuroticism, conscientiousness) as influencing factors.

Personality is described by its components called traits. These traits are not directly observable but are inferred through behavior patterns [14], [15]. Personality traits are found to have stability throughout life and have grounded into genetics [16]. Perceived personality traits of the team member have a considerable influence on the desire to develop a trust-based relationship [17]. Although to measure an individual's personality traits, there is no fully encompassing technique, nevertheless, a strong consensus in Psychology approves that the Five-Factor Model (FFM) of personality is capable of providing a broader way of measuring the traits [18], [19]. FFM covers human characteristics into five personality traits. These personality traits are found to be independent of language and culture, therefore this model captures the traits universally. These traits include openness, agreeableness, conscientiousness, extraversion, and neuroticism. These personality traits can generate an individual's worldview from a broader perspective and help assess his trustworthiness.

McCrae and Costa [20] defines openness to experience as an extent to which an individual accepts new ideas, develops new approaches, and ready to experience new happenings. A curious person who scores higher in these aspects has curiosity towards exploration, original in thoughts, and clear in imaginations, whereas low scorers are found to be more conservative and careful. Researchers in cognitive science argues that the more a person possesses this aspect the more he is open-minded and tolerant found to show trustworthy behavior [21].

Persons, more towards rationalism and well informed think themselves more incompetence is considered to be more conscientious. These individuals are known for their organization, thoroughness, and ambitions [22]. In contrast to conscientious people, there are people at lower levels of maturity, patience, and are careless. Conscientious individuals seem to be better informed and make better decisions under diverse situations. Persons with high levels of conscientiousness do not rely easily on information provided by others [21].

Extraversion is a special trait of social, active, and lively individuals, opposite to it, a shy and passive person has a lower score for the traits and is classified as introverts [20]. Extraverts exhibit more trustworthy behavior than introverts, which makes them more desired to be trusted in social communication.

People showing more tendencies towards cooperating with others are known to be more agreeable. Among factors of personality traits agreeableness is highly correlated to trustworthiness [23]–[25]. This property is found to influence trustworthiness less in presence of very low neuroticism [25]. Agreeableness is therefore influential on trustworthiness considering its correlation with other personality traits.

Neurotic people have been found to show more distrust [26], they evaluate threat more keenly and often leads to a decision that their opponents are malicious [27].

Mayer *et al.* [28] and Falcone and Castelfranchi [29], has provided a notion of trust that is applicable in the dynamical analysis. They define trust in terms of the willingness of the trustee to develop trust relations and is influenced by a trustee's ability, benevolence, and integrity. Where ability defines the capabilities of a trustee in a specific area, benevolence is the degree to which a trustee is assumed to be good for the trustor whereas integrity is a perception that how trustee follows rules and regulations.

Psychology, philosophy, management, and economics have widely studied the concept of trust and trustworthiness [30]. The importance of trust in human-agent cooperative environments has received much contemplation [31]. Trust can be classified and studied under three domains; credentials, past experiences, and cognitive trust.

Trust based on credentials is developed by applying certain credentials to gain access; in e-commerce and peer-to-peer applications trust in an agent is evaluated based on the experience of interactions with other agents [32], [33]; in contrast, cognitive trust captures human social norms for trust-based decisions in social interactions [34], [35].

There has been a debate between personality researchers that personality traits influence trustworthiness in opponents. Agreeableness has a strong influence on trust-building [36]–[38], along with agreeableness, conscientiousness and openness are also helpful parameters in defining the trustworthiness of the opponent. Every personality trait has a certain level of influence on trust development [21], whereas, according to Yamagata *et al.* [39], extraversion has lower impacts on developing trustworthiness.

Perception of trustworthiness dimensions has been under research, ability, benevolence, and integrity; also known as the ABI model, are considered to be the dominant paradigm of understanding trustworthiness. Individuals are perceived to be trustworthy if they have professed to be intelligent and capable (able); caring and kind (benevolent) and consistent and well behaved (integral). Trustors make judgments for these three dimensions through social, personal, and situational cues [40], as well as how the trustee has been behaving in contradicting situations [12], [13], [41]. Trust findings may seem to appear with some divergence, the reason is assumed simply due to different sizes of samples and (or) measurements, whereas they are certainly worth a closer look.

Personality factors are one of the crucial factors in developing a trust relationship among team members, especially when an agent has to interact with diverse team members and has a strong impact on developing trust. Several studies have focused on the estimation of the trustworthiness of collaborative teammates [42]–[44]. Major research work has been conducted to assess the trustworthiness of teammates in psychology and social sciences [1], [45], [46] whereas few attempts have been made to assess computationally the trustworthiness of human mate [47], [48], also in existing research in human-agent systems, personality traits as antecedents of trustworthiness assessment has been missing.

The current study takes these limitations a step further with inspiration from fuzzy inference systems leading towards the suboptimal fuzzy inference system (FIS). Fuzzy sets and fuzzy logic are powerful mathematical tools to model uncertain industrial, human and natural systems. Fuzzy models facilitate decision-making by the means of approximate reasoning and linguistic terms. Fuzzy inference systems can play an important role when applied to complex cognitive phenomena that are not easy to describe by conventional mathematics [43], [49]–[52].

The proposed FIS is expanded with the cognitive ability to infer personality traits of the human teammate and incorporating them into an artificial cognitive agent that can distinguish trustworthy and untrustworthy sources of information based on the opponent's personality measures. The cognitive agent will be capable of modifying its behavior according to its belief, by adopting a probabilistic approach to model trust towards the opponent. Therefore it is believed that the current study will able to reproduce the results with more accuracy and reliability.

III. PROPOSED MULTI-LAYERED TRUSTWORTHINESS ASSESSMENT MODEL

The research model has been proposed and designed following the aforementioned theoretical design and is schemed in Figure-1. The model utilizes the perceived personality traits of human collaborators as a predictor of trustworthiness. A brief description of each of the blocks is provided as under.

A. AUTOMATIC PERSONALITY TRAITS ASSESSMENT MODULE

The model consists of an automatic personality detection system through textual conversation. Since text often reflects various aspects of human personality, this module has been constructed with the influence of the work of Poria *et al.* [53]. Using a convolutional neural network (CNN), the process of personality detection has been conducted through the stream of consciousness essays. For personality traits prediction two types of features extraction (i.e. word level and document level) is performed. Furthermore, for the five traits of personality, five different neural networks were trained and the corresponding output of each network (representing a particular personality trait) is obtained as a probability distribution through the softmax layer.

B. TRUSTWORTHINESS DIMENSIONS ASSESSMENT MODULE

The preliminary literature review has already shown that personality traits of an individual are associated with those of dimensions of trustworthiness therefore FFM has a strong influence on trustworthiness [28].

Generally, it is accepted that agreeableness is positively related to the perception of trustworthiness. In the trust game, research on trustworthiness has revealed that agreeableness is a strong predictor of trustworthiness [24], [54]. A high degree of conscientiousness in individuals is sensitive to ability-based violations. Therefore general carefulness leads to lower perceptions of trustworthiness. Neuroticism is related to higher threat evaluations leading to perceiving others as being malicious [26], [27]. Extravert persons have a higher tendency of risk-taking and possess positive emotions, both of which these qualities lead to high trustworthiness perceptions. In the last, openness to experience takes to a wider vision to accept values and customs and particularly useful in predicting integrity, since the trustor views the values of his collaborator as consistent with their own.

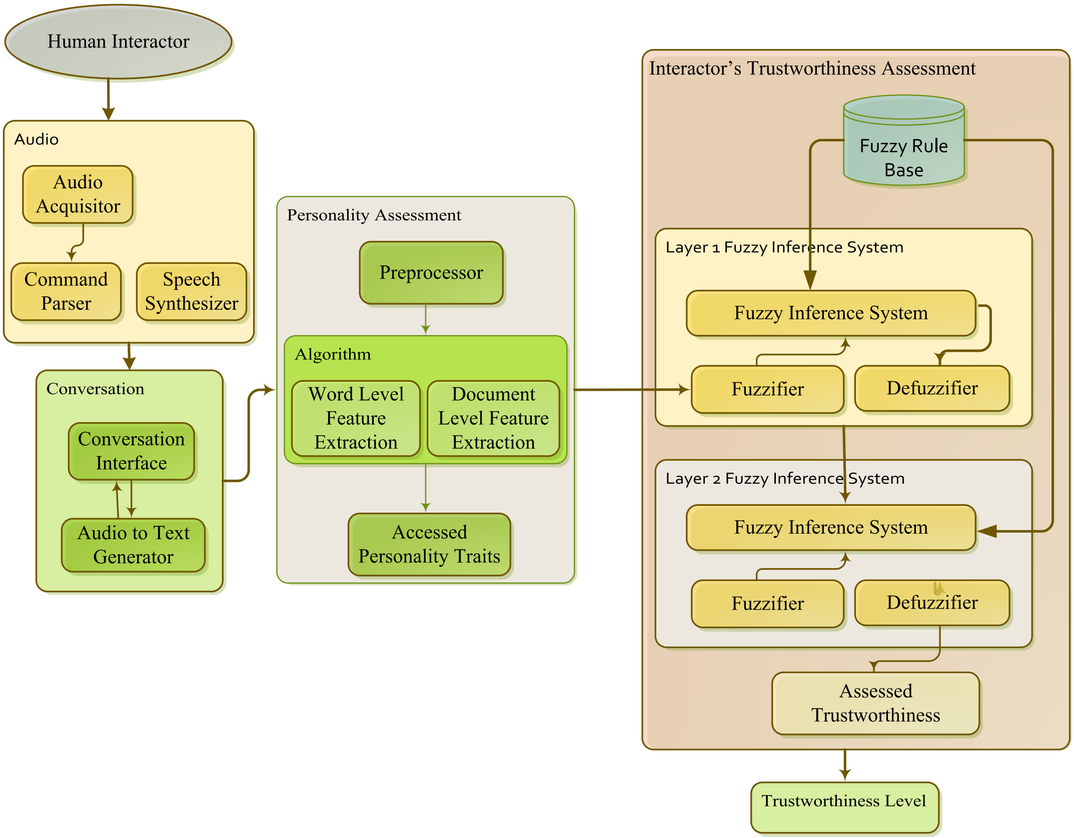


FIGURE 1. Proposed personality traits oriented trustworthiness assessment model.

