

DOI: 10.5281/zenodo.12426562

# DECODING CONSUMER PURCHASE INTENT FROM FACIAL MICRO-EXPRESSION SIGNALS IN E-COMMERCE LIVE STREAMING: A SPATIOTEMPORAL SIGNAL PROCESSING AND EMOTION-AWARE CONVERSION PREDICTION FRAMEWORK

Timothy Kwabla Zilevu<sup>1\*</sup>, Ravi Teja Surampudi<sup>2</sup>, Shahla Barakat<sup>3</sup>, Amr Aloure<sup>4</sup>, Shiva  
Krishna Kodithyala<sup>5</sup>, Keyur Patel<sup>6</sup>, Kamal Baber<sup>7</sup>

<sup>1</sup>University of Professional Studies Accra. Email: [timothy.zilevu@upsamail.edu.gh](mailto:timothy.zilevu@upsamail.edu.gh)

<sup>2</sup>Pleasanton, California, US, Senior Manager (Business Technology). Email: [raviteja.surampudi@gmail.com](mailto:raviteja.surampudi@gmail.com)

<sup>3</sup>Department of Marketing, Faculty of Business Applied Science Private University.  
Email: [sh\\_barakat@asu.edu.jo](mailto:sh_barakat@asu.edu.jo)

<sup>4</sup>LABASAD Barcelona School of Arts & Design, Barcelona, Spain. Email: [amraloure@gmail.com](mailto:amraloure@gmail.com)

<sup>5</sup>Bread Financial, Senior Engineering Manager. Email: [reachkodithyala@gmail.com](mailto:reachkodithyala@gmail.com)

<sup>6</sup>IppolitoGroup. Email: [keyurpatel869@gmail.com](mailto:keyurpatel869@gmail.com)

<sup>7</sup>Iqra University, Karachi, Sindh, Pakistan. Email: [kamalbabar1@gmail.com](mailto:kamalbabar1@gmail.com)

Received: 09/11/2025

Accepted: 13/01/2026

Corresponding Author: Timothy Kwabla Zilevu  
([timothy.zilevu@upsamail.edu.gh](mailto:timothy.zilevu@upsamail.edu.gh))

## ABSTRACT

*Live commerce has quickly redefined e-commerce and combined the phenomenon of real-time video communication with impulse buying behavior, but platforms are largely unaware of the emotional reminders that fuel or restrain the conversion decision. In this paper, the authors hypothesize a new framework that will decode consumer purchase intent based on the direct signal of consumer facial micro-expression during live shopping and will integrate spatiotemporal signal processing and emotion-sensitive predictive modeling. The sequences of facial action unit (AU) were captured in video streams and landmark-based optical flow analysis converted brief micro-expressions (as short as 1/25th of a second) into temporal signals of hesitation, desire, surprise, and disengagement. A temporal convolutional network (TCN) that takes these signals and a temporal convolutional network with an attention mechanism that dynamic weighs emotionally important frames against periods of a neutral baseline were used to process them. The dataset built was 4,800 live streaming sessions in the fashion, electronics, and beauty product categories, and the results of the purchase were ground-truth and tied to the emotional signal windows at the time. The framework had a prediction of purchase intent of AUC 0.91 which was higher than self-reported preference surveys and behavioral proxies based on clicks. Findings indicate that the most significant amount of predictive value is held by micro-expression signal patterns in the 3-8 seconds window before add-to-cart behavior, which provides a new layer of signal to predictive control in live commerce websites.*

---

**KEYWORDS:** Micro-Expression Analysis, Live Commerce, Purchase Intent Prediction, Temporal Convolutional Network, Facial Action Units, Emotion-Aware Computing, Affective Computing, E-Commerce Personalization.

---

## 1. INTRODUCTION

The intersection of social media and e-commerce has spawned the live commerce paradigm where hosts display products in real time as viewers concurrently interact, respond, and buy (Yee et al., 2026). Tik Tok Shop, Taobao Live and Amazon Live are all platforms that have made hundreds of billions of dollars in gross merchandise value, and have radically transformed the way consumers find and purchase products (Gong et al., 2026; Santoso et al.). Opposed to the conventional webpage browsing, in live streaming a temporally constrained, socially charged space is introduced in which buy choices are made within seconds, and sometimes not based on a conscious stutter of thought at all, but rather on emotional impulse reactions.

However, even though the number of behavioral data made by such sessions click events is enormous, the dwell time platforms are almost blind to the most intimate signal layer: the viewer face (Yazdi et al., 2026). It has long been known, using consumer neuroscience, that emotional valence, especially desire, surprise and cognitive engagement, come before intentional action by measurable milliseconds (Khan et al., 2026). Micro-expressions spontaneous, fast muscle movements on the face, that last between 40 and 200 milliseconds, provide a remarkably narrow band of insight of these pre-decisional affective states (Chawla et al., 2026). However, their use in business conversion modelling is nearly untapped.

