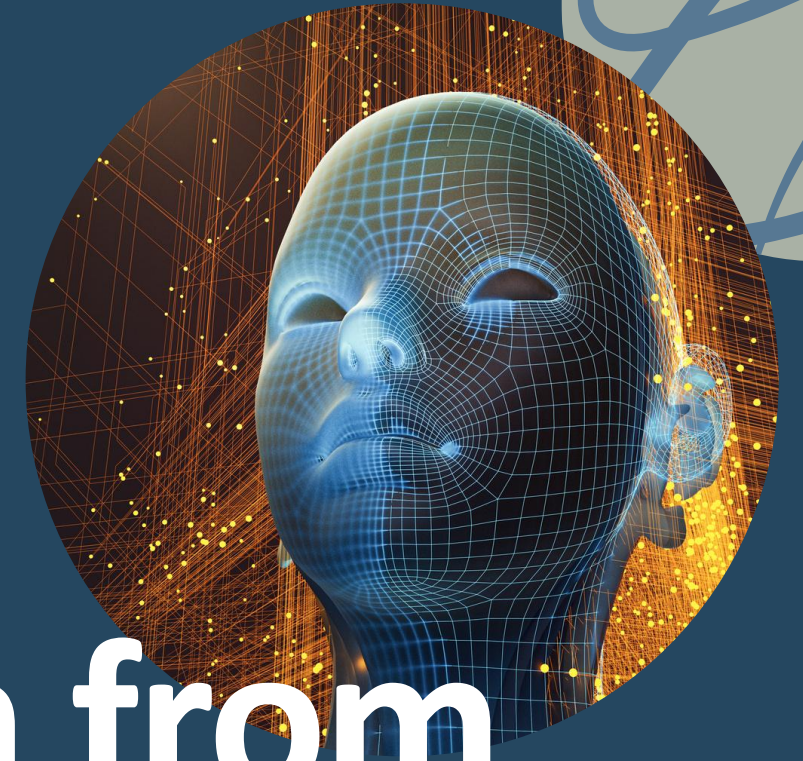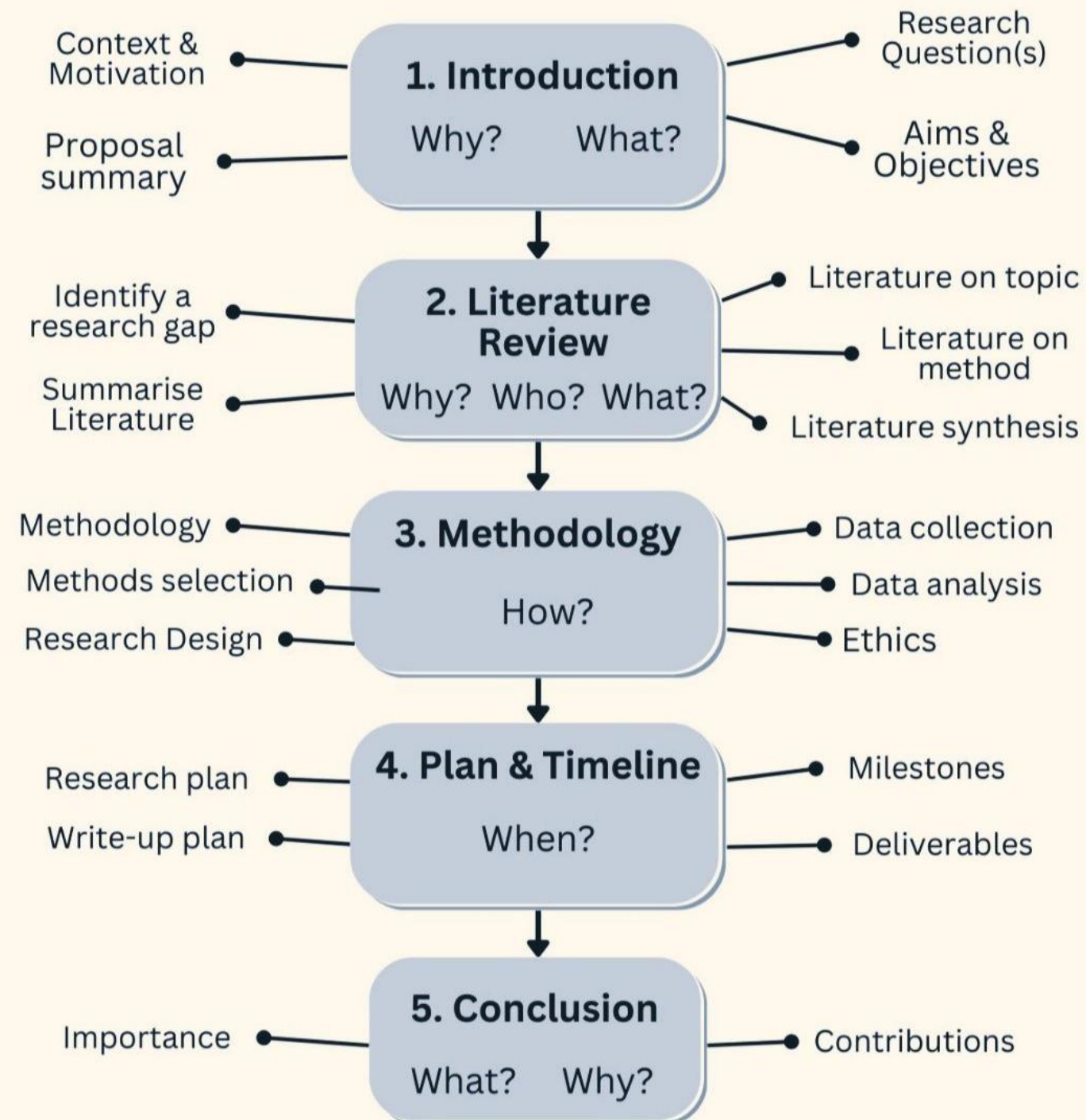# Emotion Recognition from Multimodal Media Content

