ISSN 2007-9737

Emotion-Aware Explainable Artificial Intelligence for Personality, Emotion, and Mood Simulation

David Valle-Cruz, Asdrúbal López-Chau, Marco Antonio Ramos-Corchado, Vianney Muñoz-Jiménez

Universidad Autónoma del Estado de México, Mexico

davacr@uaemex.mx, alchau@uaemex.mx, maramosc@uaemex.mx, vmunozj@uaemex.mx

Abstract. This paper is twofold. First, the authors analysed the state of the art on the explainable artificial intelligence (XAI) applied to affective computing, emotion analysis, and sentiment analysis. Second, based on Kaptein's approach and using the personality, emotion, and mood model proposed by Egges, the authors tested an Emotion-aware XAI framework to determine the affective state of predefined-behaviour autonomous entities in uncertainty contexts. The affective state analysis of eight selected formal use cases was divided into four stages: personality settings, event evaluation, emotions and moods generation, and results classification. Findings suggest that the affective state depends on the type of predominant personality and the influence of the events in the environment.

Keywords. Explainable artificial intelligence, emotion analysis, sentiment analysis, personality, emotion, mood.

1 Introduction

The artificial intelligence (AI) field is booming, experts predict that by 2050 AI will be able to realise any intellectual task a human can perform [16]. AI and data-focused machine learning (ML) have been able to achieve exceptional performance on important and difficult learning problems [19]. The applications of AI techniques have been used in "high-level" domains such as healthcare, the criminal justice system, finance, security, and military decision-making [35], providing the possibility of exploiting Big Data.

People connect every day simultaneously in Big Data, sharing emotions, sentiments, and opinions. Nowadays, everyone can express on Internet applications their sentiments from joy to anger, and from depression to surprise [4]. In this regard, emotion and sentiment analyses combine techniques and tools from neuroscience to linguistics and computer science. For this reason, emotion analysis and sentiment analysis have been of interest to different researchers in the AI field [1, 6, 24]. Affective computing enables the automatic extraction of emotions and sentiments in data.

For retailers and marketing analysts, such methods can support the understanding of customers' attitudes towards brands, especially to handle crises that cause behavioural changes in customers, and for the personality profiles understanding for recruitment [5].

AI combined with neuroscience allows us to understand and simulate human beings' behaviour, for instance, some brain capabilities, personality, emotions, sentiments, and mood. Nevertheless, the results obtained in some affective algorithms represent a black box for simulation, sentiment analysis, and decision making.

For this reason, the aim of this paper is twofold: first, the authors analyse the state of the art on the Explainable Artificial Intelligence (XAI) applied to affective computing, emotion analysis, and sentiment analysis. Second, the authors implement an Emotion-aware XAI framework to determine the affective state of predefined-behaviour autonomous entities in contexts of uncertainty. The structure of the article is as follows:

The second section shows the characteristics of the studies related to XAI applied to affective computing, emotion analysis, and sentiment

ISSN 2007-9737

analysis. The third section presents the state of the art of XAI applied to affective computing, emotion analysis, and sentiment analysis. The fourth section describes a framework, based on personality, emotion, and mood, for modelling individuals' affective states based on autonomous entities towards an Emotion-aware Explainable Artificial Intelligence (EXAI). The fifth section presents the implementation of the affective framework of EXAI based on the analysis of eight selected formal use cases of autonomous entities. The paper ends up with the conclusions and future directions to continue with the study of EXAI.

2 XAI Applied to Affective Computing, Emotion Analysis and Sentiment Analysis

This section aims to describe the characteristics of XAI studies applied to affective computing, emotion analysis, and sentiment analysis with the existing papers found on the Web of Science, Scopus, and Google Scholar.

The authors developed a systematic literature review related to XAI. Most studies related to XAI have focused on analysing machine learning (ML) techniques applied to medicine and law and to foster the humans in the loop idea towards a more responsible, accountable, transparent, and inclusive AI [11]. In the area of affective computing, emotion analysis, and sentiment analysis, there is a gap, since XAI is an emerging area. To narrow down the document search, the authors applied a query for each repository as shown in Table 1.

For the selection of the papers, the authors systematically review the query results applied in Web of Science, Scopus, and Google Scholar. First, the authors discarded non-related and repeated documents. Reports, Web sites, and theses were also omitted to obtain selected literature for the state of the art. Finally, the systematic search resulted in 16 papers.

XAI applied to affective computing, emotion analysis, and sentiment analysis, is a recent topic on the research agenda of academics throughout the world. There is an existing gap in the state of the art. The first document dates back to 2017, and there was a research boom in 2019. However, the number of studies resulting after 2019 has decreased.

Table 1. Queries applied in Web of Science, Scopus, and Google Scholar

Database	Query
Web of Science	TS("Explainable Artificial Intelligence" AND ("Sentiment Analysis" OR "Emotion Analysis" OR "Affective computing"))
Scopus	TITLE-ABS-KEY("Explainable Artificial Intelligence" AND ("Sentiment Analysis" OR "Emotion Analysis" OR "Affective computing"))
Google Scholar	intitle("Explainable Artificial Intelligence" AND ("Sentiment Analysis" OR "Emotion Analysis" OR "Affective computing"))

Most of the state-of-the-art studies have used deep learning techniques such as Deep Neural Networks, Recurrent Neural Networks, and Convolutional Neural Networks [17, 18, 20, 22, 37]. Another proportion of the studies have focused on the study of Natural Language Processing and Intelligent Agents [4, 19]. In addition, the authors found one study related to distributed linguistic representation [38], another related to pattern recognition [29], and one more paper that deals with prototype sequence networks [22]. Some documents are preprints stored on arXiv.org, for this reason, several of the studies are ongoing research.

Most state-of-the-art studies have focused on analysing the application of XAI in sentiment analysis [5, 31]. While other applications that have been studied are related to affective computing, facial expressions, and emotions [12, 17, 23, 29, 37]. Other areas of application that have been analysed are social interaction, pervasive computing, decision making, and bias detection [15].

