Affective Relevance

Tuukka Ruotsalo[1,2], V. Javier Traver[3], Aleksandra Kawala-Sterniuk[4], Luis A. Leiva[5]

[1]University of Copenhagen, Denmark
[2]LUT University, Finland
[3]INIT, Universitat Jaume I, Spain
[4]Opole University of Technology, Poland
[5]University of Luxembourg, Luxembourg

***Abstract*—Modeling information relevance aims to construct a conceptual understanding of information significant for users' goals. Today, a myriad of relevance estimation methods are extensively used in various systems and services, mostly using behavioral signals such as dwell-time and clickthrough data and computational models of visual or textual correspondence to these behavioral signals. Consequently, these signals have become integral for personalizing social media, search engine results, recommender systems, and even supporting critical decision-making. However, behavioral signals can only be used to produce rough estimations of the actual underlying affective states that the users experience. Here, we provide an overview of recent alternative approaches for measuring and modeling more nuanced relevance based on physiological and neurophysiological sensing. Physiological and neurophysiological signals have the advantage of directly measuring users' affective responses to information and provide rich data that are not accessible via behavioral measurements. With these data, it is possible to account for users' affective experience and attentional correlates toward information.**

***Index Terms*—affective computing; physiological sensing; Brain-Computer Interfaces; wearables**

What if we could infer affective experiences toward information in addition to conventional relevance signals? Affective responses are directly accessible from users' physiology, e.g., via measurements of gaze, body movements, facial expressions, and brain activity. These signals could mitigate the reliance on behavioral signals, such as click-through data or speech, which are not always available.

I. INTRODUCTION

PREDICTING whether a piece of information is relevant to users is a cornerstone of personalized services. These range from optimizing the results of search engines, chatbot responses, and social media feed content, ubiquitous access of information in pervasive environments to media and product recommendations. Typically, information relevance is indirectly estimated from user behavior, such as how long a user spends browsing some content, when or what a user clicks or purchases, or how a user explicitly rates information. Despite the success of these behavioral signals in research and practical implementations, user behavior is only a proxy of the real underlying experiences of the users, all of which emerge from how users understand and emotionally react to the information they perceive. This may be in contrast to how users' digital behavior is associated with information.

First of all, even implicit cues that are measured as a side product of users' everyday activity are tied to behavior measured from explicit interaction with computing systems. However, emotional experiences are not always predictable based on behavior. For example, despite users spending more time investigating some content, it does not always imply they would find it relevant and prefer similar content in the future. Moreover, despite dwelling on content, users may experience the content as offending, frightening, or even outrageous and, therefore, might prefer to avoid such content in the future.

Here, we present a complementary view of relevance: *affective relevance*. It refers to the level of emotional significance or state that information holds in relation to an individual's task, topic, or goal. Emotional experiences are often quantified by the dimensional theory of emotions [10], with a general consensus among theories on two fundamental dimensions: arousal and valence. Arousal signifies the extent of autonomic activation induced by an event, spanning from a state of calmness (or low intensity) to one of excitement (or high intensity). In contrast, valence signifies the degree of pleasantness evoked by an event, spanning a spectrum from negative to positive. Emotions can then be quantified on their position on the valence-arousal coordinates.

Figure 1 illustrates an example of a user watching a basketball game. Consider a scenario where one team holds a significant lead, making it impossible for the opposing team to secure victory within the remaining duration. While users may perceive the same content as highly relevant as they are interested in the game (high arousal), their emotional responses could diverge. For instance, a basketball enthusiast supporting the leading team might interpret their team's success positively (positive valence) due to the imminent victory. Conversely, another individual supporting the opposing team might acknowledge the relevance of the content but have a negative emotional experience (negative valence) due to the impending defeat of their favored team. These affective states can be reflected in the users' facial expressions or brain recordings in response to perceiving the content. This example illustrates that relevance, or interest in information, may have a strong affective component that can be highly indicative of the user's experience toward information.