Presented By:

Md.Mahfujul Karim Sheikh

Prerona Tanjin Kotha

Full Roadmap

1. Introduction
Why?   What?

- Context & Motivation
- Proposal summary
- Research Question(s)
- Aims & Objectives

2. Literature Review
Why? Who? What?

- Identify a research gap
- Summarise Literature
- Literature on topic
- Literature on method
- Literature synthesis

3. Methodology
How?

- Methodology
- Methods selection
- Research Design
- Data collection
- Data analysis
- Ethics

4. Plan & Timeline
When?

- Research plan
- Write-up plan
- Milestones
- Deliverables

5. Conclusion
What?   Why?

- Importance
- Contributions

Motivation

Importance of Emotion Recognition

Emotion recognition is **vital** in enhancing human-computer interaction and understanding behavioral dynamics. It allows for improved communication and tailored experiences in various applications such as healthcare, marketing, and education.

Benefits Across Disciplines

Recognizing emotions can transform sectors like **customer service** and mental health support. By accurately interpreting emotional states, businesses and professionals can better address needs and improve overall satisfaction and

The Role of Technology

Advanced technologies, including deep learning, have enabled more accurate and efficient emotion recognition. These tools help analyze **multimodal data**, ensuring a comprehensive understanding of emotional expressions.

Literature Review: Historical Context of ER

Emotion recognition research has a long history, predating the recent surge in AI advancements.

ⓘ Early Efforts

Relied on traditional machine learning algorithms, emphasizing manual feature extraction from modalities like facial expressions, speech patterns, and physiological signals.

⚠ Limitations

These methods were often constrained by human biases in feature engineering and a limited feature space, hindering comprehensive emotional understanding.

Literature Review: Categorization - Unimodal Approaches

Facial Expression Recognition

Primary nonverbal communication (55% of communication). Pioneering work by Ekman et al. on six basic emotions. Early methods include Gabor wavelets, LBP, WLD.

Speech Emotion Recognition

Analyzes vocal characteristics such as pitch, amplitude, and spectral tilt to infer emotional states from spoken language.

Physiological Signals

Objective but less overt. Includes analysis of heart rate variability, skin conductance, and brain activity (e.g., EEG readings).

⊗ **Overall Unimodal Limitation:** Struggles to accurately judge emotion due to the complex interplay of cues and vulnerability to noise. Capturing the full spectrum of human emotion requires more.

Literature Review: Categorization – Multimodal Approaches

Recognition: Human emotions are expressed through a rich combination of signals, making unimodal approaches insufficient.

Goal of MER: Integrate information from sound, language, images, and physiological data for comprehensive and accurate understanding of human emotion.

Combinations: Common integrations include facial expressions + audio, audio + written text, and physiological signals + visual cues.