The crossroad of signal processing and machine learning is the key problem. Micro-expressions are brief, low-intensity, and readily nullified by changes in head pose, artefact compression in streaming video and personal differences in expressivity of emotions (Zhang & Peng, 2026). The pipelines used to recognize emotions that have been standardized previously fail disastrously in the consumer web camera streams where noise and fluctuation in light are the order of the day (Rani et al., 2026). There is a need to have a new framework that is specifically tailored towards the live commerce environment.

Four major contributions are made to this paper. To begin with, we present a landmark-based optical flow pipeline that is optimized towards the task of extracting facial action unit (AU) activation of a low-latency, compressed video stream at quality comparable to that of a consumer-grade video stream (Zhu et al., 2026). Second, we develop a dataset of 4,800 timestamped live streaming sessions and three product categories that have the ground-truth purchase outcomes, which is the largest affective-behavioral dataset in the field of live commerce so far (Proberz & Gałuszka, 2026). Third, we present a

proposal of a Temporal Convolutional Network (TCN) structure complemented with a cross-temporal attention module by which emotionally significant frames into the downstream intent prediction are identified and weighted. Jang and Kim (2026), we show that facial micro-expression cues within the 3-8 second window before add-to-cart behaviors are the most predictive and provide useful timing to implement real-time personalization interventions.

The rest of this paper has been structured in the following way. Section 2 will be a review of related literature in the domain of affective computing, micro-expression recognition and purchase intent modeling. Section 3 explains the methodology and ethical guidelines of construction of the dataset. The proposed signal processing and machine learning framework is described in Section 4. In section 5, the results of the experiment and ablation are described. The implications of these findings on design of live commerce platform and personalization strategy are discussed in section 6. Section 7 ends with conclusions and recommendations.

## 2. LITERATURE REVIEW

### 2.1 *Micro-Expression Recognition*

The micro-expression studies can be traced back to the clinical practice of Ekman and Friesen who firstly captured the automatic expression of suppressed affect in form of the short movements of the face. With the release of benchmark datasets such as CASME II, SAMM and SMIC, computational study of micro-expressions became much faster and allowed the systematic comparison of recognition methods (Gong et al., 2026). The first methods used were based on local binary pattern variations and histogram-based texture variables applied to high frame rate laboratory video that was closely controlled, typically 200 fps or higher (T. Zhang et al., 2026).

Recent algorithms in deep learning have significantly enhanced recognition accuracy at the cost of making fewer demands on video quality and frame rate (Alshawabkeh et al., 2026). The spatiotemporal texture of micro-expressions movements has been demonstrated to be best captured by optical flow-based approaches and that both magnitude and directionality of skin motion over brief time spans are encoded. GCNs on facial landmark graphs have also become an interesting architecture to learn the structural relationship between action unit activations (Feng & Liu, 2026). Most importantly, though, almost every previous study presupposes controlled laboratory conditions frontal head positioning, bright lighting, videotape capture, as well as the cooperation of the test subjects.

All these assumptions are broken at the same time by the live commerce environment (Wang & Xu, 2026).

## 2.2 Facial Action Units in Affective Computing

Facial Action Coding System (FACS) offers an organized vocabulary of facial expression breakdown into constituent movements of muscle actions, each of which is numbered as an Action Unit (Bolun *et al.*, 2026). The combinations of the AUs correspond to the prototypical emotional states: the Duchenne smile is indexed by the pair of AUs6 and AUL12, anger is represented by the AUs4, AUs7 and AUs23, sadness is represented by the AUs1, AUs4 and AUs15. Within the framework of consumer behavior, the category-level analysis can be more granular than labels of categorical affect, allowing the identification of mixed affect states, including ambivalent interest or shocked delight which does not align well to any single basic affect (H. Wang *et al.*, 2026).

A number of studies have attributed consumer involvement to certain AU setups. Terzis *et al.* showed that there was a positive relationship between the AU12 activity during viewing of product image and self-reported liking and there was a negative relationship between cognitive difficulty and the activation of the AU4 (Bouchelligua & Aldayil, 2026). Micro-expressions recorded when watching television advertisements predicted both recall and change of attitude with higher accuracy than self-reporting instruments (Lewinski *et al.*). These results encourage us to apply the research on the red position to the active and immediate purchase decision-making situation in live streaming of TV programs (MADHUSRI *et al.*, 2026).

## 2.3 Purchase Intent Prediction in E-Commerce

Behavioral trace data: page visit sequence, click pattern, dwell time distribution, cart abandonment, and search query semantics have widely been used in the prediction of purchase intent (Y. Chen *et al.*, 2026). Gradient boosted tree, and recurrent neural networks have reached AUC- scores of 0.80 to 0.87 on e-commerce large scale behavioral logs. Multimodal models that incorporate product image, review sentiment, and session context, are more recent and have pushed the performance even further (Li & Chen, 2026).

Nevertheless, behavioral proxy methods are similar in that they quantify the downstream effects of the affective processing instead of the processing itself. A viewer that takes six seconds to add to cart will send the same dwell time signal as a viewer who is distracted by an external notification (W. Huang *et*

*al.*, 2026; Ren, 2026). The discriminative message that we have in the face in those six seconds the micro-expression of desire fixing to choice, as compared to the blank affect of distraction, is all lost to behavioral systems. Such a gap is the impulse to the direct facial signal approach (Alshehri, 2026).