The mismatch between individual users' subjective affective and emotional experiences and their behavior cues is often called the '*affective gap*' [20]. Bridging the affective gap

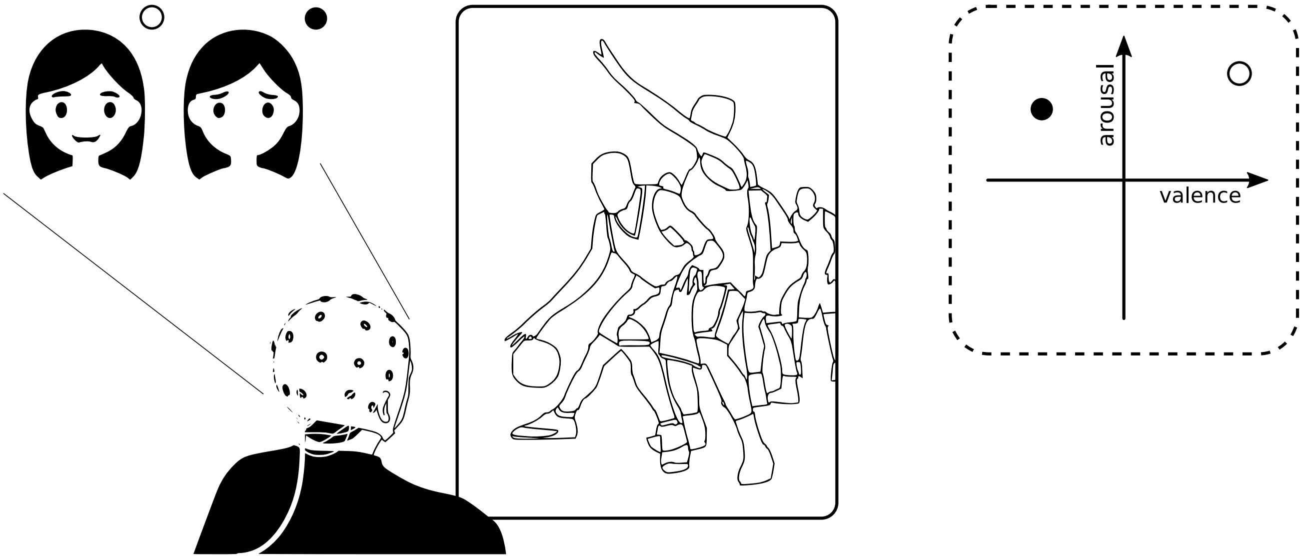

Fig. 1: An affective dimension of relevance enables computing systems to interpret users' emotional and affective responses, in addition to conventional measures of relevance or interest. In the scenario depicted here, a user is watching a basketball game. User signals, such as brain recordings or facial expressions, can be used to determine which affective reactions the content evokes in the user. For example, depending on whether a user is a fan of a winning team (positive valence marked with a white dot) or a fan of a losing team (negative valence marked with a dark dot), the same relevant and interesting (high arousal) content can evoke different affective reactions.

requires fundamentally new methodological approaches that can reveal and estimate fine-grained affective and cognitive features from user signals beyond simple measurements of user behavior.

Researchers have aimed to estimate affective features from users' physiological reactions, ranging from facial expressions to brain measurements. While various signals and methods have been studied, the affective gap has turned out to be challenging to address. As a result, affect recognition has primarily remained an isolated research task rather than a practical tool for extending the present notion of information relevance. However, recent developments in wearable sensor technology and machine learning models that can decode affective information in the presence of noise have brought us closer to understanding the affective dimension of relevance in realistic settings.

Nowadays, physiological sensors are making their way into consumer electronics. For example, smartwatches can record an electrocardiogram, and low-cost cameras can accurately detect facial expressions. This opens the door to novel ways of measuring user interest toward digital content. Pushing the boundaries of wearable computing and physiological sensing, we argue that computing systems are becoming able to detect users' affective and cognitive judgments toward information. At best, such technology can completely change how we interact with computers and computer-mediated services by allowing objective measurements of "human affect". Furthermore, the affect and even opinions of larger crowds of people that can be estimated implicitly may have broader societal benefits by revealing implicit attitudes and biases, and informing artificial intelligence systems that cope with these critical challenges.

In this article, we contribute critical analysis of the potential of affective information on user modeling, review the user signals and sensor technology for monitoring affective responses, highlight example applications on combining affective data with quantification of relevance, and discuss the ethics and implications of the technology for service providers, individuals, and society at large. These analyses can be critical for our understanding of how people interact with information and how affective dimensions of user engagement can be incorporated to reveal how our attention is allocated.

II. AFFECTIVE MONITORING FOR USER MODELING

In the last two decades, thanks to the pervasiveness of web browsers and mobile applications, researchers have focused on explicit behavioral signals, such as click-through logs, or search queries, and implicit signals extracted from explicit behavior, such as dwelling time on web pages or social media posts. These signals have the potential to uncover latent factors about the user, can be collected unobtrusively, at a large scale, and without having to instrument the user's working environment.