Crucial for: Achieving a complete and nuanced comprehension of human emotions in real-world scenarios.

Unimodal   vs.   Multimodal Datasets

LibriSpeech
TIMIT
ESC-50
UrbanSound8
K
VoxCeleb

MNIST
CIFAR-100
ImageNet
Fashion-
MNIST
COIL-20

Speech Recognition

Autonomous Driving

Basic Image Classificatio n

Sentiment Analysis from Text and Audio

IEMOCAP
CMU-MOSEI
SEED
DEAP
EIMT16
MELD
EXTENDED
COGNIMUSE
EXTENDTED
MUSIC VIDEO
DATASET

Technical Overview: Feature Extraction - Textual & Visual

Textual Modality

Capturing semantic and contextual information from dialogue is crucial for understanding underlying emotions.

Word Embeddings

- Pre-trained language models like ALBERT-base.
- Captures token-level meaning.

Sentence Embeddings

- GPT for comprehensive contextual understanding.
- Considers dialogue history, current sentence, and reply.

Visual Modality (Video Frames/Facial Expressions)

Visual cues from video frames and facial expressions offer critical information for emotion recognition.

Frame Analysis

- CNNs (VGG19) for raw video frames and Mel spectrograms.
- ResNet-50, RGB-stream I3D for spatio-temporal features.

Facial & Motion Analysis

- MTCNN for robust face detection.
- RNNs/GRUs for temporal relationships in expressions.
- FlowNetS for subtle motion features.

Technical Overview: Deep Learning Architectures

CNNs
Cornerstone for grid-like data (visual, acoustic, EEG), learning hierarchical spatial features effectively.

RNNs & LSTMs
Adept at sequential data, capturing temporal information from speech and video sequences.