The current research work is mainly focused on The trustworthiness assessment module to address the above-mentioned relationships. The module is designed to and assesses human trustworthiness based on the perceived personality traits assessment module. The details of this module are provided in the preceding sections.

C. AUDIO MODULE

The audio module is used for vocal outputs synthesis to receive the vocal message from the human and to guide him through the course of the interaction.

D. CONVERSATION MODULE

Conversation module processes vocal commands to generate textual data for the personality assessment module for personality trait perception.

IV. FUZZY-BASED SYSTEM MODEL

The proposed model is designed to assess the trustworthiness of a human agent based on his inferred personality traits using the Multi-Layer Mamdani Fuzzy Inference System (MFIS). Figure-2 shows the flow of the proposed system which consists of five parameters (the personality traits), initially to assess the trustworthiness dimensions (ability, benevolence, integrity) in layer 1.

The proposed model assesses trustworthiness (HD = Highly_Deceptive, D = Deceptive, PT = Partially_ Trustworthy, T = Trustworthy, VT = Very_Trutworthy) using five input variables (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) in Figure-2. The assessed trustworthiness dimensions are then fed to layer-2 MFIS for the final assessment of human trustworthiness. Personality factors have been found to be in mutual correlation that collectively influences the outcome of the proposed model. The values of input parameters i.e. the personality traits are used to form a lookup table for trustworthiness assessment. Since the proposed trustworthiness assessment model comprises of fuzzy "and" rules-based knowledge base, therefore, the proposed automated trustworthiness assessment model using Mamdani Fuzzy Inference based system for layer-1 and layer-2 can be written mathematically regarding t-norm as

$$t : [0,1] \times [0,1] \times [0,1] \times [0,1] \times [0,1] \rightarrow [0,1] \times [0,1] \times [0,1] \tag{1}$$

$$t : [0,1] \times [0,1] \times [0,1] \rightarrow [0,1] \tag{2}$$

For the proposed TAMFIS the fuzzy sets in layer-1, Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism, and for layer-2, Ability, Integrity, and Benevolence, the membership functions are transformed into their intersection in Eq.(1) and eq. (2) respectively. A membership

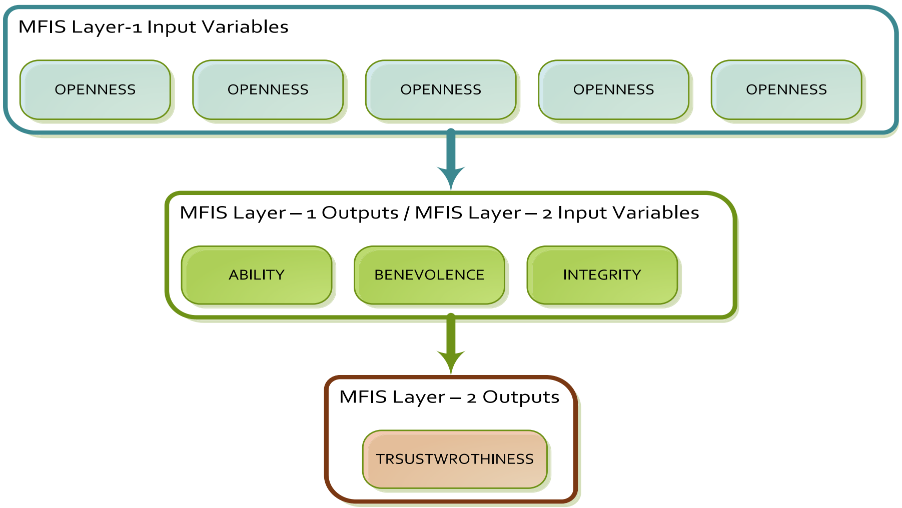


FIGURE 2. **Proposed trustworthiness assessment MFIS layers.**

TABLE 1. **Proposed fuzzy inference system ranges for calculating weights of trustworthiness dimensions.**

Input /Output Parameters	Levels	Ranges
Openness (**OPN**)	*Conventional* *Moderate* *Curious*	< 0.08 0.06 – 0.17 >0.15
Conscientiousness (**CON**)	*Careless* *Moderate* *Organized*	<0.15 0.1– 0.3 > 0.25
Extroversion (**EXT**)	*Reserved,* *Moderate* *Outgoing*	<0.07 0.05 – 0.13 >0.12
Agreeableness (AGR)	Challenging Moderate Friendly	<0.2 0.15 – 0.5 >0.45
Neuroticism (**NEO**)	*Confident* *Moderate* *Nervous*	<0.02 0.01 – 0.06 >0.05
Ability **(A)**	*Low* *Medium* *High*	<0.25 0.15 – 0.65 > 0.55
Benevolence **(B)**	*Low* *Medium* *High*	< 0.3 0.25 – 0.7 > 0.65
Integrity **(I)**	*Low* *Medium* *High*	< 0.25 0.15 – 0.8 >0.7
Trustworthiness (T)	**HD** **D** **PT** **T** **HT**	<0.2 0.5 – 0.40 0.35 – 0.6 0.55 – 0.88 >0.75

function maps each point to the degree of membership between 0 and 1 in-universe of discourse.

TAMFIS membership functions are given in table 1, whereas eq. (3) and (4) represent mathematically the membership functions for layer-1 and layer-2.

$$\mu_{OPN\cap CON\cap EXT\cap AGR\cap NEO}\,(\text{o}, \text{C}, \varepsilon, \mathcal{A}, \mathcal{N}) = \min\left[\mu_{OPN(O)}, \mu_{CON()}, \mu_{EXT(\varepsilon)}, \mu_{AGR(\mathcal{A})}, \mu_{NEO(\mathcal{N})}\right] \quad (3)$$

$$\mu_{Ability\cap Benevolence\cap Integrity}\,(a, b, i) = \min\left[\mu_{A(a)}, \mu_{B(b)}, \mu_{I(i)}\right] \quad (4)$$

A. INPUT/OUTPUT VARIABLES

Statistical values are used in a fuzzy system for assessment of trustworthiness, yielding statistical values for the outputs of the proposed fuzzy inference system. Inputs and outputs for the system along with the defined ranges are shown in Table. 1.

The ranges have been designed in accordance with the work of Lyons *et al.* [46], showing that openness to experiences and neuroticism of trustee lowers his trustworthiness whereas conscientiousness, extraversion, and agreeableness contribute to enriching trustworthiness

B. FUZZY IF-THEN RULES

Fuzzy logic if-then statements are utilized to design conditional statements to hold fuzzy logic. These statements describe the core grounds for the construction of a fuzzy rule base. A few of the rules designed for layer-1 of the fuzzy inference system are provided as under.

The proposed TAMFIS rules table is designed to follow a cognitive theory of trust psychology, the system comprises 243 input and output rules for layer-1and 27 input and output rules for layer-2. A few of these input and output rules are presented in Table. IV. Rules to measure an agent's perception of their human mate's ability, benevolence, and integrity are designed following the work of Lyons *et al.* [46]. Here we have also considered that the trustee is a teammate and therefore familiar. Fuzzy rules have been designed considering the correlation between the personality traits themselves.

IF*(Openness is Conventional) and*
(Conscientiousness is Careless) and
(Extroversion is Reserved) and
(Agreeableness is Challenging) and (Neuroticism is
Moderate)

THEN
(Ability is Low)(Benevolence is Low)(Integrity is Low)
IF*(Openness is Conventional) and*
(Conscientiousness is Careless) and
(Extroversion is Reserved) and
(Agreeableness is Moderate) and
(Neuroticism is Moderate)
THEN *(Ability is Low) (Benevolence is Low)*
(Integrity is Low)

.
.
.

IF *(Openness is Curious) and*
(Conscientiousness is Organized) and
(Extroversion is Outgoing) and
(Agreeableness is Friendly) and
(Neuroticism is Moderate)
THEN *(Ability is High) (Benevolence is High)*
(Integrity is Medium)
similarly, for layer-2:
IF *(Ability is Medium) and (Benevolence is Medium)*
and (Integrity is Medium) **THEN (Trustworthiness is** *T*)
IF *(Ability is High) and (Benevolence is Medium)*
and (Integrity is Medium) **THEN (Trustworthiness is** *T*)

.
.
.

IF*(Ability is High) and (Benevolence is High)*
and (Integrity is Medium) **THEN (Trustworthiness** *is VT)*

C. MEMBERSHIP FUNCTIONS

The membership functions for levels of proposed MFIS are given in Table. 2.

To use these rules realistically and proficiently, the major constituent of TAFIS i.e., the fuzzy rule base is created. The efficiency of an expert system is based on the implemented fuzzy rules set. Under experts' opinion, all possible fuzzy relations between inputs and outputs are included in the fuzzy rule base. These rules are written in an IF-THEN context covering all possible facets for an agent needed to develop a trust relationship with the human.