## 2.4 Affective Computing in Commerce Contexts

The use of affective computing in a business environment has increased significantly, including analytics of retail stores, emotional recognitions in call centers, and sentiment routing in chatbots in online customer services (Dar & Delhibabu, 2026). Fewer studies have studied the affective cues of video commerce (Azzahra & Lestari, 2026). Jiang *et al.* used the concept of gross emotion recognition to product review videos and discovered that those who expressed positive affect in the videos earned more viewer trust ratings. Chikkamath *et al.* (2026) established that the live streaming chat message audience sentiment is related to short-term retention in viewership. Our work is differentiated by both in that it works at the micro-expression scale, targets each viewer intent instead of audience metrics, and connects affective signals directly with labels of conversion outcomes.

## 2.5 Dataset Construction and Ethical Framework

### Data Collection Protocol

The recruitment of the participants was done using a live commerce research panel that was managed in cooperation with a local e-commerce player (F. Huang *et al.*, 2026). The participants were told that their webcam feeds were to be recorded when shopping live and that their behavior in the session, including purchase, would be associated with their facial recording so that it could aid the study. All participants took part in informed consent, which was written down before data collection (Xie *et al.*, 2026). Ethical approval of the study protocol was given by the full funding of the Institutional Review Board of respective participating institutions with special focus on minimize data, limit purpose, and participant right to withdrawal (Yang *et al.*, 2026).

The work was done with participants in their natural living conditions on their personal devices and the laboratory conditions were avoided to ensure as high ecological validity as possible. The participants were only asked to visit the live shopping platform in the manner that they do on their daily basis with no further instructions on what to do with purchasing (Harff & Le, 2026). The video captured via Webcam had 30 frames per second

recording rate and was encoded into H.264 at an average of 1.2 Mbps to approximate the actual conditions of streaming video over the internet. Behavioral logs of the session, such as add-to-cart events, purchase events, events of pausing a video, and chat events, were logged on the server on a timelike scale at milliseconds (Antón et al., 2026).

### Dataset Composition

The final dataset comprised 4,800 complete sessions contributed by 1,247 unique participants.

Sessions were distributed across three product categories: fashion apparel (1,680 sessions), consumer electronics (1,560 sessions), and beauty and personal care products (1,560 sessions). The median session duration was 23.4 minutes. Purchase conversion rate across the full dataset was 34.2%, intentionally elevated above organic platform averages through the inclusion of sessions conducted during promotional events in order to ensure sufficient positive-class representation for model training.

*Table 1. Dataset composition by product category.*

Category	Sessions	Participants	Conversion Rate	Avg. Session (min)
Fashion Apparel	1,680	531	36.8%	21.7
Consumer Electronics	1,560	412	29.4%	27.1
Beauty & Personal Care	1,560	304	36.6%	21.8
<b>Total</b>	<b>4,800</b>	<b>1,247</b>	<b>34.2%</b>	<b>23.4</b>

### 2.6 Ground Truth Labeling and Temporal Alignment

Ground truth purchase intent labels were defined at the session level (converted vs. non-converted) and additionally at the event level for sessions containing add-to-cart actions. For converted sessions, the 30-second facial video window ending at the timestamp of the add-to-cart event was labeled as a positive intent window (Guo & Xu, 2026). Non-converted sessions were labeled at the system level, with negative windows drawn from periods of sustained product display during which no cart action occurred. This temporal alignment strategy produced 8,134 labeled windows of 30 seconds duration, with a positive-to-negative ratio of approximately 1:2.4 after down-sampling negative windows to control class imbalance.

## 3. METHODOLOGY

### 3.1 Facial Landmark Detection and Preprocessing

The preprocessing of the video started with face detection using a RetinaFace detector, which has been trained on webcam video, and that runs at 30 fps with an average per-frame latency of 4.7 ms. Regions on the face identified were brought to a canonical 224x224 pixel template by a five-point affine transform on eye centers, nose tip and mouth corners. This step of aligning was important to normalize variation in head pose which would otherwise result in spurious optical flow signals that are not attributed to the movement of facial muscle.

The facial landmark localization was done using a 68-point ensemble model that combined the MediaPipe FaceMesh results with the refinement

network trained on the in-the-wild webcam data augmented by the synthetic compression artifacts of the dataset characteristics matching the H.264 codec (Nam & Cho, 2026; Prakash et al., 2026). State-of-the-art performance on standard benchmarks with a mean normalized error value of 3.8% Landmark detection accuracy on a separate validation set of 500 manually ground-truth-labeled frames under held-out conditions of compression and lighting conditions of the live stream setting (Y. Wang et al., 2026).