Transformers
Revolutionized sequence-to-sequence tasks with multi-head self-attention for temporal and cross-modal fusion.
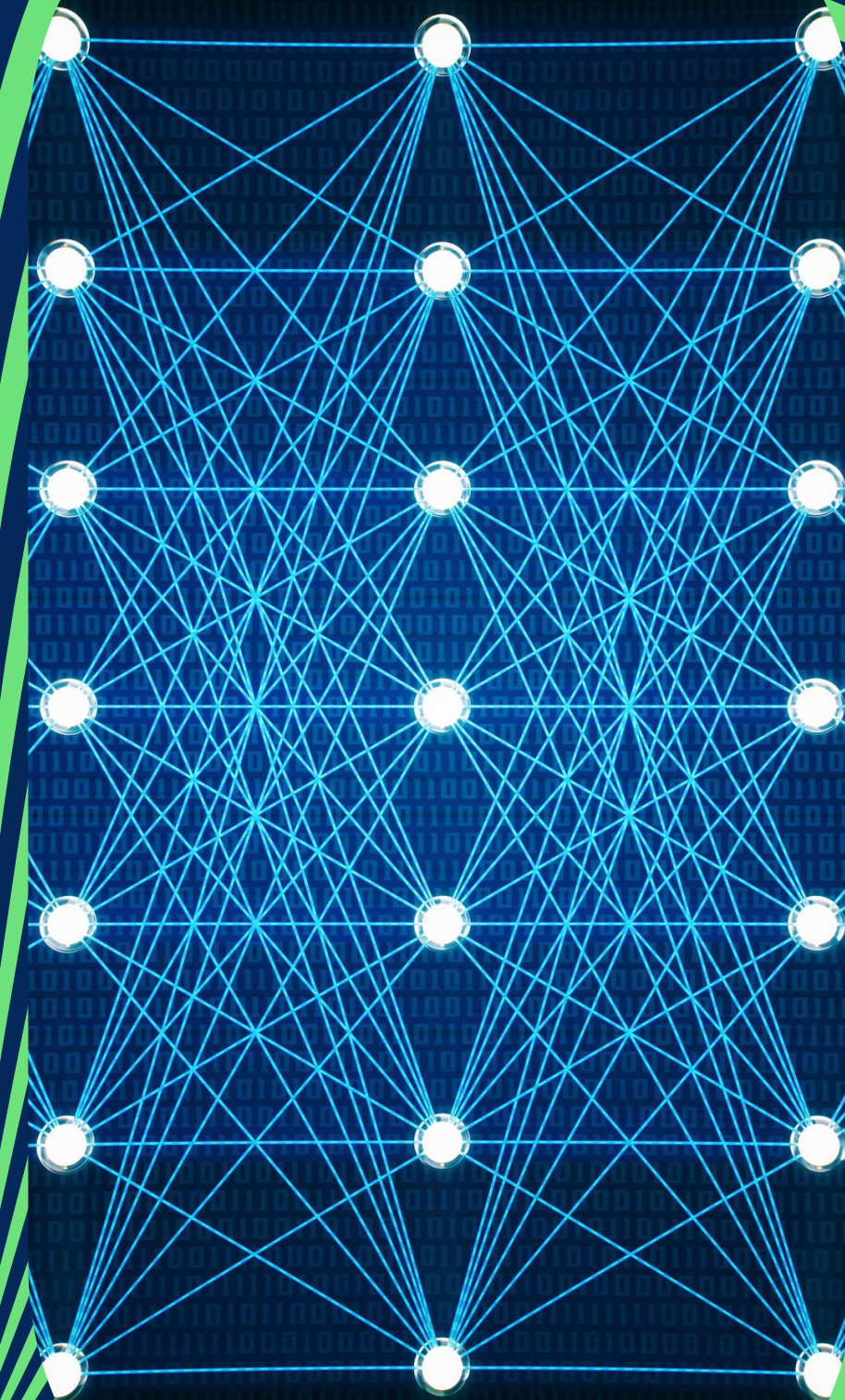
Autoencoders
Unsupervised learning for efficient data representations and shared features from physiological signals.
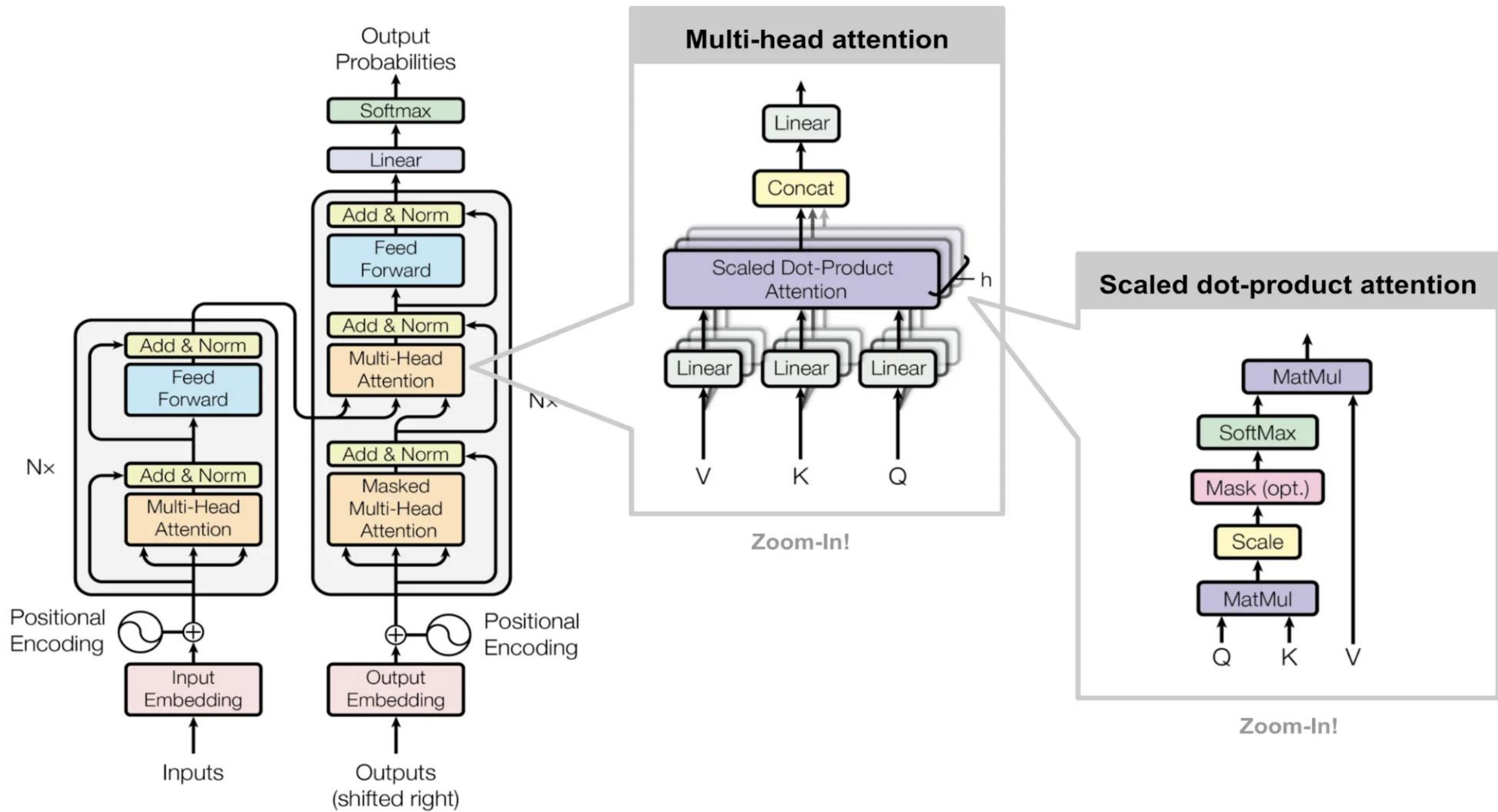
Hybrid Models
Combine CNNs, RNNs, and Transformers for robust feature extraction and integration.