However, behavioral signals are limited to those that can be reliably collected, such as what users click, which contents users spend time on, or what users type, all dependent on eventual explicit interactions with computers. To this end, the essential information on how users perceive information and what affective responses the information they perceive evokes in them remains largely uncharted. Simply put, we

Glossary

Affect and emotions. The scientific community has no consensus on the definition of emotions or affect. Some researchers have supported an understanding that emotions are discrete, measurable, and physiologically distinct (see **discrete theories** below). In contrast, others have supported an idea that places each emotional response onto a more limited number of affective dimensions, typically valence and arousal (see **dimensional theories** below).

Discrete theories of emotions. The discrete theory of emotion considers that people share a basic set of emotions, an example being Ekman's categories: anger, disgust, fear, happiness, sadness, and surprise. There is a debate on what constitutes a *basic* emotion, and it is generally accepted that the context (such as the user's cultural background) plays a significant role.

Dimensional theories of emotions. The dimensional models of emotions represent those based on a few continuous dimensions. One popular 2D model includes *valence* (how positive, negative, or neutral an emotion is) and *arousal* (how strong or weak the emotional response is). Other 2D, 3D, and higher-dimensional models have been proposed.

Affective decoding and affective annotation. Affective *decoding* is the problem of estimating the emotion of individuals from their responses to some stimulus. Affective *annotation* is the problem of labeling contents using affective decoding.

Experimental data regimes. Predictive models for affective decoding and annotation are categorized into three main regimes. In the *participant-dependent* regime, individual models are trained (and tested) separately for each participant. In the *participant-independent* regime, a single model is learned with multiple participants, which may include subjects used in testing. In the *cross-participant* regime, data from test participants are not used during training, which is the hardest problem and aims at studying to what extent the models can generalize to new subjects. This represents an ideal future scenario for calibration-free "plug-and-play" systems.

Relevance. Relevance refers to the level of significance that some information has to a particular context, task, or goal.

Affective relevance. Affective relevance refers to the level of emotional significance that some information holds in relation to an individual's task, topic, or goal.

Behavioral signals. Behavioral signals are intentional, observable interactions with computing systems. They can be *explicit* interactions, but sometimes measured *implicitly* as side information of a primary activity. For example, a popular approach has been to record click-through data to monitor which links users follow, dwelling time to measure how long users spend on content, facial expressions and gestures that communicate emotional cues, or gaze patterns that indicate what users focus their attention on.

Neurophysiological signals. Neurophysiological signals allow the measurement of brain activity. Popular non-invasive neurophysiological signals are electroencephalogram (EEG) and functional near Infrared Spectroscopy (fNIRS). EEG measures the activity of synchronously firing populations of neurons with electrodes placed on the scalp. fNIRS is an optical imaging technique that uses near-infrared (NIR) light to detect changes in cerebral blood flow as a proxy for neural activation.

Peripheral physiology signals. Psychophysiological processes often directly relate to how the human body reacts to psychological states or external events. This is particularly noticeable with emotions. Popular approaches to measuring peripheral signals include, among others, electrodermal activity (EDA), heart rate variability (HRV), pupil dilation, and extraocular muscle movement (EMG).

can observe what content users interact with, but we cannot observe how they react and feel when interacting with such content. Thus, current technology to predict affective-level responses is based on behavioral probes that may be unreliable and thus may not provide accurate information about nuanced affective experiences.

Measuring attention at the neural and peripheral processes level would enable more accurate and new types of signals that can reveal more fine-grained affective and cognitive features of user attention. What if we could collect affective responses from users and reliably measure their reactions to content? What if we could reveal how people perceive the increasing amount of information available, and what if we could automatically interpret and detect reliable, positive, threatening, or fake content from the natural responses of people toward digital media? While this may, at first, sound like science fiction, we are not far away from deploying affective sensing technology for ordinary use. In sum, computers should be able to detect users' cognitive and affective experiences toward digital content instead of relying on sparse (and possibly unreliable) behavioral signals.

There exists a variety of technologies for physiological affect monitoring that can complement behavioral signals. In addition to their underlying operating principles, they differ in several important practical respects, such as reliability of the estimates, usability and acceptability, and affordability. Table I summarizes the present behavioral, peripheral, and neurophysiological signals used for recognizing affect, which are also discussed below.

A. Behavioral signals: face, gaze, and speech

Human behavior can be easily captured from natural human-computer interaction, such as clicks or cursor movements, and external sensors, such as eye-tracking equipment or video cameras. While conventional human-computer interaction signals, such as clicks, are less helpful in detecting affective information, the computer vision field has experienced significant progress in the last decade.

TABLE I: Summary of advantages and limitations of different signal sources for estimating affective relevance.