The existing concern in the state-of-the-art papers is related to the generation of models and systems that have explicability and interpretability in data analysis, learning, deicison-making, emotions and sentiments analysis, as well as in explain prediction methods and simulations.

The black box is an issue that generates a lack of understanding into the generation of scenarios and decision making, algorithmic biases, and in the confidence in artificial-intelligence-based systems.

3 The State of the Art of XAI applied to Affective Computing, Emotion Analysis and Sentiment Analysis

The state of the art of XAI applied to affective computing, emotion analysis, and sentiment analysis consists of 16 papers. Kaptein and colleagues wrote a seminal paper, identifying three ways in which emotions can be used to enhance XAI for cognitive agents, arguing that emotions have a strong potential to enhance XAI [14].

In 2018, Weitz et al., investigate the XAI method Layer-wise Relevance Propagation (LRP) and apply it to understand how a deep learning network distinguishes facial expressions of pain from the facial expression of emotions.

Their findings represent an initial step into the research of making decisions of black-box systems comprehensible for humans [37].

Mathews demonstrates that the explainability of ML models is vital and highlights fundamental directions for research, finding that an explainability framework furnishes a useful tool for identifying both fine-grained analyses at the feature level for malware and tumour data classification and as a text dataset introspection across documents and words that are important to a classifier’s decision [19].

Henriques and colleagues, contribute with a smartphone sensing-based system able to predict emotional valence states, reinforcing smartphone sensing contribution as a tool for a continuous, passive, and personalised health check, such as emotional disturbances, in spatial and temporal context [12].

Lin et al., propose a deep learning model to process multimodal-multisensory bio-signals for affect recognition, showing significant improvement compared to the state of the art [17].

Atzmueller outlines first perspectives on designing computational methods taking into account cognitive, human–machine, and computational requirements in an integrated way, outlaying and discussing important concepts of transparent, interpretable, and explainable analytics social interactions [2].

Ming et al., using ProSeNet (Prototype Sequence Network), demonstrate that involvements of domain users can help obtain more interpretable models with concise prototypes while retaining similar accuracy [22].

Schuller sustains that automatic recognition of emotion is essentially a pattern recognition problem and that such problems are usually solved by some sort of (statistical) ML. This author argue that technical solutions towards measurement of uncertainty in automatic emotion recognition exist but need to be extended to respect a range of so far underrepresented sources of uncertainty [29].

Again, Weitz and colleagues applied XAI to explain how a deep neural network distinguishes facial expressions of pain from facial expressions of emotions such as happiness and disgust [36].

Silveira et al., propose a framework to verify if predictions produced by a trained Aspect-based model can be used to explain Document-level Sentiment classifications, by calculating an agreement metric between the two models [31].

In 2020, Bodría and colleagues concluded that many XAI techniques can be applied to the field of Natural Language Processing (NLP) to understand better the sentiment classification process (and other NLP tasks in general, however, it is an area in its infancy and where there is much room for improvement [4].

Kumar and colleagues formalise the requirements of trustworthy AI systems through an ethics perspective, focusing on the aspects that can be integrated into the design and development of AI systems [15].

Marzban and Crick train a convolutional model on textual data and analyse the global logic of the model by studying its filter values, identifying the most important words in corpus removing 95% [18].

Their results suggest that trained on just the 5% most important words can achieve the same performance as the original model while reducing training time by more than half, improving by finding blind spots and biases.

Wu and colleagues present a review of the hot topics of distributed linguistic representations in decision making, including the taxonomy of distributed linguistic representations, the key elements in distributed linguistic decision making, and some challenges and future research directions from the perspective of data science and XAI [38].

Cirqueira et al., develop an Explainable Sentiment Analysis (XSA) application for Twitter data and propose research propositions focused on evaluating such application in a hypothetical crisis management scenario [5].

Finally, Mikołajczyk and colleagues argue that bias in data is defined as any trend or deviation from the truth in data collection that can lead to false conclusions and might cause misinterpretation not only for highly data-dependable deep learning models but also for human experts, which makes identifying and avoiding bias in the research a long-standing topic in general. However, in practice, it is impossible to gather all possible cases from the whole population [21].

Our analysis shows that there is a lack of studies in the state of the art of XAI application in affective computing, emotion analysis, and sentiment analysis. For this reason, the following section presents an EXAI framework to determine the affective state of autonomous entities.

4 Revisiting Personality and Emotion Models towards Emotion-aware XAI

This section aims to continue the approach of Kaptein et al [14], presenting a framework based on personality, emotion, and mood to explain autonomous entities behaviour through EXAI.

4.1 Personality Models

The study of emotions and sentiments can be approached from two major and complementary perspectives [8, 27, 28, 30]. First, from neuroscience, analysing personality and emotion models that explain the behaviour of human beings. Second, from AI, specifically from the areas of affective computing, emotion analysis, and sentiment analysis, in which formal and mathematical models map into personality, emotions, and sentiments.

In "The Expression of Emotions," Darwin suggests that many human emotions are "vestiges," as they once had some function in our evolutionary past [7]. Emotions can be understood as switches in the treatment of information and can be classified according to the architectural location of the information that triggers the mechanism and the resulting reorganization of behaviour. The interruptions generate emotions on three levels, according to the perceived event, for example, if the processing is interrupted by an alarm mechanism, a reactive process is generated manifesting itself in primary emotions.

On the other hand, if there is deliberative processing of information, secondary emotions are manifested, and eventually, with the passage of time and immersion in the environment, tertiary emotions are gradually triggered [32]. Interruptions are clearly adaptative and allow responding to urgent needs in real-time as required by the environment and personality.