Fuzzy if-then rules $\mathcal{P}u^e$ ($1 \leq e \leq 243$) comprising fuzzy rule base for layer-1 are written as:

$\mathcal{P}u^1$ = **IF** OPN is Conventional and CON is Careless and EXT is Reserved and AGR is Challenging and NEO is Confident **THEN** A is Low B is Low I is Low

$\mathcal{P}u^2$ = **IF** OPN is Conventional and CON is Careless and EXT is Reserved and AGR is Challenging and NEO is Moderate **THEN** A is Low B is Low I is Low

.
.
.

$\mathcal{P}u^{243}$ = **IF** OPN is Curious and CON is Organized and EXT is Outgoing and AGR is Friendly and NEO is Confident **THEN** A is High B is High I is High

Similarly, for layer-2, rules, donated by $\mathcal{T}u^x$ where $1 \leq x \leq 27$:

$\mathcal{T}u^1$ = **IF** *(Ability is Medium) and (Benevolence is Medium) and (Integrity is Medium)* **THEN** *(Trustworthiness is T)*

$\mathcal{T}u^2$ = **IF** *(Ability is High) and (Benevolence is Medium) and (Integrity is Medium)* **THEN** *(Trustworthiness is T)*

.
.
.

$\mathcal{T}u^{27}$ = **IF** *(Ability is High) and (Benevolence is High) and (Integrity is Medium)* **THEN** *(Trustworthiness is VT)*

The canonical form of these fuzzy if-then rules is partial rules consisting of fuzzy prepositions and fuzzy rules.

1) FUZZY INFERENCE ENGINE

Our Mamdani TAMFIS uses fuzzy set theory to map input features (personality traits in layer-1 and trustworthiness dimensions in layer-2) to the corresponding outputs (trustworthiness dimensions in layer-1 and trustworthiness level in layer-2). In input-output product space, if-then rules implemented in TAMFIS are construed as fuzzy relations. This set of rules can be inferred in two ways; composition-based and rule-based. In current implementation composition-based inference is applicable; it combines fuzzy rules through their inner product and views it as a single if-then rule.

$\mathcal{P}u^e$ and $\mathcal{T}u^1$ be the fuzzy relation representing an arbitrary fuzzy if-then rule from the rule base, i.e.

$$\mathcal{P}u^e = \mathcal{O}^e \times C^e \times \varepsilon^e \times \mathcal{A}^e \times \mathcal{N}^e \tag{5}$$
$$\mathcal{T}u^x = a^x \times b^x \times i^x \tag{6}$$

Eq. (5) and (6) holds as:

$$\mu_{OPN \cap \mathrm{CON} \cap \mathrm{EXT} \cap \mathrm{AGR} \cap \mathrm{NEO}}$$
$$= \mu_{OPN(\mathcal{O})} \cap \mu_{CON()} \cap \mu_{EXT(\varepsilon)} \cap \mu_{AGR(\mathcal{A})} \cap \mu_{AGR(\mathcal{N})} \tag{7}$$

$$\mu_{ability \cap benevolence \cap integrity}$$
$$= \mu_{ability(a)} \cap \mu_{benevolence(b)} \cap \mu_{integrity(i)} \tag{8}$$

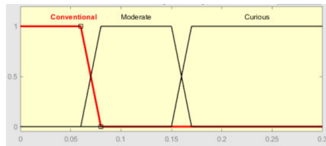
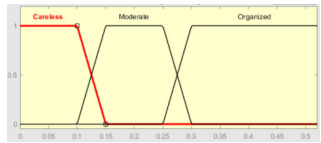
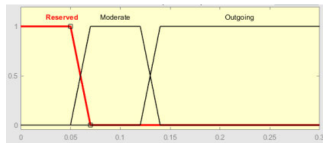
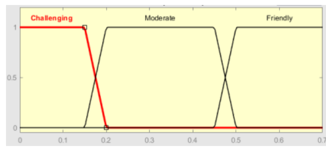
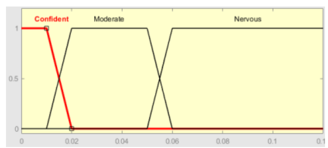
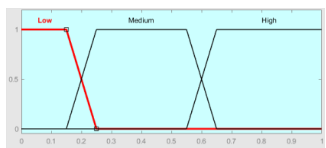
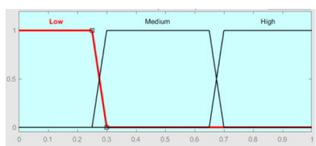
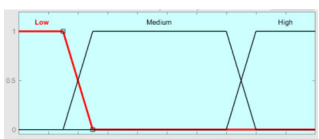
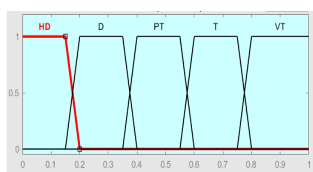
The rules from section 4.6 are then deduced as a fuzzy relation $\mathcal{Q}_{243}$ as:

$$\mathcal{Q}_{243} = \bigcup_{e=1}^{243} \mathcal{P}u^e \tag{9}$$

$$\mathcal{Q}_{27} = \bigcup_{e=1}^{27} \mathcal{T}u^e \tag{10}$$

Eq. (9) and (10) show Mamdani combinations for layer-1 and layer-2 of the proposed trustworthiness assessment model.

TABLE 2. Membership functions for TAMFIS layer-1 and layer-2.

Layer	Variable	Membership functions	Plot
LAYER – 1 Input Variables	OPN $(\mu_{OPN(\mathcal{O})})$	$\mu_{OPN,Conventional}(\mathcal{O}) = \left\{max\left(min\left(1,\frac{0.08-\mathcal{O}}{0.02}\right),0\right)\right\}$ $\mu_{OPN,Moderate}(\mathcal{O}) = \left\{max\left(min\left(\frac{\mathcal{O}-0.06}{0.02},1,\frac{0.17-\mathcal{O}}{0.02}\right),0\right)\right\}$ $\mu_{OPN,Curious}(\mathcal{O}) = \left\{max\left(min\left(\frac{\mathcal{O}-0.17}{0.02},1\right),0\right)\right\}$	
LAYER – 1 Input Variables	CON $(\mu_{CON(\mathbb{C})})$	$\mu_{CON,Careless}(\mathbb{C}) = \left\{max\left(min\left(1,\frac{0.15-\mathbb{C}}{0.05}\right),0\right)\right\}$ $\mu_{CON,Moderate}(\mathbb{C}) = \left\{max\left(min\left(\frac{\mathbb{C}-0.1}{0.05},1,\frac{0.3-\mathbb{C}}{0.05}\right),0\right)\right\}$ $\mu_{CON,Organized}(\mathbb{C}) = \left\{max\left(min\left(\frac{\mathbb{C}-0.3}{0.05},1\right),0\right)\right\}$	
LAYER – 1 Input Variables	EXT $(\mu_{EXT(\varepsilon)})$	$\mu_{EXT,Reserved}(\varepsilon) = \left\{max\left(min\left(1,\frac{0.05-\varepsilon}{0.02}\right),0\right)\right\}$ $\mu_{EXT,Moderate}(\varepsilon) = \left\{max\left(min\left(\frac{\varepsilon-0.1}{0.02},1,\frac{0.14-\varepsilon}{0.02}\right),0\right)\right\}$ $\mu_{EXT,Outgoing}(\varepsilon) = \left\{max\left(min\left(\frac{\varepsilon-0.14}{0.02},1\right),0\right)\right\}$	
LAYER – 1 Input Variables	AGR $(\mu_{AGR(\mathcal{A})})$	$\mu_{AGR,Challenging}(\mathcal{A}) = \left\{max\left(min\left(1,\frac{0.2-\mathcal{A}}{0.05}\right),0\right)\right\}$ $\mu_{AGR,Moderate}(\mathcal{A}) = \left\{max\left(min\left(\frac{\mathcal{A}-0.15}{0.05},1,\frac{0.5-\mathcal{A}}{0.05}\right),0\right)\right\}$ $\mu_{AGR,Friendly}(\mathcal{A}) = \left\{max\left(min\left(\frac{\mathcal{A}-0.5}{0.05},1\right),0\right)\right\}$	
LAYER – 1 Input Variables	NEO $(\mu_{AGR(\mathcal{N})})$	$\mu_{NEO,Confident}(\mathcal{N}) = \left\{max\left(min\left(1,\frac{0.02-\mathcal{N}}{0.01}\right),0\right)\right\}$ $\mu_{NEO,Moderate}(\mathcal{N}) = \left\{max\left(min\left(\frac{\mathcal{N}-0.01}{0.01},1,\frac{0.06-\mathcal{N}}{0.01}\right),0\right)\right\}$ $\mu_{NEO,Nervous}(\mathcal{N}) = \left\{max\left(min\left(\frac{\mathcal{N}-0.06}{0.01},1\right),0\right)\right\}$	
LAYER – 1 Input Variables / LAYER – 2 Input Variables	Ability $(\mu_{ability(a)})$	$\mu_{abulity,low}(a) = \left\{max\left(min\left(1,\frac{0.25-a}{0.1}\right),0\right)\right\}$ $\mu_{abulity,medium}(a) = \left\{max\left(min\left(\frac{a-0.15}{0.1},1,\frac{0.65-a}{0.1}\right),0\right)\right\}$ $\mu_{abulity,high}(a) = \left\{max\left(min\left(\frac{a-0.65}{0.1},1\right),0\right)\right\}$	
LAYER – 1 Input Variables / LAYER – 2 Input Variables	Benevolence $(\mu_{benevolence(b)})$	$\mu_{benevolence,low}(b) = \left\{max\left(min\left(1,\frac{0.3-b}{0.05}\right),0\right)\right\}$ $\mu_{benevolence,medium}(b) = \left\{max\left(min\left(\frac{b-0.25}{0.05},1,\frac{0.7-b}{0.05}\right),0\right)\right\}$ $\mu_{benevolence,high}(b) = \left\{max\left(min\left(\frac{b-0.7}{0.05},1\right),0\right)\right\}$	
LAYER – 1 Input Variables / LAYER – 2 Input Variables	Integrity $(\mu_{integrity(i)})$	$\mu_{integrity,low}(i) = \left\{max\left(min\left(1,\frac{0.15-i}{0.1}\right),0\right)\right\}$ $\mu_{integrity,medium}(i) = \left\{max\left(min\left(\frac{i-0.15}{0.1},1,\frac{0.8-i}{0.1}\right),0\right)\right\}$ $\mu_{integrity,high}(i) = \left\{max\left(min\left(\frac{i-0.7}{0.1},1\right),0\right)\right\}$	
LAYER – 2 Output Variable	Trustworthiness $(\mu_{trustworthiness(\mathcal{T})})$	$\mu_{trustworthiness,HD}(\mathcal{T}) = \left\{max\left(min\left(1,\frac{0.15-\mathcal{T}}{0.05}\right),0\right)\right\}$ $\mu_{trustworthiness,D}(\mathcal{T}) = \left\{max\left(min\left(\frac{\mathcal{T}-0.15}{0.05},1,\frac{0.4-\mathcal{T}}{0.05}\right),0\right)\right\}$ $\mu_{trustworthiness,PT}(\mathcal{T}) = \left\{max\left(min\left(\frac{\mathcal{T}-0.35}{0.05},1,\frac{0.6-\mathcal{T}}{0.05}\right),0\right)\right\}$ $\mu_{trustworthiness,T}(\mathcal{T}) = \left\{max\left(min\left(\frac{\mathcal{T}-0.55}{0.1},1,\frac{0.8-\mathcal{T}}{0.1}\right),0\right)\right\}$ $\mu_{trustworthiness,high}(\mathcal{T}) = \left\{max\left(min\left(\frac{\mathcal{T}-0.75}{0.1},1\right),0\right)\right\}$	