Signal sources and types

Signal source	**Type**	**Advantages**	**Limitations**
Click-through	Behavioral	• Explicit signal that is straightforward to capture from regular interactions	• Only available for information that is explicitly interacted • Does correlate with relevance, but not with nuanced affective states
Speech	Behavioral	• Explicit signal • Very natural and straightforward to capture	• Limited to verbal communication contexts • Limited recognition accuracy of affective states
Gaze	Behavioral	• Explicit signal for attention, but some affective information can be decoded • Available via eye-tracking equipment, and to some extent also via videocameras	• Only limited cues for affective information present • More accurate for attention and interest than affective states
Body gestures	Behavioral	• Explicit signal available via cameras • Unobtrusive and straightforward to capture	• Only available from contexts where users exhibit gestures • Allows recognition of only a limited set of affective states
Pupil diameter	Peripheral physiology	• Implicit signal for affective states • Available via eye-tracking equipment, and to some extent also via videocameras	• Accurate measurement requires eye-tracking equipment • Relatively noisy signal
Facial expressions	Peripheral physiology	• Implicit signal for affect, but it must be explicitly expressed • Availability via videocameras and ultrasound	• Not always available and reliably expressed • Susceptible to voluntary or involuntary modification
EDA, EMG, HRV	Peripheral physiology	• Implicit signal for measuring physiological responses • Availability via easy-to-wear sensors	• Prone to motion artifacts • A delayed response requires a longer duration to capture a meaningful sequence of signals.
EEG	Neurophysiology	• Implicit signal measuring electrical brain activity; hard to be faked • Available, but not accurate for all nuanced affective states	• Requires relatively costly wearable devices • Prone to motion and other artifacts
fNIRS	Neurophysiology	• Implicit signal measuring blood-oxygenation in the brain; hard to be faked • Less prone to movement artifacts	• Requires costly wearable devices • A delayed response requires a longer duration to capture a meaningful sequence of signals • Unequal accuracy at different valence-arousal combinations.

Among the various data types employed in automated emotion recognition, visual data stands out as the most versatile due to several compelling factors. Primarily, facial expressions and body gestures, constituting powerful nonverbal channels of communication, are important for human emotional expression.

Furthermore, in contrast to emerging data forms like physiological signals, the process of gathering visual data is significantly less intrusive. This implies that subjects are far more likely to engage in their routine tasks without disruption during the data collection process. Moreover, recent research has also employed ultrasound for sensing facial muscle movement without the requirement of video camera data [17].

It has been criticized, however, that such systems cannot directly detect emotions, but rather expressions that do not necessarily reflect the underlying true sentiments [19]. For example, detecting a frowning face is feasible with existing technology, but associating frowning reliably with an affective state or user sentiment may be more challenging. There is no exact mapping between observable expressions and possible emotions, and the expression can also be faked, intentionally or unintentionally.

In addition to analyzing face-based signals, gaze behavior can offer valuable insights into human emotional states. One essential aspect of eye-data analyses is the examination of gaze patterns, which refer to the specific directions and points of focus when observing. Gaze patterns analysis provides valuable information about visually salient information, the attention direction of individuals, or how the visual stimuli are processed.

For example, a prolonged fixation on a specific object or region may indicate interest or emotional involvement, while rapid and frequent changes in gaze may suggest uninteresting content, alertness, or cognitive processing.

Other relevant behavioral features, which can be extracted from eye-related signals, are fixations and saccades. Fixations refer to the brief periods during which the eyes remain relatively stable and focused on a specific point of interest. In contrast, saccades are rapid, involuntary eye movements between fixations when the eyes shift their focus from one point to another. Fixations can provide insights into cognitive processing, information gathering, and decision-making processes, while the saccades can reveal aspects such as visual exploration, attentional shifts, and response to stimuli.

Other signals measurable from eye-gaze, such as pupillometry or electrooculography, also hold significance as indicators of emotional experiences. While these signals are usually not considered to be behavioral, but rather peripheral physiology, they allow the measurement of information valuable for the detection of affective states. For example, users' emotions can be approximated by examining changes in their pupil size. An increase in pupil diameter is an indicator of positive valence. These signals are easy to capture and can aid in detecting affective relevance in realistic human-computer interaction, enabling the development of more effective emotion-aware systems and interfaces [5].

Speech is a versatile and essential human communication mode for expressing thoughts, emotions, and intentions. Speech perception involves decoding auditory signals to interpret the intended expressions. To this end, speech is crucial in conveying complex messages between computers and machines, and speech-based human-computer interaction has become widely accepted, especially in interacting with mobile devices. Due to its distant and hands-free operation, systems that process sounds can have generally high acceptance.