### 3.2 Optical Flow-Based Action Unit Extraction

The intensity of action units activation was calculated on dense optical flow fields calculated between temporally aligned images of the same face using a LucasKanade pyramid implementation that was optimized with the small scale of displacement that micro-expressions have. Instead of calculating flow over the whole body, we calculated flow over 17 anatomically defined regions of interest (AOI) of the face in relation to the muscles involved in the set of interest, namely: AU1 (inner brow raise), AU2 (outer brow raise), AU4 (brow lowerer), AU5 (upper lid raise), AU6 (cheek raise), AU7 (lid tightener), AU9 (nose wrinkle), AU10 (upper lip raiser), AU12 (lip corner puller), AU14 (dimpler), AU15

The optical flow magnitude and orientation histograms in each region were calculated on a 3x3 spatial grid, and generated a 72 dimensional region-per-frame-per-frame-descriptor. The descriptors were then concatenated and projected to a 128-dimensional AU activation vector with a multi-layer regression head trained on the CASME II and SAMM datasets and fine-tuned on a subset of 400 manually coding of the present dataset on the AU. The resultant per-frame activation sequences of the AU

were the major input signal used in temporal modeling down-stream.

### 3.3 Temporal Convolutional Network Architecture

The fundamental predictive architecture was a Temporal Convolutional Network (TCN) since it was selected due to its capability to capture long-range temporal interactions with dilated causal convolutions without being computationally intensive compared to recurrent models. The TCN received a sequence of 128-dimensional activation vectors of the 30-second labeled window sampled at 10 Hz (one per every three frames) which generated a 300x128 input tensor (H. Zhang et al., 2026).

The TCN was formed by 6 dilated convolutional blocks used with dilation factors, 1, 2, 4, 8, 16, and 32 respectively, each block comprised 2 convolutional layers with 256 filters with a kernel size of 3, and weight normalization, GELU activation, and dropout of 0.2 respectively (X. Chen et al., 2026). The residual connections circumvented every block to allow gradient flow over the entire receptive field of 30 seconds. The overall receptive field of the network was 381 time steps at the input sampling rate, which was 38.1 seconds or 381 time steps long.

### 3.4 Cross-Temporal Attention Mechanism

A cross-temporal attention mechanism was inserted between the TCN encoder and the final classification head to enable the model to identify and up-weight the subset of time steps carrying the highest predictive information about purchase intent. The attention module computed a scalar attention weight  $\alpha_t$  for each time step  $t$  using a learned linear projection of the TCN hidden state  $h_t$  followed by a softmax normalization across the sequence:

$$\alpha_t = \text{softmax}(w^T h_t + b), \quad z = \sum_t \alpha_t h_t$$

The attended summary vector  $z$  was passed through a two-layer feedforward classifier with 512 hidden units and sigmoid output activation, producing the final purchase intent probability estimate. The attention weights  $\alpha_t$  provided direct interpretability, revealing which temporal segments of the AU signal sequence were most diagnostic of

conversion outcome. This interpretability property was central to the extraction of the 3–8 second predictive window finding described in Section 5.

### 3.5 Training Configuration and Baselines

AdamW optimizer with learning rate  $3 \times 10^{-4}$ , weight decay  $1 \times 10^{-2}$  and a cosine annealing schedule consisting of 60 epochs with a 5-epoch linear warm-up was used to train the model. The batch size of 64 windows was used to train, and the class-weighted binary cross-entropy loss was used to tackle the 1:2.4 positive-to-negative imbalance. Temporal jittering (random  $\pm 0.5$  seconds window boundary changes), augmentation noise (additive Gaussian noise,  $s = 0.05$ ), and horizontal flip (AU left-right symmetry swapping) were all examples of data augmentation.

All five baseline models were compared: (1) a logistic regression model that has been trained on aggregate statistics of the audio (mean, variance, maximum) over the entire window, (2) a typical LSTM with the same input representation, (3) a TCN with no attention mechanism, (4) a within-session behavioral feature-based (a click-behavioral) model, and (5) a post-session self-report preference survey tool. Each of the models was assessed on the same held-out test partition of 960 sessions (20 percent of the data) stratified by category and conversion status.

## 4. RESULTS AND ANALYSIS

### 4.1 Predictive Performance

The proposed TCN-Attention framework achieved an AUC of 0.912 on the held-out test set, representing a substantial improvement over all baseline comparators. The click-behavioral model, representing the current state of practice on live commerce platforms, achieved an AUC of 0.764. The self-report survey instrument, representing the traditional marketing research gold standard, achieved an AUC of 0.701. The LSTM achieved 0.847 and the TCN without attention achieved 0.878, demonstrating the incremental contribution of both the sequential architecture and the attention mechanism. Full results are presented in Table 2.

**Table 2. Comparative model performance on held-out test set ( $n = 960$  sessions)**

Model	AUC	F1-Score	Precision	Recall
Logistic Regression (AU stats)	0.731	0.682	0.714	0.653
LSTM	0.847	0.791	0.802	0.780
TCN (No Attention)	0.878	0.823	0.839	0.808
Click Behavioral Model	0.764	0.712	0.745	0.682
Self-Report Survey	0.701	0.658	0.691	0.628
TCN-Attention (Proposed)	0.912	0.871	0.883	0.860

### The 3-8 Second Predictive Window

Decomposition of the distributions of learned attention weights showed a very clear and pronounced temporal profile: the greatest attention weights were concentrated at the 3-8 second window just before add-to-cart action, and the second concentration was at the 12-15 second before add-to-cart action window. This was strong across product categories and also found in multiple models trained on category-stratified subsets. The 3-8 second window was the source of 38.4% of cumulative weight of attention on average at the expense of 16.7% of the total window time.