2) PRODUCT INFERENCE ENGINE

Let $\wp$, $\wp'$ and φ are the fuzzy sets and the inputs and outputs for the proposed TAMFIS respectively, then by observing $\mathcal{Q}_{243}$ and $\mathcal{Q}_{27}$ as a single fuzzy rule, the output of fuzzy inference is obtained. Composition based Mamdani fuzzy inference engine is therefore used here. The product inference engine is obtained through union combination of the individual rule base, product of all t-norm operators, and Mamdani product implications as:

$$\mu_{\wp}\,(Trust\ Dimensions) = max_{1\leq z\leq 243}\left[\prod_{y=1}^{243}\begin{pmatrix}\left(\mu_{OPN_y,\mathrm{CON}_y,\mathrm{EXT}_y,\mathrm{AGR}_y,\mathrm{NEO}_y}(\mathcal{O},C,\varepsilon,\mathcal{A},\mathcal{N})\right)\\ \left(\mu_{ability_z,benevolence_z,integrity_z}(\mathrm{a,b,i})\right)\end{pmatrix}\right] \quad (11)$$

$$\mu_{\varphi}\,(Trust\ worthiness) = max_{1\leq v\leq 27}\left[\prod_{v=1}^{27}\left(\left(\mu_{ability_v,benevolence_v,integrity_v}(\mathrm{a,b,i})\right)\left(\mu_{x(x)}\right)\right)\right] \quad (12)$$

Here x represents the domain of discourse for output, provided the fuzzy sets $\wp$ and $\wp'$ are obtained as inputs to layer-1 and layer-2 to calculate φ.

3) DE-FUZZIFIER

Fuzzy outputs are converted to receive their corresponding single crisp values. Several methods of defuzzification are available; defuzzifier can be implemented through some common techniques like max or mean-max membership principle, weighted average, centroid method. The current study has utilized a centroid type of De-fuzzifier. Centroid defuzzifier describes the transformation of the fuzzy output generated by the trustworthiness assessment inference engine to frangible using analogous membership functionalities in contrast to those used by the fuzzifier.

Defuzzifier maps the fuzzy set φ in eq.12 to a crisp point ξ^* for layer-1 and ξ^{**} for layer-2. Defuzzifier specifies a point in the output universe of discourse that gives the best representation of the fuzzy set φ.

$$\xi^* = \frac{\int \wp\mu_{\wp}(\wp)d_{\wp}}{\int \mu_{\wp}(\wp)d_{\wp}} \quad (13)$$

The layer-2 output will then take the form:

$$\xi^{**} = \frac{\int \varphi\mu_{\varphi}(\varphi)d_{\varphi}}{\int \mu_{\varphi}(\varphi)d_{\varphi}} \quad (14)$$

Eq. (13) and (14) calculates the crisp output values for the trustworthiness dimensions, provided the fuzzy set of personality traits and trustworthiness of human collaborator respectively.

4) FUZZY LOGIC SIMULATION AND RESULTS

Figure-3(a–f) represents the defuzzifier's graphical representations of the proposed TAMFIS. Figure 4(a), depicts the cognitive behavior to ability trustworthiness dimension concerning conscientiousness and openness. Since it has been observed that conscientiousness holds a direct relationship with the ability of a human whereas more open the person is lesser ability to hold the secret is observed. The proposed trustworthiness assessment model, therefore, portrays similar behavior as described in [46]. Similar trends have been observed for the other two trustworthiness dimensions (benevolence and integrity).

With rising levels of agreeableness and extroversion, a high rise in benevolence can be seen in figure 4-b, whereas, since the high value of openness descends human's trustworthiness level, therefore, its presence in figure 4-c and 4-d has irregular effects on overall trends of benevolence and integrity. Since the personality traits are mutually related and affect the trustworthiness dimensions and are seen to effects the cognitive trust levels in a collective fashion. The proposed model is found to exhibit similar behavior under these considerations.

Trustworthiness dimensions have been found to influence the overall trustworthiness perception of human collaborators in a linear fashion. Whereas considering the three dimensions are correlated to each other and influence trustworthiness accordingly, Figure- 3(e) and 3(f) depict the effects of ability, benevolence, and integrity on the trustworthiness of human mate.

5) DEMPSTER SHAFER THEORY (DST) BASED TRUSTWORTHINESS ASSESSMENT

Dempster–Shafer theory (DST) provides a general basis for reasoning under uncertainty [55]–[57]. DST has a deep underlying connection with probability theories in the context of statistical inference. The technique used in [43] using Dempster Shafer Theory is preowned here for the comparison of results obtained in section 4.1 from proposed TAMFIS. Accumulatively, there are eight influencing factors (openness "$\mathcal{O}$", conscientiousness "", extroversion "$\mathcal{E}$", agreeableness "$\mathcal{A}$", neuroticism "$\mathcal{N}$", ability "a", benevolence "b" and integrity "i", serving as trustworthiness dimensions. We assumed,

i. There are two agent's artificial cognitive agent and a human.
ii. An artificial cognitive agent is a trustor whereas a human is a trustee.
iii. Artificial agent assesses human personality traits to predict the trustworthiness of human mate thereby developing a trust relationship.

DST application for trustworthiness assessment follows definitions leading to define belief intervals:

Def.-1: V, decrement frame is a set consist of $\left\{\mathrm{D}\ (\text{dependence})\ , \tilde{\mathrm{D}}(\text{independence})\right\}$, we may write:

$$P\,(V) = \left\{\varphi, \{D\}, \left\{\tilde{D}\right\}, \left\{D, \tilde{D}\right\}\right\} \quad (15)$$

Def.-2: $\mathrm{m}_{\rho t}$ be the function for probability assignment function from cognitive agent to human, is defined as,

$$m_{\rho\tau} : P\,(V) \to [0, 1]\,,$$
$$\text{Where} \quad m_{\rho\tau} = m_{\rho t} \circ m_{t\tau}$$