Personality is the combination of an individual's behaviour, emotion, motivation, and the characteristics of an individual thought pattern [20]. In the field of personality psychology, one perspective states that personality consists of three dimensions or super factors [10]. Regarding this idea, Eysenck proposes three dimensions of personality: extroversion, neuroticism, and psychoticism, which are related to the motors of behaviour: reproduction, conservation, and self-defence respectively. Besides, Eysenck added a fourth factor: intelligence [10]. Barrick and Mount propose a five-dimensional model (Big Five) that includes extroversion, anxiety (neuroticism), self-control or consciousness, hostility/affability, and intellect [3]. Another five-dimensional model of personality is the OCEAN, conformed by openness, conscientiousness, extraversion, agreeableness, and neuroticism [1].

In 1930 the MMPI model, by its acronym in English Minnesota Multiphasic Personality Inventory model, was created. MMPI was designed to find personality profiles, and it has

Table 2. Primary scales of MMPI-2 RF model

Scales	Variables	Description
Hypochondriasis	Hs	Person's perception of their health or illnesses
Depression	D	Person's level of discouragement or motivation.
Hysteria	Hy	Person's degree of emotionality or excitement.
Psychopathic deviation	Pd	Person's need for control or rebellion against control
Masculinity/ Femininity	MF	The stereotype of a person by indicating how to compare men and women
Paranoia	Pa	Degree of disturbance of looking at an idea in or out of reality
Psychasthenia	Pt	Someone that lives with or without unresolved problems
Schizophrenia	Sc	Person's ability to have original or unique thoughts and whether they can get out of it
Hypomania	Ma	Person's psychic energy
Social Introversion	Si	The pleasure of being surrounded by more people

been one of the most used models to determine personality profiles.

The latest version of this model is the MMPI-2-RF (restructured), made in 2003 by Tellegen and Kaemmer [34] having three validation scales and ten basic scales to map a personality profile.

This model also contains additional scales that allow for the generation of better personality profiles, making it the best personality evaluator.

Despite being a rather complicated test at the time of its evaluation and interpretation, it is also the most valid and reliable. Table 2 shows the ten basic scales of the MMPI-2 RF model.

4.2 Emotional Frameworks

Several psychological models allow the representation of emotions, such as the double dimensional representation of the affective state, in which emotions are shown simply, this model describes emotions universally and unambiguously [33]. Emotions are classified from positive to negative and vary depending on levels of arousal. Roseman proposes a cognitive model that consists of categorizing the appreciations that people make, as positive or negative, towards the events that cause emotions.

This model does not respond to complex situations that are classified as certain, uncertain, and unexpected [26].

Ortony, Clore, and Collins grouped emotions considering that communication is an essential part of the same emotion, if a person is unable to communicate emotions, he/her is considered to lose their empathy. This model (OCC) specifies 22 emotional categories, including consequences of events, actions, and reactions to objects [25]. Finally, Ekman proposes six basic emotions: joy, fear, sadness, disgust, anger, and surprise, which are widely used, with other similar models, in affective computing, emotion analysis, sentiment analysis, and lexicons [9].

4.3 Personality, Emotion, and Mood Framework

AI and neuroscience have meant that personality and emotion models can be adapted to generate an approach to human behaviour in controlled environments or unique cases [23]. The empirical results of these approaches show that the inclusion of an emotional influence in the reasoning model helps to explain and better understand the behaviours observed in reality, allowing the generation of emotionally oriented architectures [13]. Egges, Kshirsagar, and Magnenat-Thalmann [8] proposed a model of personality and emotion (PE model), integrating the OCEAN and the OCC models. Egges et al. developed a mathematical model that allows mapping the personality and emotional states of the individual ***I*** in time $\boldsymbol{t}$ to real values, see Eq. (1):

$$I_t = (p, e_t), \tag{1}$$

where:

$\boldsymbol{I_t}$: entity that represents the individual in time $\boldsymbol{t}$.
$\boldsymbol{p}$: personality vector of the individual (***n**-dimensions*), initialized to $\mathbf{0}$ in $\boldsymbol{t} = 0$.
$\boldsymbol{e_t}$: vector that contains all the emotional dimensions that vary over time (***m**-dimensions*).
In such a way that Eq. (2) is defined as:

$$p^T = (\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_i)\ ,\ i \in [1, n] : \alpha_i \in [0,1]\ , \tag{2}$$

where α_i is the i-th value for the personality vector and Eq. (3) as:

$$\boldsymbol{e}_t^T = \begin{cases} (\beta_1, \beta_2, \beta_3, \cdots, \beta_j)\ , j \in [1, m] : \beta_j \in [0,1] & if\ \boldsymbol{t} > 0, \\ \mathbf{0} & if\ \boldsymbol{t} = 0, \end{cases} \tag{3}$$

where β_j is the j-th value for the emotion vector within the model a history of all emotional states is stored, which is a vector of all $\boldsymbol{e}_i$, see Eq. (4):

$$\omega_t = (e_0, e_1, e_2, \cdots e_t), \tag{4}$$

this array is called the emotional history ($\boldsymbol{\omega}_t$).

To determine the initial influence of the personality on the emotions is defined the matrix $\boldsymbol{P_0}$ (of $\boldsymbol{m} \times \boldsymbol{n}$ dimensions), in which each value indicates how each personality factor influences each emotion, $\boldsymbol{m}$ is the number of emotions, and $\boldsymbol{n}$ is the number of samples. The product of the matrix $\boldsymbol{P_0}$ with the personality vector $\boldsymbol{p}$, results in the vector $\boldsymbol{u}$. It contains the result of the personality influence on each emotion, generated in time $\boldsymbol{t}$, see Eq. (5):

$$u = P_0 \cdot p = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{bmatrix}. \tag{5}$$

Based on the vector $\boldsymbol{u}$ and with the aim to normalize the data obtained, the matrix $\boldsymbol{P}$ is defined, which contains all the values of $\boldsymbol{u}$ in its diagonal, see Eq. (6):

$$P = \begin{bmatrix} \varepsilon_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \epsilon_n \end{bmatrix}. \tag{6}$$

In this manner the product of this matrix by the influence of emotion ($\boldsymbol{a}$) results in the change of the emotional state, see Eq. (7):

$$\Psi_e(p, \omega_t, a) = P \cdot a\ . \tag{7}$$

Decay of the emotional state is a decrease in the emotional threshold, as a result of emotional deterioration over time, represented by a small change in an emotional state, as shown Eq. (8):

$$\Omega_e(p, \omega_t) = \begin{bmatrix} -C_e \\ \vdots \\ -C_e \end{bmatrix}. \tag{8}$$

Normally the decay (C_e) is a small value close to 0.01 [8, 23].