Although speech and voice recognition are already mature technologies, their application for emotion estimation is significantly more challenging. The difficulties may arise from the diversity of human cultures and languages, speaking styles, and particular sentences being utteranced. Furthermore, it is not uncommon for a single utterance to evoke multiple emotions, and ascribing distinct emotions to each segment of the utterance is typically challenging. Speech can also be of limited utility for detecting affective relevance, as speech is usually used only when sending direct messages with intentional communication and it is not available for information that users only perceive, but do not intentionally act upon. For example, a user might see an interesting social media posting, but they do not necessarily respond to and any speech interaction may not be associated with observing the posting. Moreover, as estimating emotions from speech can be difficult even for human subjects [15], making speech a challenging data source for emotion recognition. However, using speech in some well-contextualized settings, such as detecting affective relevance during conversations or social events, may facilitate the task.

Finally, body gesture recognition techniques can identify a reduced set of emotions and affective states via visual and ultrasound sensing data. The current methods utilizing image sequences and skeletal data often neglect to account for spatial connections and graphical structures explicitly. As a result, the ability to accurately interpret user expressions conveyed through physical movements is somewhat constrained and has allowed the classification of only a narrow set of emotional states. This reduced scope implies that it affects the accuracy of gesture signals for general affect decoding [4].

Research has also provided resources and evidence that affective decoding can be performed using data collected in-the-wild [7] and from realistic video recordings [12] without reliance on artificially curated data or wearable sensors. This marks a way for detecting affective relevance as it occurs as a part of our everyday information interaction.

B. *Neurophysiological signals*

In contrast to external observations, such as facial expressions or body language, neurophysiological signals offer a more direct and objective access to the internal emotional state of humans. These signals provide insights into the underlying neural processes and physiological responses associated with emotions. This advantage has sparked significant interest in research and application of Brain-Computer Interfaces (BCIs).

However, the analysis of brain signals is an extremely difficult task. These signals are noisy as they are prone to various (internal and external) artifacts, with much intra-subject and inter-subject variability, which significantly complicates the design of robust affective decoding systems.

Despite the above-mentioned difficulties, BCI-based technology can be one of the most reliable sources of true affective states, since they are less susceptible to voluntary or involuntary modification by the user. In the last few years, numerous machine learning approaches have been applied that hold the promise to yield high estimation performance. Currently, recording these signals still represents a costly and obtrusive procedure, although some more usable, cost-effective, and wearable systems have been developed over the years [13]. Two main technologies have been used: Electroencephalography (EEG) and functional Near-Infrared Spectroscopy (fNIRS).

The EEG technology is currently the most popular non-invasive brain imaging technology; it is very well studied, and relies on the electrical activity of the brain. This technique is older than fNIRS, and the EEG signals recorded from the surface of the scalp are noisier than the intracranial EEG recordings. The fNIRS technology, in contrast, is based on blood flow changes in the brain tissue. EEG and fNIRS differ in their spatial and temporal recording resolution: EEG has high temporal resolution but poor spatial resolution, and the opposite happens with fNIRS.

Other brain imaging methods, which exhibit either high spatial or temporal resolution, are functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG), but they are prohibitively costly, and their usage requires specific laboratory environments. This makes them less usable for scenarios that aim for affective decoding in everyday human-computer interaction.

In summary, neurophysiological signals applied in BCIs can offer direct and objective access to the internal emotional state of individuals. BCIs have wide-ranging applications, from healthcare to gaming and affective computing, enabling personalized and adaptive systems that can respond to users' emotional needs. The recent advancements in BCIs highlight the growing recognition of the significance of emotions in human-computer interaction and the potential for the design and development of innovative and impactful applications.

C. Peripheral physiology signals

A third important source for affective monitoring are peripheral physiological signals, which lie between the external observation of human behavior and the analysis of the brain.

These signals include, for example, heart rate variability (HRV), measured with electrical (electrocardiogram, ECG) or optical (photoplethysmography, PPG) sensors and electrodermal activity (EDA), measured through galvanic skin response (GSR) signals [6].

Heart rate is regulated by the two autonomous nervous system divisions, namely the sympathetic and parasympathetic components. An increased activity of the sympathetic component is typically characterized by elevated heart rate and decreased HRV, while increased parasympathetic activity is characterized by decreased heart rate and increased HRV. There is evidence that evaluating the momentary changes in HRV can provide strong cues about fluctuations in cognitive processes, but their application in affective computing scenarios remains to be explored.