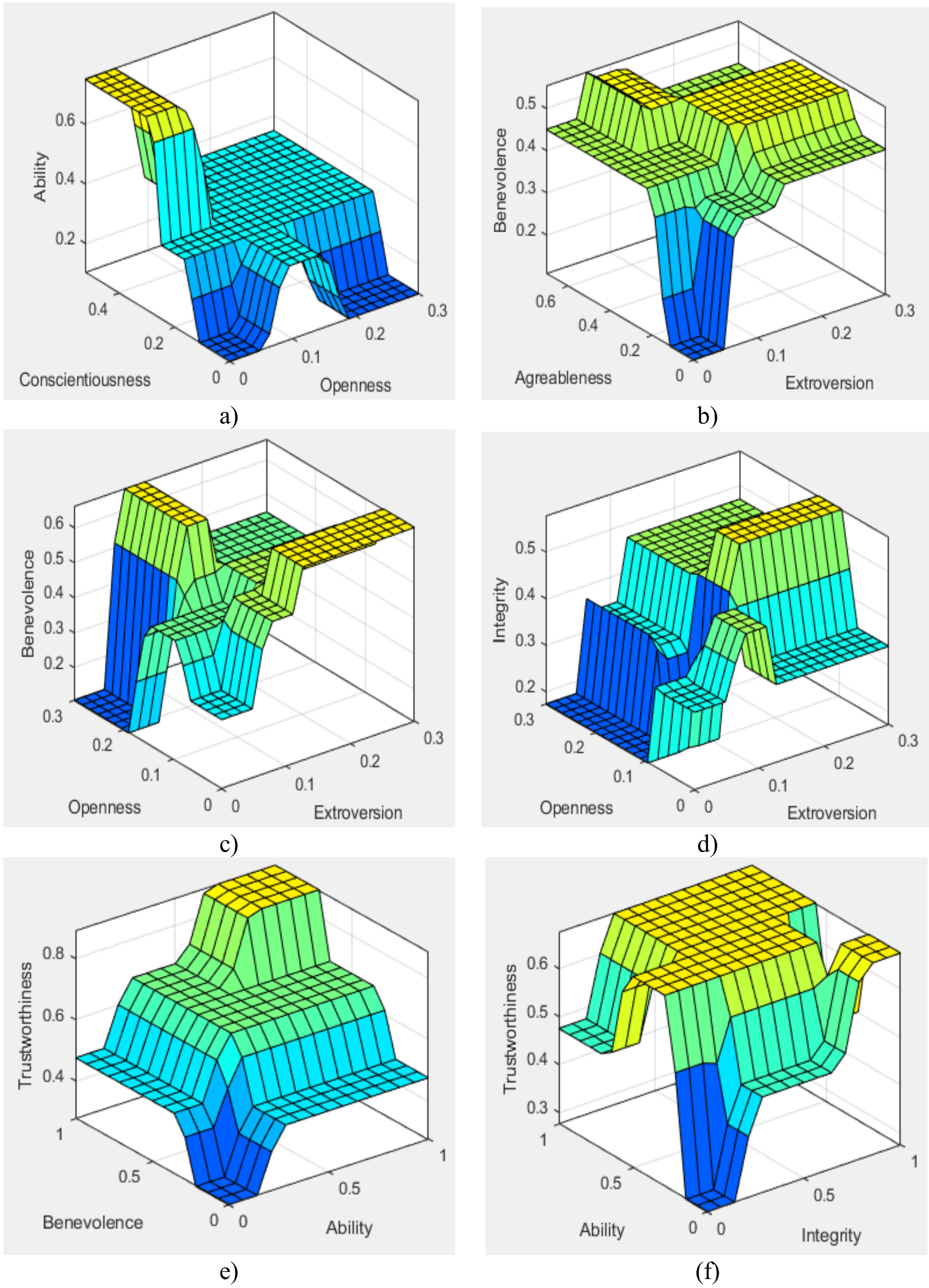


FIGURE 3. **Proposed rule surfaces of trustworthiness assessment.**

We have, $\rho = \{\mathcal{O}, C, \varepsilon, \mathcal{A}, \mathcal{N}\}$ and $t = \{a,b,i\}$, $\tau =$ trustworthiness

$$\mathrm{m}_{\rho\tau}(\emptyset) = 0;$$

$$\sum\nolimits_{w\subseteq P(V)} m_{\rho\tau}(w) = m_{\rho\tau}(\{D\})$$
$$+m_{\rho\tau}(\left\{\tilde{D}\right\} + m_{\rho\tau}(\left\{D, \tilde{D}\right\}) = 1$$

where, $\mathrm{m}_{\rho t}$ represents how trustworthy the human is perceived by a cognitive agent.

Def.-3: $\mathrm{Dep}_{\rho\tau}$, dependence function is defined as, $\mathrm{Dep}_{\rho\tau} : \mathrm{P}(\mathrm{V}) \rightarrow [0, 1]$;

$$\mathrm{Dep}_{\rho\tau}(\{\mathrm{D}\}) = \sum_{\mathrm{w}\subseteq\mathrm{D}\subseteq\mathrm{P(V)}} \mathrm{m}_{\rho\tau}(\mathrm{w})$$

Def.-4: Plausibility function $\mathrm{Pl}_{\rho\tau} : \mathrm{P}(\mathrm{V}) \rightarrow [0, 1]$; $\mathrm{Pl}_{\rho\tau}(\mathrm{w}) = 1 - \mathrm{Dep}_{\rho\tau}(\tilde{\mathrm{w}})$ is given as:

$$\mathrm{Pl}_{\rho\tau}(\{\mathrm{D}\}) = \mathrm{m}_{\rho\tau}(\{\mathrm{D}\}) + \mathrm{m}_{\rho\tau}\left(\{\mathrm{D}, \tilde{\mathrm{D}}\}\right) = 1 - \mathrm{Dep}_{\rho\tau}(\tilde{\mathrm{w}})$$

$\mathrm{Pl}_{\rho\tau}(\{D\})$ depicts the degree to which cognitive agent is not independent of human, therefore interval for interdependence is $\left[\mathrm{Dep}_{\rho\tau}(\{D\}), \mathrm{Pl}_{\rho\tau}(\{D\})\right]$. Interdependence transfer and interdependence clustering mechanism needed to association pieces of evidence are derived here.

In the proposed system, trustworthiness (τ) is assessed as a composition of perceived personality traits (ρ) mapped onto the trustworthiness dimensions (t). If the level of trustworthiness τ depends on personality traits ρ is represented by interdependence interval $\left[\mathrm{Dep}_{\rho t}(\{D\}), \mathrm{Pl}_{\rho t}(\{D\})\right]$ and interdependence interval $\left[\mathrm{Dep}_{t\tau}(\{D\}), \mathrm{Pl}_{t\tau}(\{D\})\right]$ shows to what extent trustworthiness depend on trustworthiness dimensions, the principle of attenuation gives,

$$\mathrm{m}_{\rho t}(\{D\}) = \mathrm{m}_{\rho t}(\{D\})\,\mathrm{m}_{\rho t}(\{D\})\,\mathrm{m}_{\rho t}(\{D\}) \tag{16}$$

$$\mathrm{m}_{\rho t}\left(\left\{\tilde{D}\right\}\right) = \mathrm{m}_{\rho t}\left(\left\{\tilde{D}\right\}\right)\mathrm{m}_{\rho t}\left(\left\{\tilde{D}\right\}\right)\mathrm{m}_{\rho t}\left(\left\{\tilde{D}\right\}\right) \tag{17}$$

$$\mathrm{m}_{t\tau}(\{D\}) = \mathrm{m}_{t\tau}(\{D\})\,\mathrm{m}_{t\tau}(\{D\})\,\mathrm{m}_{t\tau}(\{D\}) \tag{18}$$

$$\mathrm{m}_{t\tau}\left(\left\{\tilde{D}\right\}\right) = \mathrm{m}_{t\tau}\left(\left\{\tilde{D}\right\}\right)\mathrm{m}_{t\tau}\left(\left\{\tilde{D}\right\}\right)\mathrm{m}_{t\tau}\left(\left\{\tilde{D}\right\}\right) \tag{19}$$

whereas Plausibility could be found out as:

$$\begin{aligned}\mathrm{Pl}_{\rho\tau}(\{D\}) &= 1 - \mathrm{Dep}_{\rho\tau}\left(\left\{\tilde{D}\right\}\right)\\ &= \mathrm{Pl}_{\rho t}(\{D\}) + \mathrm{Pl}_{t\tau}(\{D\})\\ &\quad -\mathrm{Pl}_{\rho t}(\{D\})\,\mathrm{Pl}_{t\tau}(\{D\})\end{aligned} \tag{20}$$

The two independence sets of probability assignment m_ρ can be combined as.

For the given five evidence, to support interdependence, there are five intervals $\left[\mathrm{Dep}_\gamma(\{D\}), \mathrm{Pl}_\gamma(\{D\})\right]$, $1 \le \gamma \le 5$ and their joint basic belief assignment

$$\mathrm{m}_{\rho t} = \mathrm{m}_{\mathcal{O}} \oplus \mathrm{m} \oplus \mathrm{m}_{\varepsilon} \oplus \mathrm{m}_{\mathcal{A}} \oplus \mathrm{m}_{\mathcal{N}} \tag{21}$$

Similarly for trustworthiness dimensions,

$$\mathrm{m}_{t\tau} = \mathrm{m}_a \oplus \mathrm{m}_b \oplus \mathrm{m}_i \tag{22}$$

and can be written as,

$$\mathrm{m}_{\rho t}(F) = \begin{cases} 0, & \text{if } F = \phi \\ K \sum\limits_{\cap F_i = F} \prod\limits_{1 \le i \le 5} \mathrm{m}_i(F), & \text{if } F \ne \phi \end{cases} \tag{23}$$

$$\mathrm{m}_{t\tau}(F') = \begin{cases} 0, & \text{if } F' = \phi \\ K \sum\limits_{\cap F'_i = F'} \prod\limits_{1 \le i \le 3} \mathrm{m}_i(F'), & \text{if } F' \ne \phi \end{cases} \tag{24}$$

where F and F′ are the intersections of all the subsets, whereas K^{-1} is normalized factor and

$$K^{-1} = \sum_{\cap F_{i \ne \phi}} m_1(F_1)m_2(F_2)m_3(F_3)m_4(F_4)m_5(F_5) \tag{25}$$