Finally, the emotional state ($\boldsymbol{e}_{t+1}$) depends recursively on the former emotional state ($\boldsymbol{e}_t$), the personality (p), of the emotional history (ωt) and of the emotional influence (a) [1, 6, 24], see Eq. (9):

$$e_{t+1} = e_t + \Psi_e(p, \omega_t, a) + \Omega_e(p, \omega_t)\ , \tag{9}$$

where $\boldsymbol{\Psi_e(p, \omega_t, a)}$ is the update of emotional state, $\boldsymbol{\Omega_e(p, \omega_t)}$ is the internal change (decay of the emotional state), a is a vector that contains the values that indicate how each personality dimension affects the emotions.

Adding mood to the model ($\boldsymbol{m}_t$), the emotional state ($\Psi_e(\boldsymbol{p}, \boldsymbol{\omega}_t, \boldsymbol{a})$) changes, modelling more realistic behaviours [1, 6, 24]. Similar to the change in emotional state, the mood can be updated over time depending on the event in which the individual is immersed. The updating of the mood is very similar to the emotional state update and is defined by the Eq. (10):

$$m_{t+1} = m_t + \Psi_m(p, \omega_t, \sigma_t, a) + \Omega_e(p, \omega_t, \sigma_t)\ , \tag{10}$$

ISSN 2007-9737

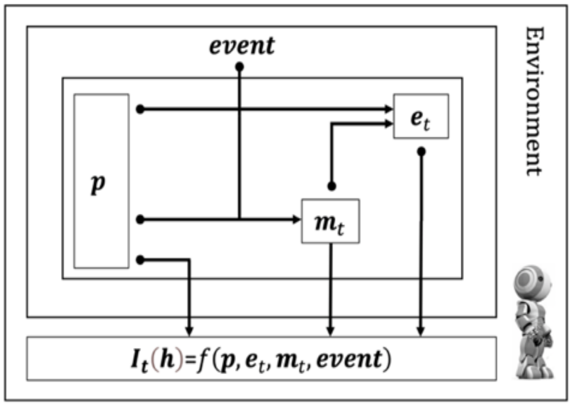


Fig. 1. Affective state update of an individual in time t

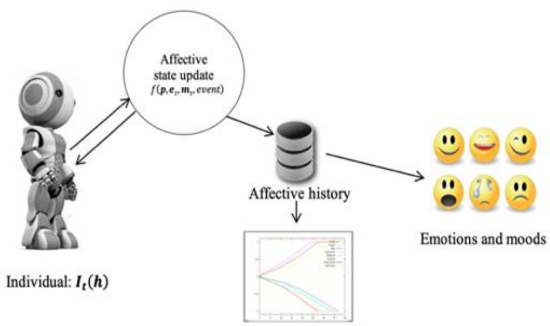


Fig. 2. Affective state framework

where $\boldsymbol{\sigma}_t$ is an array that contains the history of all mood states.

An emotion is an alteration by a shock or impulse in the brain caused by impressions of senses, ideas, or memories.

In this case, the personality and the event are important factors that will generate diversity in the emotional states for each profile, resulting in sentiments of anger, fear, disgust, surprise, joy, sadness, or some other kinds of emotions.

Thus, a generalisation of the modelling of the emotional state at time $\boldsymbol{t}$, is given by Eq. (11):

$$e_t = \begin{cases} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_r \end{bmatrix}, j \in [1, r] : \beta_j \in [-1, 1] & if\ t > 0, \\ 0 & if\ t = 0\,. \end{cases} \quad (11)$$

For the Ekman model [9], with six emotions, the emotional state vector have six emotions and is made up of scales of anger, fear, disgust, surprise, joy, sadness. In a similar way, the vector of mood in time $\boldsymbol{t}$ is generalized, which allows the modeling of the emotional state, since it can be noticed that depending on the mood a person has, the degree or the way of perceiving the environment changes, as well as the magnitude of individuals' emotional state, since individuals' emotional state is disturbed [1, 6, 24]. In this way, the mood vector is represented by Eq. (12):

$$m_t = \begin{cases} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_k \end{bmatrix}, \quad l \in [1, k] : \gamma_l \in [-1, 1] & if\ t > 0, \\ 0 & if\ t = 0\,, \end{cases} \quad (12)$$

where the values that take the dimensions of mood vary in the range of -1 to 1.

4.4 Affective State of an Individual

Based on Egges' work, the authors develop a model to describe how the affective state of an individual evolves in time as a result of external stimuli. The affective state of an individual at time $\boldsymbol{t}$ ($\boldsymbol{I_t(h)}$) is initially determined by the influence of personality and evaluation of the event, considering autonomous entities. These two factors affect emotional levels ($\boldsymbol{e}_t$), by exposing individuals to a similar event in the environment ($\boldsymbol{event}$), personality ($\boldsymbol{p}$) influences emotional ($\boldsymbol{e}_t$) and mood ($\boldsymbol{m}_t$) levels. Simultaneously, mood ($\boldsymbol{m}_t$) affects the emotional levels ($\boldsymbol{e}_t$) of the individual generating an affective state at time $\boldsymbol{t}$ ($\boldsymbol{I_t(h)}$). Recursively, the affective state at time $\boldsymbol{t+1}$ ($\boldsymbol{I_{t+1}(h)}$) depends on the previous affective state ($\boldsymbol{I_t(h)}$) and all the factors that modify it: the event, the personality, the emotions, and the mood at time $\boldsymbol{t+1}$ (see Figure 1).