These architectures form the backbone of advanced MER systems, each contributing unique strengths to processing diverse data types.
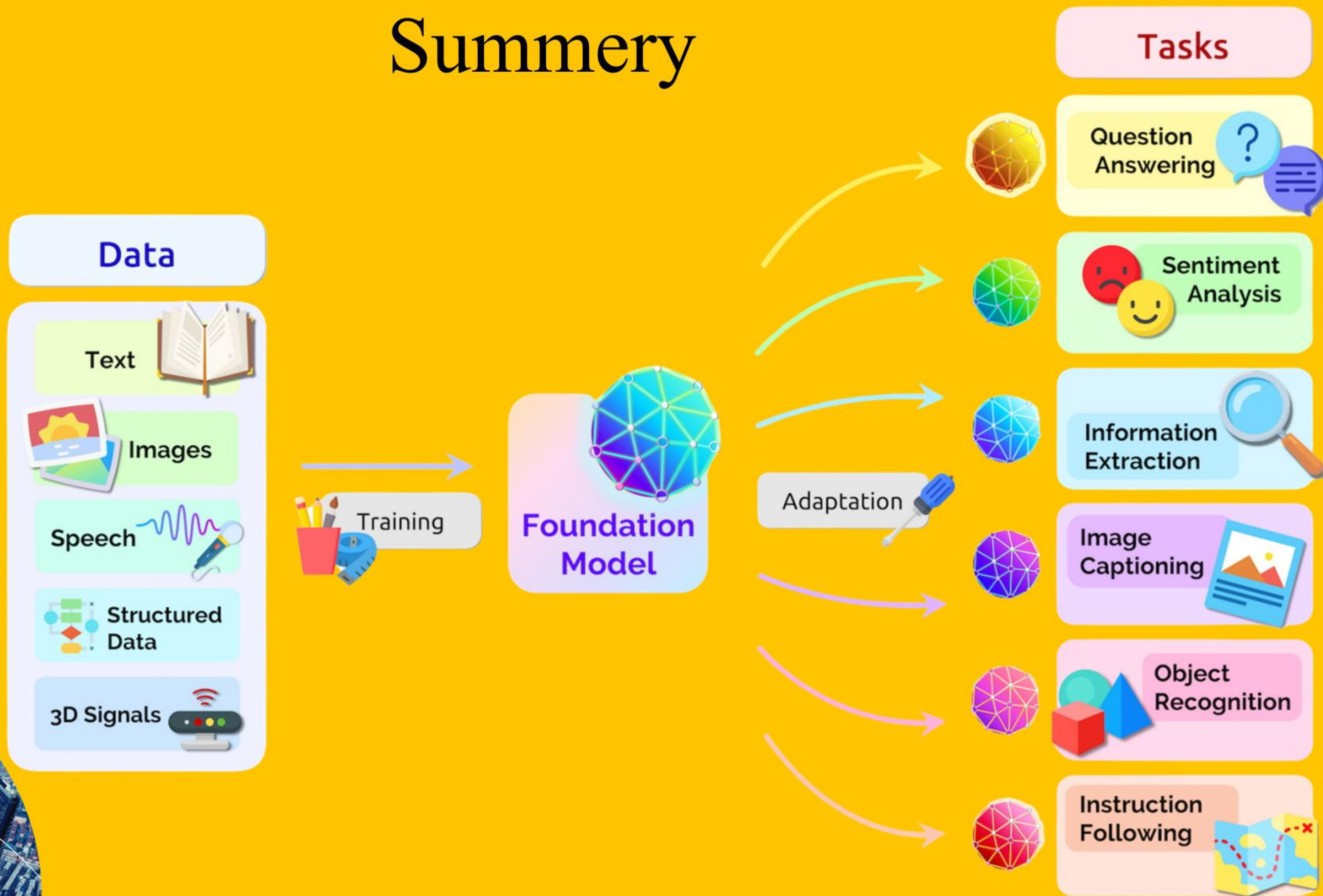
Summery