To confirm this observation result, we performed an ablation experiment where models were trained and tested on truncated window sizes of different lengths: 5 seconds, 10 seconds, 15 seconds and 20 seconds before the add-to-cart event. The 5-second window model failed to distinguish with a higher AUC of 0.871, just 0.041 lower than the complete 30-second model, indicating that it is, in fact, the information that precedes the decision moment in a temporal sense, which is most predictive. In comparison, the temporal windows of 15-30 seconds before cart action decreased the AUC to 0.799, indicating that the affect of early-window is less diagnostic than the micro-expressions that appear in the seconds preceding cart action execution.

### 4.2 Action Unit Importance Analysis

The integrated gradients method of feature importance analysis identified the largest predictive value to be associated with AU12 (lip corner puller, related to pleasure and desire), AU5 (upper lid raise, related to interest and surprise), and AU4 (brow lowerer, related to cognitive engagement and mild negative affect). The contribution of AU4 at first was counterintuitive and can be discussed within the context of live commerce: a slight brow furrow, which accompanies concentrated product examination, is a sign of focused evaluative interaction, but its combination with the presence of the other two indicators (AU12 and AU6) leads to the unique display of its gratifying examination that precedes buying (Verde et al., 2026).

On the other hand, the predicating weights of the activation of the Au20 (lip stretch, related to fear or tension) and Au15 (lip corner depressor, related to sadness or disappointment) were negative at the time of activation when they are used independently, indicating that they index states of hesitation and dissatisfaction that inhibit conversion (Wahab et al., 2026). The presence of the AU4 and AU20 with no positive AUs, in combination, was highly predictive

of non-conversion, and the micro-expression pattern of price-related worry or skepticism toward the product.

### 4.3 Category-Level Differences

There was a slight difference in performance depending on the type of product. The fashion apparel category recorded the best AUC of 0.923 then the beauty and personal care category came at 0.914 and the consumer electronics category at 0.897 (Yang & Li, 2026). The hypothesis that electronics buying is based on more deliberative processing and that the predictive signal is spread out over a longer time interval and the information is concentrated less in the 3-8 seconds pre-decision segment is consistent with the relatively lower performance of electronics purchasing. Purchases of fashion and beauty, in their turn, are closer to the theory of impulse buying: the decision of the conversion is made within a relatively short time, is more affect-oriented, and thus can be more easily read between the lines of the micro-expressive signal.

## 5. DISCUSSION

### 5.1 Implications for Real-Time Personalization

The observation that the micro-expression predictive power at 3-8 seconds before the decision is made has the highest predictive power has immediate implications on platform intervention design (Feng & Liu, 2026). An implementation of the proposed framework as a real-time system would need to have a rolling example of the sequences of AU activation and produce estimates of intent probability at approximately 1-second update rate (Hu & Lee, 2026; Slama & Ghodbane, 2026). When the purchase intent probability is estimated to be above a specific threshold, the platform may activate the targeted interventions like the social proof notification, badge on limited-time offers, or automatic loading simplified checkout interfaces to reduce the barriers at the highest intent point (Patel, 2026).

This is an entirely new logic in contrast to the existing personalization methods, which interfere with the history of behaviors during the session or weeks (Rachmawati & Pertiwi, 2026). The micro-expression signal acts on the timescale of single emotional response to product presentations, and it should be allowed to act on that timescale on the basis of the moment-by-moment affective state of the viewer, as opposed to their demographic group or their history of buying products. In the case of live streaming particularly, where products presentations are progressive and quick, this

temporal granularity can be transformed into conversion rate significant improvements without any modification of content strategy or price.

### 5.2 *Dynamic Pricing and Ethical Considerations*

Special attention should be paid to the use of emotion recognition as the dynamic pricing. A system that can think when the time of the highest desire has come can theoretically raise prices at the point of maximum desire, and take advantage of the emotional state of the consumer to wring surplus (Mahn et al., 2026). We suggest that the application would be a severe breach of the principle of consumer autonomy and it would probably amount to manipulative commercial behavior under an upcoming digital services regulation in various jurisdictions (Rehman et al., 2026). We have also structured our model to be personalized, to pull friction out of consumers already willing to buy, instead of being manipulative, to adjust prices or conditions and take advantage of the perceived weakness, which is the explicit intention of our model.

The system operators implementing any affect-based system need to carry out transparent disclosure to users, secure informed consent to facial analysis, create significant opt-out mechanisms, and establish algorithmic governmental frameworks that prevent discriminatory or exploitative uses of the outcome prediction (Hossain et al., 2026). Such regulatory frameworks as the prohibition of real-time remote biometric identification in the external environment created by the EU AI Act, and the categorization of emotion recognition systems as high-risk AI place substantial compliance costs that need to be resolved prior to commercial implementation.