Assuming ρ as evidences between agent and human that support interdependence at interdependence intervals [0.60, 0.92], [0.60, 0.90], [0.60, 0.95], [0.65, 0.95] and [0.7, 0.85] Clustering mechanism from evidences $\mathcal{O}$" and "C" then provides:

$$\mathrm{Dep}_{\mathcal{O}}(\{D\}) = 0.60; \quad \mathrm{Pl}_{\mathcal{O}}(\{D\}) = 0.92$$

$$\mathrm{Dep}_{C}(\{D\}) = 0.60; \quad \mathrm{Pl}_{C}(\{D\}) = 0.90$$

clustering mechanism gives in Eq. (24)

$$\begin{aligned}K^{-1} &= 1 - \mathrm{Dep}_{\mathcal{O}}(\{D\}) - \mathrm{Dep}_i(\{D\})\\ &\quad +\mathrm{Dep}_{\mathcal{O}}(\{D\})\,\mathrm{Pl}_i(\{D\})\\ &\quad +\mathrm{Dep}_i(\{D\})\,\mathrm{Pl}_{\mathcal{O}}(\{D\})\,K^{-1} = 0.892\\ \mathrm{Dep}_{\mathcal{O}C}(\{D\}) &= K[\mathrm{Dep}_{\mathcal{O}}(\{D\})\,\mathrm{Pl}_C(\{D\})\\ &\quad +\mathrm{Dep}_C(\{D\})\,\mathrm{Pl}_{\mathcal{O}}(\{D\})\\ &\quad -\mathrm{Dep}_{\mathcal{O}}(\{D\})\,\mathrm{Dep}_C(\{D\})]\\ &\quad \times \mathrm{Dep}_{\mathcal{O}C}(\{D\}) \approx 0.821\\ \mathrm{Pl}_{\mathcal{O}C}(\{D\}) &= K\mathrm{Pl}_{\mathcal{O}}(\{D\})\,\mathrm{Pl}_C(\{D\}) \approx 0.928\end{aligned}$$

For "$\mathcal{O}$" and "C", interdependence interval is $[\mathrm{Dep}_{\mathcal{O}C}(\{D\}), \mathrm{Pl}_{\mathcal{O}C}(\{D\})] = [0.821, 0.928]$

Similarly "ε" and "$\mathcal{A}$" evidences can also be combined as:

$$\mathrm{Dep}_{\varepsilon}(\{D\}) = 0.60; \quad \mathrm{Pl}_{\varepsilon}(\{D\}) = 0.92$$

$$\mathrm{Dep}_{\mathcal{A}}(\{D\}) = 0.65; \quad \mathrm{Pl}_{\mathcal{A}}(\{D\}) = 0.95$$

Then

$$\begin{aligned}K^{-1} &= 1 - \mathrm{Dep}_{\mathcal{A}}(\{D\}) - \mathrm{Dep}_{\varepsilon}(\{D\})\\ &\quad +\mathrm{Dep}_{\varepsilon}(\{D\})\,\mathrm{Pl}_{\mathcal{A}}(\{D\})\\ &\quad +\mathrm{Dep}_{\mathcal{A}}(\{D\})\,\mathrm{Pl}_{\varepsilon}(\{D\})\,K^{-1} = 0.937\\ \mathrm{Dep}_{\varepsilon\mathcal{A}}(\{D\}) &= K[\mathrm{Dep}_{\varepsilon}(\{D\})\,\mathrm{Pl}_{\mathcal{A}}(\{D\})\\ &\quad +\mathrm{Dep}_{\mathcal{A}}(\{D\})\,\mathrm{Pl}_{\varepsilon}(\{D\})\\ &\quad -\mathrm{Dep}_{\varepsilon}(\{D\})\,\mathrm{Dep}_{\mathcal{A}}(\{D\})]\\ &\quad \times \mathrm{Dep}_{\varepsilon\mathcal{A}}(\{D\}) \approx 0.851\\ Pl_{\varepsilon\mathcal{A}}(\{D\}) &= KPl_{\varepsilon}(\{D\})\,Pl_{\mathcal{A}}(\{D\}) \approx 0.963\end{aligned}$$

"ε" and "$\mathcal{A}$" have, therefore, the interdependence interval $[\mathrm{Dep}_{\varepsilon\mathcal{A}}(\{D\}), \mathrm{Pl}_{\varepsilon\mathcal{A}}(\{D\})] = [0.851, 0.963]$

We have,

$$\mathrm{Dep}_{\mathcal{O}C}(\{D\}) = 0.821; \quad Pl_{\mathcal{O}C}(\{D\}) = 0.928$$

$$Dep_{\varepsilon\mathcal{A}}(\{D\}) = 0.851; \quad Pl_{\varepsilon\mathcal{A}}(\{D\}) = 0.963$$

The combined effect of trustworthiness dimensions from agent to human, $[\mathrm{Dep}_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\}), \mathrm{Pl}_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\})]$ is calculated

$$\begin{aligned}K^{-1} &= 1 - Dep_{\mathcal{O}C}(\{D\}) - Dep_{\varepsilon\mathcal{A}}(\{D\})\\ &\quad +Dep_{\mathcal{O}C}(\{D\})\,Pl_{\varepsilon\mathcal{A}}(\{D\})\\ &\quad +Dep_{\varepsilon\mathcal{A}}(\{D\})\,Pl_{\mathcal{O}C}(\{D\})\\ Dep_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\}) &= K[Dep_{\mathcal{O}C}(\{D\})\,Pl_{\varepsilon\mathcal{A}}(\{D\})\\ &\quad +Dep_{\varepsilon\mathcal{A}}(\{D\})\,Pl_{\mathcal{O}C}(\{D\})\\ &\quad -Dep_{\mathcal{O}C}(\{D\})\,Dep_{\varepsilon\mathcal{A}}(\{D\})]\\ Dep_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\}) &\approx 0.971 Pl_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\}) = KPl_{\mathcal{O}C}(\{D\})\\ Pl_{\varepsilon\mathcal{A}}(\{D\}) &\approx 0.984[Dep_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\}), Pl_{\mathcal{O}C\varepsilon\mathcal{A}}(\{D\})]\\ &= [0.971, 0.984]\end{aligned}$$

TABLE 3. Simulation results comparison table of TAMFIS with Dempster Shafer theory.

PROPOSED TRUSTWORTHINESS ASSESSMENT MFIS						DST		
Personality Traits		Assessed Trustworthiness Dimensions		Trustworthiness		Evidence	Trustworthiness Dimensions	Assessed Trust-worthiness
CRISP	LINGUISTIC	CRISP	LINGUISTIC	CRISP	LINGUISTIC			
$\mathcal{O}=0.04$	Conventional	a =0.19	Low	0.174	Highly Deceptive	$\mathcal{O}=[0.1,0.4]$	a =[0.2,0.8] b =[0.25,0.7] i =[0.5,0.8]	$\mathcal{T}=[0.174,0.176]$
C= 0.10	Careless					C= [0.1,0.5]		
$\mathcal{E}=0.05$	Reserved	b =0.22	Low			$\mathcal{E}=[0.3,0.45]$		
$\mathcal{A}=0.10$	Challenging	i =0.15	low			$\mathcal{A}=[0.5,0.7]$		
$\mathcal{N}=0.09$	Nervous					$\mathcal{N}=[0.2,0.5]$		
$\mathcal{O}=0.05$	Conventional	a =0.16	Low	0.377	Deceptive	$\mathcal{O}=[0.2,0.55]$	a =[0.5,0.8] b =[0.25,0.9] i =[0.3,0.6]	[27] = [0.372,0.38]
C= 0.20	Moderate					C= [0.25,0.52]		
$\mathcal{E}=0.06$	Reserved	b =0.31	Medium			$\mathcal{E}=[0.35,0.45]$		
$\mathcal{A}=0.22$	Moderate	i =0.19	low			$\mathcal{A}=[0.5,0.65]$		
$\mathcal{N}=0.11$	Nervous					$\mathcal{N}=[0.3,0.65]$		
$\mathcal{O}=0.10$	Moderate	a =0.23	Medium	0.564	Partially Trustworthy	$\mathcal{O}=[0.3,0.7]$	a =[0.5,0.8] b =[0.25,0.9] i =[0.3,0.7]	$\mathcal{T}=[0.56,0.571]$
C= 0.23	Organized					C= [0.25,0.52]		
$\mathcal{E}=0.12$	Moderate	b =0.32	Medium			$\mathcal{E}=[0.2,0.45]$		
$\mathcal{A}=0.11$	Challenging	i =0.23	Medium			$\mathcal{A}=[0.5,0.7]$		
$\mathcal{N}=0.02$	Confident					$\mathcal{N}=[0.4,0.8]$		
$\mathcal{O}=0.23$	Curious	a =0.59	Medium	0.778	Trustworthy	$\mathcal{O}=[0.25,0.65]$	a =[0.45,0.7] b =[0.45,0.5] i =[0.45,0.8]	$\mathcal{T}=[0.77,0.78]$
C= 0.31	Moderate					C= [0.55,0.75]		
$\mathcal{E}=0.13$	Outgoing	b =0.87	High			$\mathcal{E}=[0.3,0.5]$		
$\mathcal{A}=0.15$	Challenging	i =0.81	High			$\mathcal{A}=[0.4,0.7]$		
$\mathcal{N}=0.019$	Confident					$\mathcal{N}=[0.6,0.7]$		
$\mathcal{O}=0.26$	Curious	a =0.95	High	0.89	Very Trustworthy	$\mathcal{O}=[0.55,0.7]$	a =[0.5,0.9] b =[0.6,0.8] i =[0.5,0.9]	$\mathcal{T}=[0.90,0.91]$
C=0.39	Organized					C= [0.65,0.75]		
$\mathcal{E}=0.14$	Outgoing	b =0.93	High			$\mathcal{E}=[0.35,0.55]$		
$\mathcal{A}=0.65$	Friendly	i =0.98	High			$\mathcal{A}=[0.45,0.7]$		
$\mathcal{N}=0.01$	Confident					$\mathcal{N}=[0.15,0.75]$		

Interdependence interval for personality traits are then calculated combining the evidence $\mathcal{N} = [0.7, 0.95]$ to obtain,

$$[Dep_{\mathcal{OCEAN}}(\{D\}), Pl_{\mathcal{OCEAN}}(\{D\})] = [0.989, 0.992]$$

Interdependence interval for trustworthiness dimensions is therefore [0.989, 0.992].