Comparative Study of Multimodal Emotion Recognition Systems

Recent advancements highlight the diverse approaches and performance metrics across different datasets and modalities.

Le et al. (2023)	Video, Audio, Text + Transformer	85.9%	IEMOCAP
Xiong et al. (2025)	EEG + CNN-KAN-F2CA	97.9%	SEED
Xie et al. (2021)	Conversational MER + EmbraceNet	65%	MELD
Zhang et al. (2024)	Speech, Text + Gated Multimodal Unit	78.2%	MOSEI
Chen et al. (2022)	Facial, Audio + Attention Fusion	81.5%	MSP-IMPROV

Performance varies significantly based on the chosen modalities, fusion architecture, and the emotional complexity and annotation quality of the dataset. EEG-based methods often show higher accuracy for valence/arousal due to direct physiological signals.

Key Applications of Emotion Recognition

Mental Health Monitoring

Emotion recognition technology can significantly aid in **mental health assessments** by detecting emotional states, enabling timely interventions and personalized therapies for patients.

Customer Experience Enhancement

Businesses utilize emotion recognition to analyze **customer reactions**, improving service quality and tailoring experiences, which leads to higher satisfaction and loyalty rates.

Content Personalization

Emotion analysis helps content creators deliver **personalized media** experiences, ensuring that users receive engaging and relevant content based on their emotional responses and preferences.
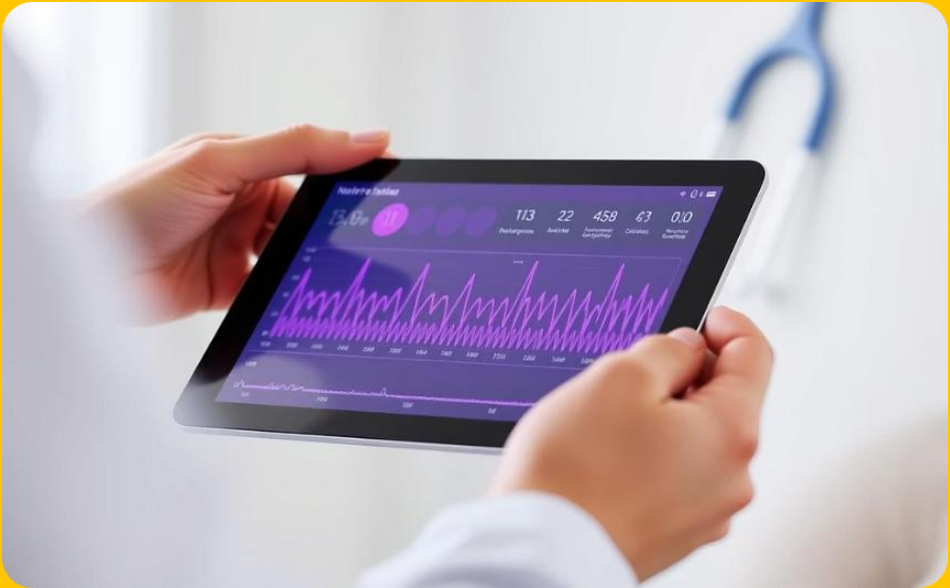
Applications of MER

Robotics & Intelligent Systems

Enabling Socially Assistive Robots (SARs) to interpret human affect and respond emotionally, fostering natural human-robot communication.

Healthcare & Well-being

Monitoring emotional status for mental health and general well-being, aiding in earlier intervention and personalized care.

Education & Learning

Developing adaptive learning systems that respond to student emotions, optimizing educational pathways and improving engagement.