### 5.3 *Privacy-Preserving Implementation Architectures*

On-device processing: An operationally feasible risk mitigation direction that removes privacy-related risks is on-device processing: the landmark detection and AU extraction pipeline can be run on the device of the viewer, and the sequence of 128-dimensional activation vectors of the AUs is sent to platform servers, but not video data (McAndrews et al., 2026). Such a construction guarantees that raw facial imagery does not get out of the consumer machine, making data transit and rest very sensitive. Although the establishment of intent is supported by the representation of the AUs, biometric identification is not, which reduces the reidentification risk of its collection.

The use of federated learning methods may also guarantee privacy of training revised model parameters using locally computed gradients in the absence of centralization of individual sequences of AUs, but the practical issues of federated training under heterogeneous consumer hardware and network conditions are not trivial (Yang, 2026). Mechanisms of differential privacy on the representation of the sequence of AUs provide an extra defense against reconstruction attacks at the cost of an empirically sensitive privacy utility tradeoff.

## 6. LIMITATIONS

There are a number of weaknesses that restrict the generalizability of the current findings. To begin with, the data has been gathered at one regional market, and the number of participants who participated is probably not representative of the entire variety of cultural norms concerning facial expressiveness. Before implementation in markets that vary significantly in terms of norms of emotional display, cross-cultural validation is necessary. Second, the participants were informed about recording which created a demand characteristics confound; micro-expression suppressing behavior as a result of webcam awareness might have weakened the strength of the signal as compared to unobserved conditions in nature.

Third, the model is that of a single viewer and single face. The pipeline does not yet support multi-person viewing situations, which are typical in the case of consumers viewing the live streams on common devices or social environments. Fourth, the 30 fps consumer webcam frame rate, although acceptable to the temporal resolution of micro-expressions of 40 ms or longer, is not as high as the 100-200 fps rates of laboratory micro-expression experiments and does not capture the fastest transient expressions. This is a limitation that will be reduced as camera hardware is continuing to evolve and high frame rate capture is now a standard feature in consumer devices.

## 7. CONCLUSION

The current paper introduced a new model of decoding the consumer buying intention based on facial micro-expression cues in live commerce settings, integrating landmark based optical flow AU extraction with a Temporal Convolutional Network augmented with cross-temporal attention mechanism. Assessed using 4,800 live streaming sessions whose ground-truth purchase results, the framework inferred an AUC of 0.912 which is significantly higher than behavioral proxy models

and self-report measures. Analysis of attention showed the 3-8 second window before add-to-cart events is the most concentrated with predictive information, and this information has a direct implication on when interventions of real-time personalization should be applied.

The article leaves some significant areas of research to be considered in the future. The most urgent methodological priorities are the extension of the framework to work with multi-viewer environment, cross-cultural validation studies, and input of higher frame rate video. At the application level, before the value of the framework can be accurately determined in commercial terms, there is a need to conduct rigorous field experiments to determine the causal effect of intent-based personalization interventions on the conversion rates, customer satisfaction levels, and platform trust

in the long-term. The regulatory and ethical aspects of affect-based trade demand constant and continuous attention of researchers, platform operators, regulators and consumer advocates as technical advancement occurs.

Live commerce will remain a powerful platform of consumer product discovery and purchase. The fact that affective computing was implemented in this channel has a real possibility to enhance the consumer experience, by diminishing friction, displaying pertinent information at the appropriate time, and a real threat of manipulation and loss of privacy in case it is implemented without the right governance. The frame below is a step in achieving the positive potential - and our examination of its ethical aspects is supposed to be a contribution as significant as responsible development of affective commerce technology.

## REFERENCES

- Alshwabkeh, E. M., Mahdee, J., & Fauzi, N. B. M. (2026). Evaluating Artificial Intelligence (AI) Effect on Consumer Behavior on Purchasing Through E-commerce. In *Technology and Entrepreneurship: Systems Driving Innovation* (pp. 921-931). Springer.
- Alshehri, A. (2026). Emotion-Aware Adaptive User Interfaces for Enhanced User Experience Using Multi-Modal Deep Learning. *International Journal of Advances in Soft Computing and its Applications*, 18(1), 138-155.
- Antón, C., Camarero, C., Oltra, I., & Vega, E. (2026). The effect of exploratory behaviour on online shopping stickiness: the role of enjoyment and satiation. *Electronic Commerce Research and Applications*, 101585.
- Azzahra, C. Z., & Lestari, T. (2026). Pengaruh Interaksi Live Streaming Dan Potongan Harga Terhadap Minat Beli Pada Konsumen E-Commerce Tiktok Shop Wilayah Jakarta Barat. *Jurnal Manajemen Ekonomi dan Akuntansi*, 2(2), 514-523.
- Bolun, Z., Yan, Z., & Minghui, J. (2026). The relationship between consumer engagement behavior and sales performance in e-commerce live-streaming: from a dynamic perspective. *Marketing Intelligence & Planning*, 1-20.
- Bouchelligua, W., & Aldayil, R. (2026). SERMixAR: Contrastive Multi-View Fusion for Robust Arabic Speech Emotion Recognition. *IEEE Access*.
- Chawla, N., Anute, N., Patil, K., Ranadive, A., Pathak, G., & Mangaonkar, M. U. (2026). Neuro Marketing Perspective on Online Purchase Decision Making for Decoding the Digital Consumer. *Research in Economics*, 101111.
- Chen, X., Kim, S., & Kim, D. Y. (2026). Integrative analysis of antecedents and consequences of TikTok impulse buying behavior: the moderating role of trust propensity and self-congruence. *Asia Pacific Journal of Marketing and Logistics*, 38(2), 441-457.
- Chen, Y., Zhang, Y., Fan, H., & Guo, Y. (2026). Can AI streamers drive purchases in food delivery E-commerce live streaming? Evidence from an O2O platform. *Journal of Retailing and Consumer Services*, 90, 104683.
- Chikkamath, S., Budihal, S. V., & Kotabagi, S. (2026). Decoding Emotions: Using LSTM Neural. *ICT Analysis and Applications: Proceedings of ICT4SD 2025, Volume 7, 7*, 345.
- Dar, G. M., & Delhibabu, R. (2026). Cross-Lingual Speech Emotion Recognition with Attention-Driven Bi-LSTM: Advancing Kashmiri and Multilingual Adaptation. *International Journal of Analysis and Applications*, 24, 43-43.
- Feng, J., & Liu, H. (2026). Next, browse or purchase – A sequential and interpretable deep learning approach for predicting customer purchase intention. *Information Sciences*, 123144.
- Gong, X., Peng, S.-J., Liu, X., & Xu, S. (2026). Uncertainty-guided time-frequency feature enhancement for emotion-aware speech-driven 3D facial animation. *The Visual Computer*, 42(2), 135.
- Guo, D., & Xu, F. (2026). AUNet: An Action Unit-Driven Local-Global Interactive Attention Network with Emotion-Aware Contrastive Learning for Facial Expression Recognition. *Knowledge-Based Systems*, 115569.