Personality traits interval is therefore calculated in eq. (27). The impact of this interval upon human's trustworthiness assessment is then calculated by using join belief assignment as $\mathrm{m}_{\rho\tau} = \mathrm{m}_{\rho t} \oplus \mathrm{m}_{t\tau}$. Here $\mathrm{m}_{t\tau}$ is calculated according to [43]. The trustworthiness interval between agent and human is then calculated as $\mathrm{m}_{\rho\tau} = [0.989, 0.992]$ for $\mathrm{m}_{t\tau} = [0.991, 0.993]$, yielding trustworthiness interval as:

$$\mathrm{m}_{\rho\tau} \approx [0.992, 0.994]$$

Table. 2 presents three trustworthiness assessment results for five random cases (Highly Deceptive, Deceptive, Partially Trustworthy, Trustworthy, and Very Trustworthy) with the ones obtained from the model implemented through DST. Estimation and comparison between resultant trustworthiness intervals from DST with crisp values of proposed Trustworthiness assessment model if performed. Both methods have been found to support the assessment of trustworthiness to high levels offering negligible difference.

V. CONCLUSION AND FUTURE WORK

The current research proposed a cognitive trustworthiness assessment fuzzy-based model (TAMFIS) to build a trust-based relationship between artificial cognitive agents and humans. The proposed model is based on perceived human personality traits. The model has utilized the personality traits assessment model to improve the cognitive trustworthiness assessment capability of an artificial agent. The system is therefore assumed to have a capability to identify trustworthy and malicious collaborators based on his personality traits even when it had limited or no previous interactions. The proposed fuzzy-based trustworthiness assessment model for an artificial agent can interact with its teammates and estimate their trustworthiness to make autonomous decisions about its actions. Further implementation for the proposed model has been carried out through Dempster Shafer Theory (DST). We have evaluated our proposed trustworthiness assessment model using DST and the results are found to be similar. Future implementation of the model is planned through the LSTM recurrent network to predict human trustworthiness.

ACKNOWLEDGMENT

(Sadaf Hussain and Rizwan Ali Naqvi are co-first authors.)

REFERENCES

[1] E. E. Levine, J. T. B. Bitterly, T. R. Cohen, and M. E. Schweitzer, "Who is trustworthy? Predicting trustworthy intentions and behavior," *J. Personality Social Psychol.*, vol. 115, no. 3, pp. 468–494, 2018.

[2] R. B. Lount, Jr. and N. C. Pettit, "The social context of trust: The role of status," *Org. Behav. Hum. Decis. Processes*, vol. 117, no. 1, pp. 15–23, Jan. 2012.

[3] Y. Zhengand, Y. Ren, and Ö. Özer, "Trust, trustworthiness, and information sharing in supply chains bridging China and the United States," in *Handbook of Information Exchange in Supply Chain Management*, vol. 60, 10th ed. Cham, Switzerland: Springer, 2016, pp. 2381–2617.

[4] Ö. Zheng and Y. Özer, "Establishing trust and trustworthiness for supply chain information sharing," in *Handbook of Information Exchange in Supply Chain Management*, 2017, pp. 287–312.

[5] N. R. Buchan, R. T. A. Croson, and S. Solnick, "Trust and gender: An examination of behavior and beliefs in the investment game," *J. Econ. Behav. Org.*, vol. 68, nos. 3–4, pp. 466–476, Dec. 2008.

[6] R. P. Larrick, "The social context of decisions," *Annu. Rev. Org. Psychol. Org. Behav.*, vol. 3, pp. 441–467, Mar. 2016.

[7] R. Derfler-Rozin, M. Pillutla, and S. Thau, "Social reconnection revisited: The effects of social exclusion risk on reciprocity, trust, and general risk-taking," *Org. Behav. Hum. Decis. Process.*, vol. 112, no. 2, pp. 140–150, Jul. 2010.

[8] M. E. Schweitzer, J. C. Hershey, and E. T. Bradlow, "Promises and lies: Restoring violated trust," *Org. Behav. Hum. Decis. Processes*, vol. 101, no. 1, pp. 1–19, Sep. 2006.

[9] M. E. Schweitzer, T.-H. Ho, and X. Zhang, "How monitoring influences trust: A tale of two faces," *Manage. Sci.*, vol. 64, no. 1, pp. 253–270, Jan. 2018.

[10] R. B. Lount, Jr., "The impact of positive mood on trust in interpersonal and intergroup interactions," *J. Personality Social Psychol.*, vol. 98, no. 3, pp. 420–433, 2010.

[11] J. E. Anderson, T. Schlösser, D. Ehlebracht, D. Dunning, and D. Fetchenhauer, "Trust at zero acquaintance: More a matter of respect than expectation of reward," *J. Personality Social Psychol.*, vol. 107, no. 1, p. 122, 2014.

[12] A. W. Brooks, H. Dai, and M. E. Schweizer, "I'm sorry about the rain! Superfluous apologies demonstrate empathic concern and increase trust," *Social Psychol. Personality Sci.*, vol. 5, no. 4, pp. 467–474, 2014.

[13] A. W. Brooks, M. Schweitzer, and A. D. Galinsky. (Sep. 2015). *The Organizational Apology: A Step-by-Step Guide Magazine*. [Online]. Available: https://hbr.org/2015/09/the-organizational-apology

[14] R. R. McCrae and P. T. Costa, "The five-factor theory of personality," in *Handbook of Personality: Theory and Research*, O. P. John, R. W. Robin, and L. A. Pervin, Eds. New York, NY, USA: The Guilford Press, 2008, pp. 159–181.

[15] G. Matthews, P. A. Hancock, J. Lin, A. R. Panganiban, L. E. Reinerman-Jones, J. L. Szalma, and R. W. Wohleber, "Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems," *Personality Individual Differences*, vol. 169, no. 1, pp. 1–11, 2020.

[16] J. J. Mondak, *Personality and the Foundations of Political Behavior*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[17] A. Capiola, G. M. Alarcon, and M. D. Pfahler, "The role of human personality on trust in human-robot interaction," in *Trust in Human-Robot Interaction*. New York, NY, USA: Academic, 2021, ch. 7, pp. 159–178.

[18] J. J. Mondak, M. V. Hibbing, D. Canache, M. A. Seligson, and M. R. Anderson, "Personality and civic engagement: An integrative framework for the study of trait effects on political behavior," *Amer. Political Sci. Rev.*, vol. 104, no. 1, pp. 85–110, Feb. 2010.

[19] A. S. Gerber, G. A. Huber, D. Doherty, and C. M. Dowling, "Personality and the strength and direction of partisan identification," *Political Behav.*, vol. 34, no. 4, pp. 653–688, Dec. 2012.

[20] R. R. McCrae and P. T. Costa, Jr., *Personality in Adulthood: A Five-Factor Theory Perspective*, 2nd ed. New York, NY, USA: Guilford Press, 2003.

[21] P. T. Dinesen, A. S. Nørgaard, and R. Klemmensen, "The civic personality: Personality and democratic citizenship," *Political Stud.*, vol. 62, no. 1, pp. 134–152, Apr. 2014.

[22] R. R. McCrae and P. T. Costa, Jr., *Personality in Adulthood: A Five-Factor Theory Perspective*. New York, NY, USA: Guilford Press, 2005.

[23] A. Becker, T. Deckers, T. Dohmen, A. Falk, and F. Kosse, "The relationship between economic preferences and psychological personality measures," *Annu. Rev. Econ.*, vol. 4, no. 1, pp. 453–478, Sep. 2012.

[24] A. Ben-Ner and F. Halldorsson, "Trusting and trustworthiness: What are they, how to measure them, and what affects them," *J. Econ. Psychol.*, vol. 31, no. 1, pp. 64–79, Feb. 2010.

[25] J.-E. Lönnqvist, M. Verkasalo, P. C. Wichardt, and G. Walkowitz, "Personality disorder categories as combinations of dimensions: Translating cooperative behavior in borderline personality disorder into the five-factor framework," *J. Personality Disorders*, vol. 26, no. 2, pp. 298–304, Apr. 2012.