ECG measures the electrical activity that arises from heart muscles during cardiac contractions and passes through the soft tissues to the superficial skin. The basic ECG pattern consists of a series of waves or deflections of electrical activity, among which the R wave is the largest one. It reflects the depolarization of the main mass of the ventricles. As such, HRV is measured based on the R-to-R (RR) intervals across the cardiac cycles. The heart rate can also be measured with a PPG sensor placed on the finger that detects changes in blood volume. PPG quantifies HRV based on the peak-to-peak (PP) intervals of the acquired signal, also known as inter-beat-intervals (IBI). Although PPG measures a hemodynamic signal rather than electrical activity, it provides traces similar to the ones obtained with ECG [14]. In addition, HRV is sometimes monitored with wearable devices (e.g., smartwatches) that are widely accepted by participants and are more suitable for naturalistic experiments that involve body movements.

EDA has been one of the most popular psychophysiological signals to acquire, not only because of the relative simplicity and the low-cost equipment needed, but also due to the fact that EDA can provide information about numerous mental constructs involving changes in sympathetic activity. For example, EDA is considered a pure arousal indicator [9] and has been used also to model relevance experience in information retrieval [1]. EDA refers to changes in skin conductivity due to sympathetic nervous system activity. The activation of the sympathetic branch of the nervous system stimulates the production of sweat in the eccrine glands located in the palms of the hands and soles of the feet, increasing skin conductivity in these areas.

When using EDA to obtain information about affective states, a fundamental aspect to consider is that, as mentioned above, EDA is a proxy for sympathetic activation, which is a component of multiple psychological processes. This makes it a useful signal for the prediction of various physiological states, but at the same time, it is particularly challenging to map EDA changes to specific affective states. Thus, depending on the state to be analyzed, the EDA signal may be especially useful when combined with other neurophysiological measures, allowing disambiguation of the meaning of the observed changes in these signals.

In fact, it is not uncommon to combine multiple biosignals as hybrid and multimodal systems for higher effectiveness, so that the biomedical data recorded from various sources and sensors can be put together to provide more reliable information. For example, speech and gaze can provide complementary emotional cues when monitoring players in video games.

III. APPLICATIONS OF AFFECTIVE RECOGNITION

Recognizing the emotional response to digital content has numerous applications, which can be divided into two broad categories depending on the potential main beneficiary, whether it is the primary user (e.g., the monitored subject of the affective technology) or a secondary user (i.e. those using the processed data and information). Certainly, sometimes the boundaries between both categories are blurred, so multiple implied stakeholders may benefit from affective recognition technologies.

A. Systems targeted at primary users

Art, digital games, and entertainment are domains where the monitored subject is the one who can benefit the most directly from affective computing technology. For example, audiovisual content such as music can be generated from users' predicted affective state. In turn, generated contents may induce new brain responses, so that stimuli and affective states relate in a closed loop. Ultimately, these approaches may enrich the affective experience by better encoding and exploiting the affective relevance of multimedia content.

In the video game industry, affective relevance has great potential not only to understand players' experience during gaming but also to modify the gameplay components according to the monitored player's affective state. Similarly, recommender systems can leverage the predicted user's affective state to better guide the recommendations. In addition, Human-Robot Interaction (HRI), and social robotics in particular, are exciting and relatively new research areas where affective relevance may play an important role: not only social robots can benefit from recognizing human emotions, but they can also act to express some emotions to which people, in turn, may react. One area of interest is the study of how the appearance and behavior of the robot can elicit emotions. For instance, recently, the robot's gender and level of anthropomorphism have been shown to play an important role [18]. Another area of research is using emotion expression as a persuasive tool, such as finding which type of robot-expressed emotion may favor people to adopt positive behaviors or habits. By building a good model of the user's expectations, a robot may be trained to produce actions that are more suitable and affectively relevant. As a result, HRI represents a rich ecosystem where affective relevance research can find novel practical applications as well as inspiration for new theoretical and methodological contributions. More generally, any user interface can be modified dynamically to reflect or change the user's experience of relevance and affect.

B. *Systems targeted at secondary users*

Medicine, education, social networks, and biometrics are only a few examples where representatives of various professional sectors (e.g., medical doctors, teachers, industry, or government employees) can obtain support from affective technology by mining collected data (e.g., from patients, trainees, customers, citizens). For example, automatic emotion recognition can assist physicians in understanding patients with difficulties in expressing emotions (e.g., due to motor impairments). On the other hand, virtual reality systems can be instrumental in helping with emotional diagnosis, assistance, or induction. In academic contexts, detecting reactions such as engagement or boredom can be useful to improve reading materials or lecture delivery methods. More generally, multimedia content can be automatically and implicitly tagged based on emotional responses from which its affective relevance can be modeled and subsequently used for improved interaction.