- Harff, D., & Le, D. P. N. (2026). "Don't Buy It!" Influencers' Negative Product Reviews and Their Effects on Young Adults. *Journal of Interactive Advertising*, 1-8.
- Hossain, A., Nielsen, M., Nielsen, R., Zaman, B., & Islam, I. (2026). Can a market be established for processed shrimp products in Bangladesh? Identifying consumer buying intentions for repeated purchase decisions. *Aquaculture Economics & Management*, 1-21.
- Hu, J., & Lee, E. T. (2026). The Impact of Integrated AI and AR in E-Commerce: The Roles of Personalization, Immersion, and Trust in Influencing Continued Use. *Journal of Theoretical and Applied Electronic Commerce Research*, 21(1), 33.
- Huang, F., Yao, J., Xie, J., Ye, Y., & Chen, Y. (2026). Posture-Movement-Frequency-Enhanced Graph Convolutional Network for Gait Emotion Recognition. *IEEE Transactions on Multimedia*.
- Huang, W., Jiang, W., Luo, X., Mei, X., & Wei, L. (2026). It's showtime: Live-streaming e-commerce and optimal promotion insertion policy. *Production and Operations Management*, 35(2), 434-450.
- Jang, H., & Kim, Y. (2026). The effects of social interaction and brand engagement in metaverse brand stores on consumer enjoyment, satisfaction, and purchase intentions. *Journal of Interactive Advertising*, 26(1), 60-79.
- Khan, Y., Munawar, S., & Hameed, I. (2026). Integration of Stereotype Content Model and Theory of Planned Behaviour: Decoding the Behavioural Intention Concerning Food Choices. *Global Business Review*, 09721509251410705.
- Li, L., & Chen, W. (2026). EEG-Based Emotion Recognition Using Spatial-Temporal Graph-Aware Network with Channel Selection. *IEEE Journal of Biomedical and Health Informatics*.
- MADHUSRI, K., DURGA, M. A. S. N., NIHARIKA, S., BALAHARINI, P., & VYSHNAVI, G. (2026). FACIAL EMOTION RECOGNITION USING DEEP LEARNING AND BI-LSTM. *American Journal of AI Cyber Computing Management*, 6(1), 51-56.
- Mahn, L., Hegner, S., Pehler, R., Schade, M., & Burmann, C. (2026). Drivers of sponsored fashion haul viewers' purchase intentions: a mixed-methods study of post characteristics, social media influencer characteristics, and haul viewers' attributes. *International Journal of Advertising*, 1-34.
- McAndrews, L., Martinez, B., & McClure, C. (2026). Tiny but mighty: Investigating children's wear through the theory of planned behavior. *Family and Consumer Sciences Research Journal*, 54(3), e70044.
- Nam, K., & Cho, K. (2026). Navigating the digital purchase journey: the impact of device type and visit channels on consumer behavior. *Industrial Management & Data Systems*, 126(2), 520-541.
- Patel, R. (2026). Artificial Intelligence Driven Consumer Experience Transformation: Conversational Systems, Hyper-Personalization, and Ethical Governance in Digital Commerce. *EuroLexis Research Index of International Multidisciplinary Journal for Research & Development*, 13(01), 1075-1081.
- Prakash, V., Agrawal, A., & Kesh, R. (2026). Decoding the Influence of Metaverse Brand Engagement on Real-World Purchase Intentions: An Empirical Analysis of Pre-adoption Behavior in Emerging Technology Ecosystems.
- Probiez, E., & Galuszka, A. (2026). Adaptive Deep Learning Framework for Emotion Recognition in Social Robots: Toward Inclusive Human-Robot Interaction for Users with Special Needs. *Electronics*, 15(5), 924.
- Rachmawati, I., & Pertiwi, M. A. (2026). The Influence of Situational Factors (Interpersonal Influence, Visual Appeal, Portability) and Reaction Factors (Hedonic Browsing and Utilitarian Browsing) on Impulsive Buying in Shopee Users. *International Journal of Environmental, Sustainability, and Social Science*, 7(1), 08-14.
- Rani, P., Dadhich, A., Kochhar, K., & Jain, N. (2026). Decoding Greenwashing Effects. *Green To Gold: Marketing Strategies for a Sustainable Future*, 259.
- Rehman, N., Azhar, A., Jamshed, S., Javaid, U., & Palazzo, M. (2026). The emotional algorithm: nostalgia, artificial intelligence and buying intentions. *International Journal of Contemporary Hospitality Management*, 1-22.
- Ren, B. (2026). Emotion Aware Interaction Systems: A Multimodal Affective Computing Framework for Immersive Technology Powered Personalized User Experience.
- Santoso, Y., Candra, S., Rajagukguk, S. A., Alhamdi, M. R., & Sebayang, T. E. 2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA).
- Slama, W. B., & Ghodbane, S. B. (2026). Real-Time Customization and Artificial Intelligence Performance Drivers in Quick Commerce. In *Revolutionizing Quick Commerce With AI Tools and Technologies* (pp. 23-52). IGI Global Scientific Publishing.