[26] P. Bobko, M. Schuelke, S. Jessup, C. Calhoun, and T. Ryan, "Suspicion, trust, and automation," Air Force Res. Lab., Wright-Patterson Air Force Base, OH, USA, Tech. Rep. AFRL-RH-WP-TR-2017-0002, 2017.
[27] T. R. Schneider, "Evaluations of stressful transactions: What's in an appraisal?" *Stress Health*, vol. 24, no. 2, pp. 151–158, 2008.
[28] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, Jul. 1995.
[29] R. Falcone and C. Castelfranchi, "Social trust: A cognitive approach," in *Trust and Deception in Virtual Societies*. Dordrecht, The Netherlands: Springer, 2001, pp. 55–90.
[30] M. Lahijanian and M. Lahijanian, "Social trust: A major challenge for the future of autonomous," in *Proc. AAAI Fall Symp. Cross-Disciplinary Challenges Auton. Syst.*, 2016, pp. 1–6.
[31] B. Kuipers, "How can we trust a robot?" *Commun. ACM*, vol. 61, no. 3, pp. 86–95, Feb. 2018.
[32] F. M. Hafizoğlu and S. Sen, "Understanding the influences of past experience on trust in human-agent teamwork," *ACM Trans. Internet Technol.*, vol. 19, no. 4, pp. 1–22, Nov. 2019.
[33] K. Mogens, M. Nielsen, and V. Sassone, "Trust models in ubiquitous computing," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 366, no. 1881, pp. 3781–3793, 2008.
[34] A. Herzig, E. Lorini, J. F. Hubner, and L. Vercouter, "A logic of trust and reputation," *Log. J. IGPL*, vol. 18, no. 1, pp. 214–244, Feb. 2010.
[35] E. Lorini, A. Herzig, and F. Moisan, "A simple logic of trust based on propositional assignments," in *The Goals of Cognition. Essays in Honor of Cristiano Castelfranchi*. London, U.K.: College Publications, 2013, ch. 1, pp. 407–419.
[36] J. J. Mondak, *Personality and the Foundations of Political Behavior*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[37] M. R. Anderson, "Community psychology, political efficacy, and trust," *Political Psychol.*, vol. 31, no. 1, pp. 59–84, Feb. 2010.
[38] T. Dohmen, A. Falk, D. Huffman, and U. Sunde, "Representative trust and reciprocity: Prevalence and determinants," *Econ. Inquiry*, vol. 46, no. 1, pp. 84–90, Jan. 2008.
[39] K. Hiraishi, S. Yamagata, C. Shikishima, and J. Ando, "Maintenance of genetic variation in personality through control of mental mechanisms: A test of trust, extraversion, and agreeableness," *Evol. Hum. Behav.*, vol. 29, no. 2, pp. 79–85, Mar. 2008.
[40] G. Loewenstein, S. Sah, and D. M. Cain, "The burden of disclosure: Increased compliance with distrusted advice," *J. Personality Social Psychol.*, vol. 104, no. 2, pp. 289–304, 2013.
[41] D. De Cremer, E. van Dijk, and M. M. Pillutla, "Explaining unfair offers in ultimatum games and their effects on trust: An experimental approach," *Bus. Ethics Quart.*, vol. 20, no. 1, pp. 107–126, Jan. 2010.
[42] A. Sapienza and R. Falcone, "Evaluating agents' trustworthiness within virtual societies in case of no direct experience," *Cognit. Syst. Res.*, vol. 64, pp. 164–173, Dec. 2020.
[43] S. Hussain, S. Abbas, T. Sohail, M. A. Khan, and A. Athar, "Estimating virtual trust of cognitive agents using multi layered socio-fuzzy inference system," *J. Intell. Fuzzy Syst.*, vol. 37, no. 2, pp. 2769–2784, Sep. 2019.
[44] M. Alkhamees, S. Alsaleem, M. Al-Qurishi, M. Al-Rubaian, and A. Hussain, "User trustworthiness in online social networks: A systematic review," *Appl. Soft Comput.*, vol. 103, pp. 107–159, May 2021.
[45] J. Müller and C. Schwieren, "Big five personality factors in the trust game," *J. Bus. Econ.*, vol. 90, no. 1, pp. 37–55, Feb. 2020.
[46] G. M. Alarcon, J. B. Lyons, J. C. Christensen, M. A. Bowers, S. L. Klosterman, and A. Capiola, "The role of propensity to trust and the five factor model across the trust process," *J. Res. Personality*, vol. 75, pp. 69–82, Aug. 2018.
[47] M. Patacchiola, A. Chella, S. Vinanzi, and A. Cangelosi, "Would a robot trust you? Developmental robotics model of trust and theory of mind," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 374, no. 1771, pp. 1–10, 2019.
[48] V. Seidita, F. Lanza, A. Chella, S. Vinanzi, A. Cangelosi, A. Chella, and V. Seidita, "A global workspace theory model for trust estimation in human-robot interaction," in *Proc. 7th Int. Workshop Artif. Intell. Cognition*, Manchester, U.K., 2019, pp. 104–112.
[49] R. Naqvi, M. Arsalan, and K. Park, "Fuzzy system-based target selection for a NIR camera-based gaze tracker," *Sensors*, vol. 17, no. 4, p. 862, Apr. 2017.
[50] F. Xiao, "EFMCDM: Evidential fuzzy multicriteria decision making based on belief entropy," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 7, pp. 1477–1491, Jul. 2020.
[51] G. Liu, F. Xiao, C.-T. Lin, and Z. Cao, "A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2677–2690, Nov. 2020.
[52] K. R. Park and R. A. Naqvi, "Discriminating between intentional and unintentional gaze fixation using multimodal-based fuzzy logic algorithm for gaze tracking system with NIR camera sensor," *Opt. Eng.*, vol. 55, no. 6, pp. 063109-1–063109-17, 2016.
[53] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 74–79, Mar. 2017.
[54] A. Becker, T. Deckers, T. Dohmen, A. Falk, and F. Kosse, "The relationship between economic preferences and psychological personality measures," *Annu. Rev. Econ.*, vol. 4, no. 1, pp. 453–478, Sep. 2012.
[55] C. Lucas and B. N. Araabi, "Generalization of the Dempster–Shafer theory: A fuzzy-valued measure," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 3, pp. 255–270, Jun. 1999.
[56] Y. Deng, "Generalized evidence theory," *Int. J. Speech Technol.*, vol. 43, no. 3, pp. 530–543, Oct. 2015.
[57] F. Xiao, "Generalization of Dempster–Shafer theory: A complex mass function," *Int. J. Speech Technol.*, vol. 50, no. 10, pp. 3266–3275, Oct. 2020.

SADAF HUSSAIN received the master's degree in physics from UET, Lahore, Pakistan, the master's degree in computer science from Iqra University, Lahore, and the M.Phil. degree in computer science from GC University, Lahore. She is currently pursuing the Ph.D. degree in computer science with the School of Computer Science, NCBA&E, Lahore. She has extensive experience in the development and teaching of computer science to undergraduate students and college students. She is also working with the Lahore Garrison Education System. Her main research interests include cognitive computing, soft computing, and artificial intelligence.

RIZWAN ALI NAQVI received the B.S. degree in computer engineering from COMSATS University, Pakistan, in 2008, the M.S. degree in electrical engineering from Karlstad University, Sweden, in 2011, and the Ph.D. degree in electronics and electrical engineering from Dongguk University, South Korea, in 2018. From 2011 to 2012, he was a Lecturer with the Department of Computer Science, Sharif College of Engineering and Technology, Pakistan. In 2012, he joined the Faculty of Engineering and Technology, The Superior College, Pakistan, as a Senior Lecturer. From 2018 to 2019, he worked as a Postdoctoral Researcher with Gachon University, South Korea. He is currently working as an Assistant Professor with Sejong University, South Korea. His research interests include gaze tracking, biometrics, computer vision, artificial intelligence, machine learning, deep learning, and medical imaging analysis.

SAGHEER ABBAS received the M.Phil. degree in computer science from the School of Computer Science, NCBA&E, Lahore, Pakistan, and the Ph.D. degree from the School of Computer Science, NCBA&E, in 2016. He is currently working as an Associate Professor with the School of Computer Science, NCBA&E. For the past eight years, he has been teaching graduate and undergraduate students in computer science and engineering. He has published about 48 research articles in international journals and reputed international conferences. His primary research interests include cloud computing, the IoT, intelligent agents, image processing, and cognitive machines.

MUHAMMAD ADNAN KHAN received the B.S. and M.Phil. degrees from the International Islamic University, Islamabad, Pakistan, and the Ph.D. degree from Isra University, Pakistan, in 2016. He is currently working as a Research Professor with the Pattern Recognition and Machine Learning Laboratory, Department of Software Engineering, Gachon University, South Korea, and an Associate Professor with the Faculty of Computing, Riphah School of Computing and Innovation, Riphah International University at Lahore, Pakistan. Before joining the Riphah International University, he has worked in various academic and industrial roles in Pakistan. For the past 12 years, he has been teaching graduate and undergraduate students in computer science and engineering. He is also guiding five Ph.D. scholars and six M.Phil. scholars. He has published more than 180 research articles with cumulative JCR-IF more than 220 in international journals and reputed international conferences. His primary research interests include machine learning, MUD, image processing and medical diagnosis, and channel estimation in multi-carrier communication systems using soft computing. He received the Scholarship Award from the Punjab Information Technology Board, Government of Punjab, Pakistan, for his B.S. and M.Phil. degrees and the Scholarship Award from the Higher Education Commission, Islamabad, Pakistan, for his Ph.D. degree.

TANWEER SOHAIL received the M.Sc. degree (Hons.) in mathematics from Government College University Lahore and the Ph.D. degree from the University of Science and Technology of China, China. He is currently working as an Assistant Professor with the Department of Mathematics, University of Jhang, Jhang. He has seven years of teaching experience at the undergraduate and graduate levels. He has supervised many students at the undergraduate level. He has also worked with the University of Sargodha at Gujranwala, as the Director of QEC. His research interests include algebraic topology, algebra, knot theory, soft computing, and probability theory and its applications. He received the Scholarship from the Chinese Academy of Science and The World Academy of Science for his Ph.D. degree.

DILDAR HUSSAIN received the B.S. degree in computer science from the Kohat University of Science and Technology, Pakistan, in 2010, and the Ph.D. degree in biomedical engineering from Kyung Hee University, South Korea, in 2019. From 2013 to 2019, he worked with YOZMA BMTech Company Ltd., South Korea (develop diagnostic imaging equipment instruments, such as DXA, Chats X-rays, and Ultrasonic), as a Research and Development Engineer. He is currently working as a Postdoctoral Research Fellow with the School of Computational Science, Korea Institute for Advanced Study (KIAS), which is a subordinate institute of KAIST, South Korea. His research interests include bioinformatics, medical imaging, medical image analysis, computer vision, biomedical natural image processing, artificial intelligence, machine learning, deep learning, and mineral and nutritional study.