Formally, and based on the framework designed by Egges and colleagues [8, 23] and based on the personality ($\boldsymbol{p}$), emotional ($\boldsymbol{e}_t$) and mood ($\boldsymbol{m}_t$) scales, the affective state due to an event in the environment ($\boldsymbol{event}$) of the $h-th$ individual ($\boldsymbol{I}$) in time $\boldsymbol{t}$ can be defined as a function of the following elements, see Eq. (13):

Table 3. Personality settings used for experimentation

Case	Personality	Features
1	Extroversion and high femininity	Tendency to relate to others and openly show feelings. High physical, psychic, and moral expression considered characteristic of a woman.
2	Medium hypomania	Behavior characterized by an expansive, hyperactive, and/or irritable mood.
3	High femininity and medium psychasthenia	High physical, psychic, and moral expression considered characteristic of a woman. Personality characterized by phobias, obsessions, anguish, feeling of strangeness before the world or oneself and intellectual and social inhibition.
4	Medium hypochondriac and high extrovert	Concern about having an illness. Tendency to relate to others and openly show feelings.
5	High depressive personality and medium hypochondriac	Feelings of constant sadness and a loss of interest in different activities. Concern about having an illness.
6	High hysterical and psychasthenia	Frequent psychic changes and emotional alterations. Personality characterized by phobias, obsessions, anguish, feeling of strangeness before the world or oneself, and intellectual and social inhibition.
7	High masculinity, extroversion, and hypomania personality	High physical, psychic, and moral expression is considered characteristic of a man. Tendency to relate to others and openly show feelings. Behavior characterized by an expansive, hyperactive, and/or irritable mood.
8	High hysterical and masculine personality	Frequent psychic changes and emotional alterations. High physical, psychic, and moral expression is considered characteristic of a man.

$$I_t(h) = f(p, e_t, m_t, event), \tag{13}$$

where:

$\boldsymbol{I_t(h)}$: The affective state of the $\boldsymbol{I_t(h)}$ individual in time $\boldsymbol{t}$.
f: Function that depends on $\boldsymbol{p}$, $\boldsymbol{e}_t$, $\boldsymbol{m}_t$, and the ***event*** for determining de affective state of the $h-th$ individual in time $\boldsymbol{t}$.
$\boldsymbol{p}$: Personality vector of the $h-th$ individual.
$\boldsymbol{e}_t$: Vector of the emotional state of the $h-th$ individual in time $\boldsymbol{t}$.
$\boldsymbol{m}_t$,: Vector of the mood of the $h-th$ individual in time $\boldsymbol{t}$.
event: The positive or negative evaluation of the event that took place in the environment.

5 Implementation and Formal use Cases

For the implementation of the affective framework, and the formal use cases the authors decided to use the basic scales of the MMPI-2 RF model for personality ($\boldsymbol{n} = 10$): see Table 3; the Ekman model for emotional states ($\boldsymbol{r} = 6$): anger, fear, disgust, surprise, joy and sadness; two dimensions of mood ($\boldsymbol{k} = 2$): good mood and bad mood applied for $\boldsymbol{t} = 100$ units of time.

For the tests, eight personality profiles based on MMPI-2-RF were designed to generate different affective states facing negative events or immerse in uncertain contexts (stress, pain, or crises) [5]. The goal was to predict how the emotional state of an individual change in time when he/she is stimulated with an external event or stimulus.

The way individuals perceived events in their environment was mapped to numerical values between -1 and 1, in this manner, negative values represent the result of undergoing events that were not very well accepted and positive values result from satisfactory events.

The resulting emotions were mapped to values between -1 to 1. Negative values indicate a low threshold of the emotion displayed by the individual or considerable emotional decay.

Positive values, close to 1, show that the emotion predominates. Values near 0 manifest emotional neutrality, these values correspond to individuals with stoic personalities. Figure 2 shows the affective state framework based on the personality, emotion, and mood model.

ISSN 2007-9737

5.1 Description of Analysed Cases

The experiments were conducted with eight personality profiles settings, shown in Table 3, using the proposed model to simulate the response or reaction of each individual subjected to the influence of a stressful situation [5]. For the sake of simplicity and clarity, the values of variables of each personality profile are shown in the results.

5.2 Experiments

Case 1. Extroversion and High Femininity

Femininity is high as well as extroversion, depending on the way the individual is expected to react, different values can be modified in order to find patterns of behaviour. As can be seen, the predominant personality scales are social introversion (highly positive) and femininity (Figure 3). The emotional evolution does not have a considerable alteration of the emotions and moods despite the succession of negative events.

Although significant changes are noticed in happiness and sadness, the alteration is easily regulated in time; this type of personality may be useful by some companies for job profiles in which the staff is subject to pressure and stress.

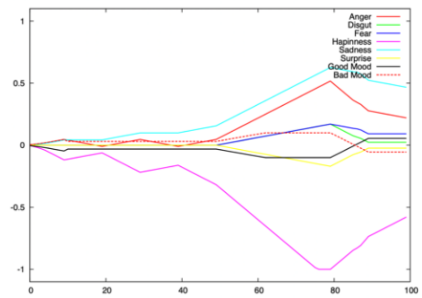
Case 2. Medium Hypomania

In this type of personality, significant changes can be noticed in the anger and disgust, altering them drastically. Good mood and happiness decline dramatically, while sadness and surprise are regulated in time. The succession of negative events is repetitive and increases all the negative emotions. The individual ends up in an affective state of sadness (Figure 4).