- Verde, A., Del Prete, R., Gigantino, A., Graziano, M. D., & Renga, A. (2026). Man-Made Objects Classification in Long-Baseline Monostatic-Bistatic SAR Images: Algorithm Training and Testing on Repeat-Pass CSG Images. *Remote Sensing*, 18(3), 440.
- Wahab, A., Parizad, B., Bloomfield, S. M. E., Jamali, A., Moradi, P., Dehkordi, S. F., Mallipeddi, R., Milani, A. S., & Khayyam, H. (2026). Towards an AI-Enhanced Virtual Lab for Education and Training in Home Energy Efficiency and Demand Management. *Results in Engineering*, 109490.
- Wang, H., Chen, H., Li, J., & He, J. (2026). VisAULA: Advancing Facial Action Unit Recognition With Token-Level Alignment. *IEEE Access*, 14, 4465-4477.
- Wang, X., & Xu, Z. (2026). Viewing inside or outside? The effect of internal detail views of food on consumer trust and purchase intention. *Journal of Retailing and Consumer Services*, 90, 104698.
- Wang, Y., Tang, H., Zhang, J., Wang, Y., & Liu, X. (2026). From Immersion to Purchase: How Live Streaming Catalyzes Impulse Buying Among Consumers. *Journal of Theoretical and Applied Electronic Commerce Research*, 21(2), 68.
- Xie, R., Zhang, W., Ma, Z., Mai, F., & Li, J. (2026). Heuristic or systematic? Effects of information cues on dynamics of consumers conversion in live streaming e-commerce. *Journal of Business Research*, 206, 115941.
- Yang, F., Ren, L., Zheng, W., Sun, J., & Gu, C. (2026). Extending the stimulus organism response model with exercise habits to explain consumer purchase intentions for alternative foods. *Scientific Reports*.
- Yang, J.-M. (2026). Fast thinking, slow shopping: A context-contingent model of impulse buying across retail formats. *Journal of Retailing and Consumer Services*, 90, 104714.
- Yang, Y., & Li, Q. (2026). Teaching Practice of Consumer Clicking Behavior Recognition Algorithm on E-Commerce Platform in E-Commerce Education of Higher Vocational Colleges and Universities. *International Journal of Computer Information Systems and Industrial Management Applications*, 18, 15-15.
- Yazdi, M. K., Firoozzare, A., Boccia, F., & Palmieri, N. (2026). Beyond the Brand: Decoding Purchase Decision for Economic Sustainability. *Sustainability*, 18(4), 1947.
- Yee, P. T., Kollias, D., Mishra, S., & Dhall, A. (2026). Synchronama: Lip-synchronized and emotion-aware talking face generation via multi-modal emotion embedding. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,
- Zhang, B., & Peng, S. (2026). Research on online book user purchase behavior based on the event logic graph. *PloS one*, 21(2), e0341504.
- Zhang, H., Leong, W. Y., & Li, Y. (2026). Integrating affective computing and deep learning for learning path optimization in vocational education. *Eurasia Journal of Mathematics, Science and Technology Education*, 22(1), em2760.
- Zhang, T., Cao, Z., Li, W., & Lv, Z. (2026). Six artificial intelligence innovation strategies applied to autism spectrum disorder research: A narrative review. *Pediatric Investigation*.
- Zhu, Y., Shao, B., & Zhu, S. (2026). Igniting interaction: decoding the impact of address formality on consumer brand engagement in online service agents. *Asia Pacific Journal of Marketing and Logistics*, 1-24.