IV. Ethics and Privacy

The rise of large-scale user monitoring, as evidenced by monitoring millions of web users, has revealed the utility of low-fidelity data for inferring detailed information about individual users. On the other hand, it has already raised concerns about the exposure of user behavior for purposes not initially consented to by users [8]. When user monitoring technology capable of recognizing affective attributes becomes as common as behavioral tracking via personal computers and smartphones, affective information may become available to service providers. Consequently, ways in which facial expression tracking, physiological tracking, and other wearable hardware monitoring can be used unethically to reveal cognitive and affective user attributes may emerge. Recent work has shown that increasing awareness of the usage of data is making users more hesitant to use technology that can reveal detailed information about additional attributes of their physical and cognitive states [11].

For example, *subliminal probing* is a technique where a user is exposed to information, and their corresponding affective reaction can be recorded without their knowledge or consent [3]. This may reveal users' opinions without them even being aware of such monitoring – for example, by measuring responses to advertisements shown very rapidly. Physiological data recorded via wearables can also be used as a *biometric identifier* to pinpoint an otherwise anonymous user despite the recording of data not initially consented to for that purpose.

Moreover, affect and opinions of crowds of users toward stimulus information can be estimated implicitly, sometimes referred to as *brainsourcing* [2]. This may bring broader societal benefits by allowing the detection of harmful attitudes and biases. However, such inference may also be used in an unethical way for identifying the development of crowd opinions, and political stances, or even for designing interventions influencing larger crowds of people.

There are already regulative actions circumventing unethical use. For instance, the EU AI act[1] prevents AI systems for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data in the workplace and educational institutions. The AI Act's prohibition seems to focus on preventing potential misuse of emotion detection technology in sensitive areas, including workplaces, education, and marketing. However, this prohibition does not encompass all possible uses but targets specific scenarios where the risk of privacy infringement is deemed high.

V. The road ahead

Affective computing is becoming increasingly accessible to the general public on several fronts, from computer vision to smartwatches and comfortable wearable sensors. Fueled by an increasing and extensible real-time mobile connectivity, now it is possible to measure expressions and biosignals from the human body and brain. These signals carry rich information about the cognitive and affective states of the user and can be used to estimate affective relevance: whether users are interested in certain information and what emotional responses that information evokes. As such, the present technology is showing the way for getting wearable and physiological sensing technology out of the laboratory environments into everyday life, with an ever-increasing affordability and signal quality.

Detecting human affect may allow for improving many applications ranging from biomedical monitoring, more accurate search and recommendation systems [16], to detecting harmful, incorrect, or dubious information online. While physiological signals may be of a more ambiguous nature than those relying on behavioral data, they allow inference of more nuanced information about users' experiences and can augment or complement the current signals.

Previous work on affective computing has largely focused on decoding affective states in laboratory conditions, but less efforts have been devoted to combining affective sensing with conventional relevance estimation methods for realistic information-intensive tasks.

This realm poses significant challenges and opportunities when implementing these technologies in practical settings, es-

[1] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

pecially on a large scale. One opportunity lies in development of the current sensor technology to be feasible for real-world deployments. Peripheral sensors are gradually transcending laboratory confines and finding applications in smartwatches, virtual reality headsets, and other human-computer interaction hardware. Nevertheless, high-precision brain-computer interface sensors are not yet widely accessible for consumer-grade headsets. On the other hand, cameras are widespread and, although not always entirely reliable sources for affective information, they offer a simple means to tap into users' emotional experiences while enabling user-control for privacy.

Another opportunity pertains to the development of decoding technology, capable of concurrently assessing emotional states and modeling the stimuli or content that elicit these emotional responses. Machine learning techniques that are easily calibrated or can even learn with minimal guidance, are pivotal for broader adoption of affective technology, complementing thus existing signals of relevance and interest. As valence (positivity or negativity) has been shown to be easier to decode for high-arousal content (content that is already recognized to be relevant or drawing user attention), models can be developed to complement current relevance estimation methods [16].