Case 3. High Femininity and Medium Psychasthenia

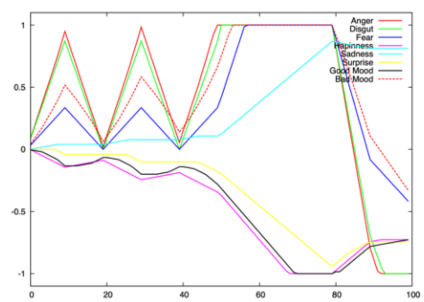
Initially, the individual shows significant changes in disgust and sadness.

The positive emotions decline and proliferate the negative ones. Another characteristic of this type of personality is how surprise evolves being the highest in value in the end. This individual is



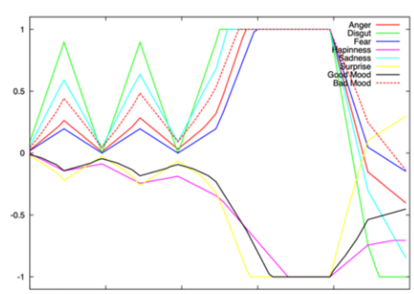
Variable	Value
Hs	0.1
D	0.1
Hy	0.3
Pd	0.1
MF	0.8
Pa	0.1
Pt	0.1
Sc	0.1
Ma	0.5
Si	0.8

Fig. 3. Affective state with an extroverted and high feminine personality



Variable	Value
Hs	0.1
D	0.2
Hy	0.1
Pd	0.1
MF	0.3
Pa	0.1
Pt	0.1
Sc	0.1
Ma	0.5
Si	0.4

Fig. 4. Affective state with medium hypomania personality



Variable	Value
Hs	0.3
D	0.4
Hy	0.4
Pd	0.4
MF	0.8
Pa	0.3
Pt	0.6
Sc	0.2
Ma	0.4
Si	0.2

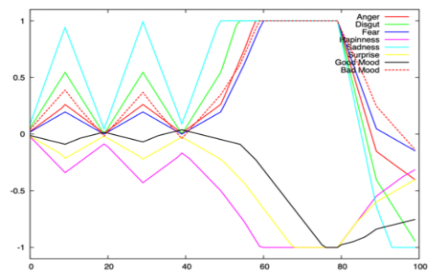
Fig. 5. Affective state with high femininity and medium psychasthenia personality

surprised after a succession of negative events (Figure 5).

Case 4. Medium Hypochondriac and High Extrovert

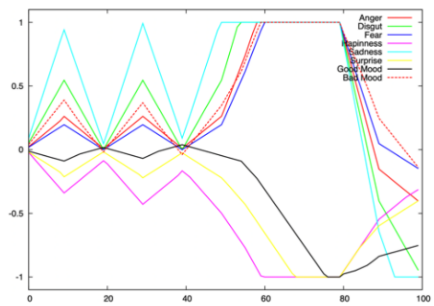
Due to the personality of the individual, there are significant changes in sadness, disgust, and happiness.

ISSN 2007-9737



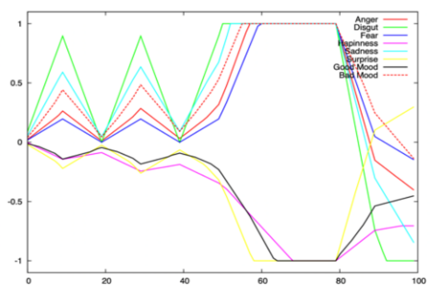
Variable	Value
Hs	0.9
D	0.6
Hy	0.5
Pd	0.7
MF	0.4
Pa	0.1
Pt	0.1
Sc	0.2
Ma	0.5
Si	0.8

Fig. 6. Affective state with medium hypochondriacal and high extrovert personality



Variable	Value
Hs	0.4
D	0.8
Hy	0.5
Pd	0.3
MF	0.3
Pa	0.2
Pt	0.5
Sc	0.2
Ma	0.5
Si	0.4

Fig. 7. Affective state with a high depressive and medium hypochondriac personality



Variable	Value
Hs	0.2
D	0.7
Hy	0.9
Pd	0.1
MF	0.8
Pa	0.2
Pt	0.8
Sc	0.3
Ma	0.4
Si	0.1

Fig. 8. Affective state with high hysterical and psychasthenic personality

Happiness, surprise, and good mood decline. Anger, fear, sadness, and bad mood predominate in the affective state. At the end of this formal use case analysis, fear and bad mood predominate (Figure 6).

Case 5. High Depressive Personality and Medium Hypochondriac

Figure 7 shows similar behaviour to the medium hypochondriacal and high extrovert personality, this may mean that hypochondriasis predominates over depression and extroversion, or that hypochondriasis and depression go hand in hand in personality.

Case 6. High Hysterical and Psychasthenia

Unlike the hypochondriacal individual where sadness is more noticeable than other emotions, the hysterical personality individual shows severe variations in disgust and sadness. Negative emotions predominate and positive emotions decline. In the end, the individual's affective state is characterized by surprise, fear, anger, and bad mood (Figure 8).

Case 7. High Masculinity, Extroversion, and Hypomania Personality

Figure 9 shows the personality of a hypomaniacal, extravert, but masculine individual. Fear and bad mood predominate. Happiness and good mood decline. In the end, the individual presents an affective state of anger and disgust, but with incremental happiness.

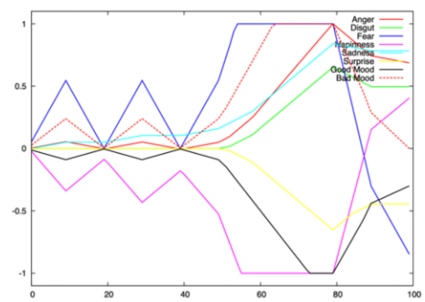
Case 8. High Hysterical and Masculine Personality

The masculinity of the individual defines some variations from the hysterical-psychasthenic individual. Anger and bad mood predominate. Happiness, surprise, and good mood decline. However, in the end, the individual is in an affective state governed by happiness (Figure 10).

6 Conclusions and Future Work

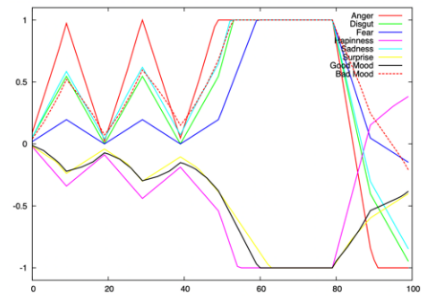
This paper analysed the state of the art on the XAI applied to affective computing, emotion analysis, and sentiment analysis, through existing literature in Web of Science, Scopus, and Google Scholar. Besides, the authors implemented an Emotion-aware Explainable AI framework to determine the affective state of individuals in contexts of uncertainty, and assessed the behaviour of individuals immersed in negative situations or events. According to each type of personality the affective state of individuals varies

ISSN 2007-9737



Variable	Value
Hs	0.5
D	0.3
Hy	0.2
Pd	0.3
MF	0.9
Pa	0.2
Pt	0.5
Sc	0.5
Ma	0.8
Si	0.8

Fig. 9. Affective state with high masculinity, extroversion, and hypomania personality



Variable	Value
Hs	0.2
D	0.5
Hy	0.8
Pd	0.5
MF	0.9
Pa	0.3
Pt	0.4
Sc	0.1
Ma	0.5
Si	0.5

Fig. 10. Affective state with high hysterical and masculine personality

in different ways. Our contribution is to propose a framework to analyse an individual's emotional state change (based on personality, emotion, and mood) when stimulated by an external event.

For the predominantly extroverted personality, the author discovered how the positive emotions and good mood are increasing, while the negative ones along with the bad mood are slowly decreasing. For the predominantly optimistic personality, the positive emotions and good mood increase, and a slight change is noticed in the evolution of these emotions, while the negative ones and the bad mood decrease rapidly. For the predominantly hypochondriacal personality, positive emotions and good moods decline rapidly, while negative emotions and bad moods increase as rapidly as positive emotions decline. This type of behaviour will be called asymptotic behaviour based on the evolution of the affective state.

For the predominantly hysterical personality, a behaviour similar to hypochondriacal behaviour is generated for its negative emotions, although sadness increases slowly as positive emotions gradually diminish. Despite these findings, some personality profiles could present similar affective states when exposed to the same critical event. Findings also showed significant changes in behaviour when the predominant personality dimensions vary, and small changes when the values are similar. In this manner, that emotions can be predicted based on the model proposed for the different personality patterns subjected to the changes of events in the environment: situations of fear, stress, and pressure.

The affective framework based on personality, emotion, and mood allows generating behavioural anticipation useful in the detection and explainability of what kind of actions a real human would carry out in certain situations. New lines of research may focus on the experimentation of cognitive agents for the human simulation behaviour with other cognitive architectures, in order to explain the emotion and sentiments simulation towards an improved Emotion-aware Explainable Artificial Intelligence.

References

1. **Aguirre, H., Corchado, F., Ruvalcaba, L., Rios, J., Thalmann, D. (2008).** Use of intelligent emotional agents in the animation of autonomous virtual creatures. Seventh Mexican International Conference on Artificial Intelligence, pp. 420–425. DOI: 10.1109/ MICAI.2008.15.
2. **Atzmueller, M. (2019).** Towards socio-technical design of explicative systems: transparent, Interpretable and Explainable Analytics and Its Perspectives in Social Interaction Contexts information. AfCAI 2019, pp. 1–8.
3. **Barrick, M.R., Mount, M.K. (1991).** The big five personality dimensions and job performance: a meta-analysis. Personnel psychology, Vol. 44, No. 1, pp. 1–26. DOI: 10.1111/j.1744-6570.1991.tb00688.x.
4. **Bodria, F., Panisson, A., Perotti, A., Piaggesi, S. (2020).** Explainability methods for natural language. Processing Applications to Sentiment Analysis. CEUR Workshop Proceedings, SEBD.
5. **Cattell, R.B. (2017).** The scientific analysis of personality. Routledge. DOI: 10.4324/97813 15134772.