Therefore, achieving consistently high accuracy in detecting discrete affective states is not always essential. Consider the scenario of modeling users' emotional responses, where merely discerning the valence of information for highly arousing (attention-captivating) content might suffice. Consequently, the existing decoding technology could already meet the requirements of numerous real-world situations, prioritizing performance in downstream tasks over the need for precise decoding accuracy. For example, robust affective annotation can be accomplished by combining signals from many individuals even when individual models are less accurate [2] or from an extensive period from an individual [1]. Moreover, information retrieval or recommender systems can benefit from affective information by complementing content-based models for information access, and also preventing potentially harmful content [16].

ACKNOWLEDGEMENTS

This work was supported by the Horizon 2020 FET program of the European Union (grant CHIST-ERA-20-BCI-001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147). This work was also supported by the Academy of Finland (grants 352915, 350323, 336085, 322653). This work is part of the project PCI2021-122036-2A, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

REFERENCES

[1] Oswald Barral, Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, page 389–399, New York, NY, USA, 2015. Association for Computing Machinery.

[2] Keith M Davis III, Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. Brainsourcing: Crowdsourcing recognition tasks via collaborative brain-computer interfacing. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1–14, 2020.

[3] Mario Frank, Tiffany Hwu, Sakshi Jain, Robert T. Knight, Ivan Martinovic, Prateek Mittal, Daniele Perito, Ivo Sluganovic, and Dawn Song. Using EEG-based BCI devices to subliminally probe for private information. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES)*, pages 133–136, 2017.

[4] Sze Chit Leong, Yuk Ming Tang, Chung Hin Lai, and CKM Lee. Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. *Computer Science Review*, 48:100545, 2023.

[5] Jia Zheng Lim, James Mountstephens, and Jason Teo. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors*, 20(8):2384, 2020.

[6] Radek Martinek, Martina Ladrova, Michaela Sidikova, Rene Jaros, Khosrow Behbehani, Radana Kahankova, and Aleksandra Kawala-Sterniuk. Advanced bioelectrical signal processing methods: Past, present, and future approach—part iii: Other biosignals. *Sensors*, 21(18):6064, 2021.

[7] Daniel McDuff, Rana el Kaliouby, Thibaud Senechal, May Amr, Jeffrey F. Cohn, and Rosalind Picard. Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected "in-the-wild". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2013.

[8] Jan Mizgajski and Mikołaj Morzy. Affective recommender systems in online news industry: how emotions influence reading choices. *User Model. User-Adapt. Interact.*, 29(2):345–379, 2019.

[9] Mimma Nardelli, Alberto Greco, Laura Sebastiani, and Enzo Pasquale Scilingo. ComEDA: A new tool for stress assessment based on electrodermal activity. *Comput. Biol. Med.*, 150, 2022.

[10] Robert Plutchik. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier, 1980.

[11] Sarah Prange, Sven Mayer, Maria-Lena Bittl, Mariam Hassib, and Florian Alt. Investigating user perceptions towards wearable mobile electromyography. In *Proceedings of the IFIP Conference on Human-Computer Interaction (INTERACT)*, pages 339–360, 2021.

[12] Zhihang Ren, Jefferson Ortega, Yifan Wang, Zhimin Chen, Yunhui Guo, Stella X. Yu, and David Whitney. VEATIC: Video-based emotion and affect tracking in context dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4467–4477, January 2024.

[13] Joshua Sabio, Nikolas S Williams, Genevieve M McArthur, and Nicholas A Badcock. A scoping review on the use of consumer-grade eeg devices for research. *Plos one*, 19(3):e0291186, 2024.

[14] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *J. Med. Eng. Technol.*, 32(6), 2008.

[15] Youddha Beer Singh and Shivani Goel. A systematic literature review of speech emotion recognition approaches. *Neurocomputing*, 2022.

[16] Ruotsalo Tuukka, Mäkelä Kalle, Spapé M. Miciel, and Leiva A. Luis. Feeling positive? predicting emotional image similarity from brain signals. In *Proc. ACM Conference on Multimedia*, 2023.

[17] Shuning Wang, Linghui Zhong, Yongjian Fu, Lili Chen, Ju Ren, and Yaoxue Zhang. Uface: Your smartphone can "hear" your facial expression! *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 8(1), mar 2024.

[18] Wenjing Yang and Yunhui Xie. Can robots elicit empathy? the effects of social robots' appearance on emotional contagion. *Computers in Human Behavior: Artificial Humans*, 2(1):100049, 2024.

[19] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, 59:103–126, 2020.

[20] Sicheng Zhao, Xingxu Yao, Jufeng Yang, Guoli Jia, Guiguang Ding, Tat-Seng Chua, Bjoern W Schuller, and Kurt Keutzer. Affective image content analysis: Two decades review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intellig.*, 44(10), 2021.