6. **Cirqueira, D., Almeida, F., Cakir, G., Jacob, A., Lobato, F., Bezbradica, M., Helfert, M. (2020).** Explainable sentiment analysis application for social media crisis management in retail. 4th International Conference on Computer-Human Interaction Research and Applications, CHIRA, pp. 319–328. DOI: 10.5220/0010215303190328.
7. **Cruz, D.V., Rodríguez, E., Corchado, M.A.R., Marcial-Romero, J.R.M., Corchado, F.R. (2013).** Facial expressions based in emotions for virtual agents. CONIELECOMP, 23rd International Conference on Electronics, Communications and Computing, pp. 7–12. DOI:10.1109/CONIELECOMP.2013.6525748.
8. **Darwin, C. (1958).** The Expression of emotions. Harvard University Press.
9. **Egges, A., Kshirsagar, S., Magnenat-Thalmann, N. (2003).** A model for personality and emotion simulation. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 453–461. DOI: 10.1007/978-3-540-45224-9_63.
10. **Ekman, P. (1994).** Moods, emotions, and traits. The nature of emotion: Fundamental questions, pp. 56–58.
11. **Eysenck, H.J. (1990).** Biological dimensions of personality. Handbook of personality: Theory and research. The Guilford Press, pp. 244–276.
12. **Hall, P., Gill, N., Cox, B. (2020).** Responsible machine learning: Actionable strategies for mitigating risks & driving adoption. O'Reilly Media.
13. **Henriques, N.A.C., Coelho, H., Garcia-Marques, L. (2019).** SensAI+ expanse adaptation on human behaviour towards emotional valence prediction. arXiv preprint arXiv:1912.10084.
14. **Imbert, R., De Antonio, A. (2005).** Cognitiva: una arquitectura con base emocional para agentes software. CEDI I Congreso Español de Informática. WAF.
15. **Kaptein, F., Broekens, J., Hindriks, K., Neerincx, M. (2017).** The role of emotion in self-explanations by cognitive agents. Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), pp. 88–93. DOI: 10.1109/ACIIW.2017.8272595.
16. **Kumar, A., Braud, T., Tarkoma, S., Hui, P. (2020).** Trustworthy ai in the age of pervasive computing and big data. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 1–6. DOI: 10.1109/PerCom Workshops48775.2020.9156127.
17. **Kurzweil, R. (2005).** The singularity is near: When humans transcend biology. Penguin.
18. **Lin, J., Pan, S., Lee, C.S., Oviatt, S. (2019).** An explainable deep fusion network for affect recognition using physiological signals. Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2069–2072. DOI: 10.1145/3357384.3358160.
19. **Marzban, R., Crick, C.J. (2020).** Interpreting convolutional networks trained on textual data. arXiv preprint arXiv:2010.13585. DOI: 10.5220/0010205901960203.
20. **Mathews, S.M. (2019).** Explainable artificial intelligence applications in NLP, biomedical, and malware classification: A literature review. Intelligent Computing, CompCom 2019, Advances in Intelligent Systems and Computing, Vol. 998, pp. 1269–1292. DOI: 10.1007/978-3-030-22868-2_90.
21. **Mehta, Y., Majumder, N., Gelbukh, A., Cambria, E. (2019).** Recent trends in deep learning based personality detection. Artificial Intelligence Review, Vol. 53, pp. 2313-2339. DOI: doi.org/10.1007/s10462-019-09770-z.
22. **Mikołajczyk, A., Grochowski, M., Kwasigroch, A. (2021).** Towards explainable classifiers using the counterfactual approach-global explanations for discovering bias in data. Journal of Artificial Intelligence and Soft Computing Research, Vol. 11, No. 1, pp. 51–67. DOI: 10.2478_jaiscr-2021-0004.
23. **Ming, Y., Xu, P., Cheng, F., Qu, H., Ren, L. (2019).** ProtoSteer: Steering deep sequence model with prototypes. IEEE Transactions on Visualization and Computer Graphics, Vol. 26, No. 1, pp. 238–248. DOI: 10.1109/TVCG. 2019.2934267.
24. **Orozco, H., Ramos, F., Ramos, M., Thalmann, D. (2010).** A fuzzy model to update the affective state of virtual humans: an approach based on personality. International Conference on Cyberworlds, pp. 406–413. DOI: 10.1109/CW.2010.40.
25. **Orozco, H., Ramos, F., Ramos, M., Thalmann, D. (2011).** An action selection process to simulate the human behavior in virtual humans with real personality. The Visual Computer, Vol. 27, No. 4, pp. 275–285. DOI: 10.1007/s00371-011-0549-z.
26. **Ortony, A., Clore, G.L., Collins, A. (1988).** The cognitive structure of emotions. Cambridge University Press.
27. **Roseman, I.J., Spindel, M.S., Jose, P.E. (1990).** Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. Journal of personality

ISSN 2007-9737

and social psychology, Vol. 59, No. 5, pp. 899–915.

28. **Rosenwein, B.H., Cristiani, R. (2017).** What is the History of Emotions? John Wiley & Sons.
29. **Schuller, B. (2019).** Responding to uncertainty in emotion recognition. Journal of Information, Communication and Ethics in Society, Vol. 17, No. 3, pp. 299–303. DOI: 10.1108/JICES-07-2019-0080.
30. **Shvo, M., Buhmann, J., Kapadia, M. (2019).** Towards modeling the interplay of personality, motivation, emotion, and mood in social agents. Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pp. 2195–2197.
31. **De Souza, S.T., Uszkoreit, H., Ai, R. (2019).** Using aspect-based analysis for explainable sentiment predictions. CCF International Conference on Natural Language Processing and Chinese Computing, pp. 617–627. DOI: 10.1007/978-3-030-32236-6_56.
32. **Simon, H.A. (1967).** Motivational and emotional controls of cognition. Psychological review. Vol. 74, No. 1, pp. 29–39. DOI: 10.1037/h0024127.
33. **Sondhi, G., Sloane, A. (2006).** Emotions And digital photo sharing in a networked smart home. proceedings from the 5th conference on design and emotion.
34. **Tellegen, A., Ben-Porath, Y.S., McNulty, J.L., Arbisi, P.A., Graham, J.R., Kaemmer, B. (2003).** MMPI-2 restructured clinical (RC) scales: Development, validation, and interpretation. University of Minnesota Press.
35. **Valle-Cruz, D., Criado, J.I., Sandoval-Almazán, R., Ruvalcaba-Gomez, E.A. (2020).** Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. Government Information Quarterly, Vol. 37, No. 4. DOI: 10.1016/j.giq.2020.101509.36.
36. **Weitz, K., Hassan, T., Schmid, U., Garbas, J.U. (2019).** Deep-learned faces of pain and emotions: Elucidating the differences of facial expressions with the help of explainable AI methods. TM-Technisches Messen, Vol. 86, No. 7–8, pp. 404–412. DOI: 10.1515/teme-2019-0024.
37. **Weitz, K., Hassan, T., Schmid, U., Garbas, J. (2018).** Towards explaining deep learning networks to distinguish facial expressions of pain and emotions. Forum Bildverarbeitung, pp. 197–208.
38. **Wu, Y., Zhang, Z., Kou, G., Zhang, H., Chao, X., Li, C.C., Don, Y., Herrera, F. (2020).** Distributed linguistic representations in decision making: Taxonomy, key elements and applications, and challenges in data science and explainable artificial intelligence. Information Fusion, Vol. 65, pp. 165–178. DOI: 10.1016/j.inffus.2020.08.018.

Article received on 29/07/2021; accepted on 30/09/2021.
Corresponding author is Asdrúbal López-Chau.