Security & Safety

Implementing smart protection systems, driver monitoring for safety, and lie detection capabilities for enhanced security control.
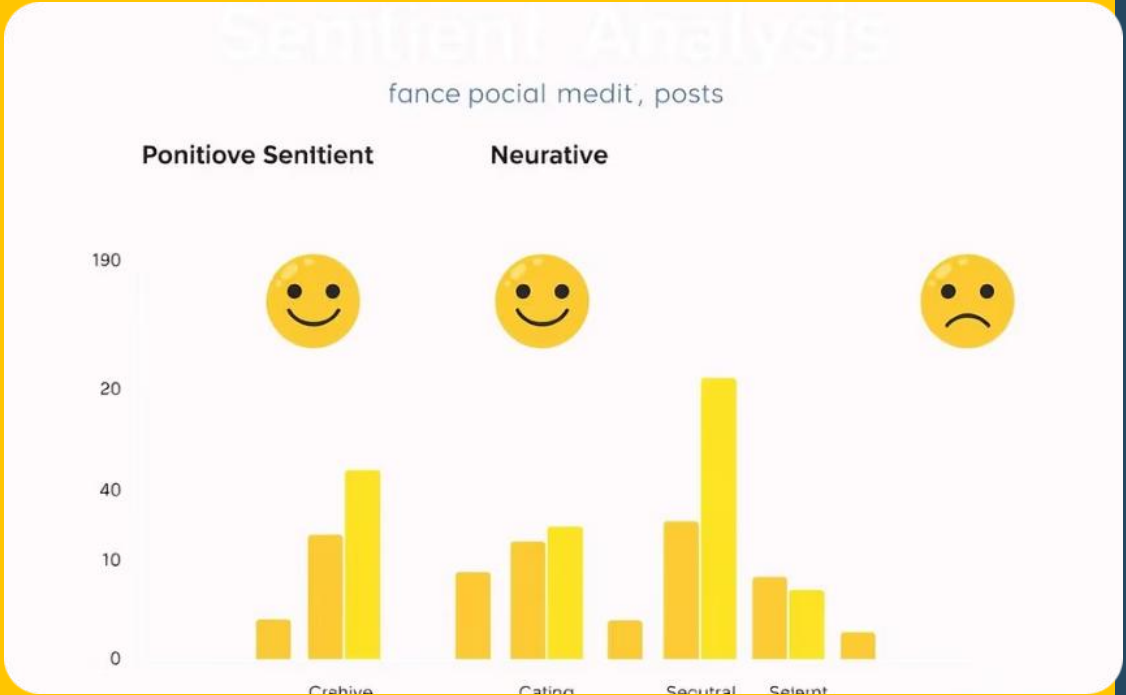
Social Media Analysis

Analyzing user-generated video content and social networking data for sentiment and emotion, enhancing content creation and recommendation systems.

Challenges in Emotion Recognition

This section explores key obstacles in emotion recognition technology.

Data Privacy Concerns
Balancing **user privacy** with data collection is critical, as sensitive emotional data raises ethical issues for individuals and organizations.

Variability in Emotional Expression
Individuals express emotions **differently**, influenced by cultural backgrounds, personal experiences, and contextual factors, complicating accurate emotion recognition.

Limited Contextual Understanding
Current models often struggle to **comprehend context**, which is essential for interpreting emotions accurately in diverse scenarios and interactions.

Technological Limitations
Existing algorithms may lack the necessary **robustness and adaptability** to handle the complexity of multimodal data in real-world applications.

Research Gaps and Future Directions in Emotion Recognition

Despite advancements in **emotion recognition**, several research gaps remain. Key areas for exploration include enhancing multimodal fusion techniques, improving data diversity, and tackling ethical implications in AI use, which can drive more robust and fair systems in future applications.

Key Takeaways & Future Directions

Key Takeaways

- Multimodal emotion recognition significantly outperforms unimodal approaches by capturing richer contextual information.

- Deep learning models, especially Transformers, are crucial for complex feature extraction and robust fusion across modalities.

- Applications span diverse fields, from enhancing HCI to critical roles in healthcare and security.

Future Research Directions

- **Ethical AI:** Addressing privacy concerns and bias in data for fair and responsible systems.

- **Real-time Processing:** Developing low-latency models for immediate emotional response in dynamic environments.

- **Cross-cultural Recognition:** Improving generalization across diverse cultural expressions of emotion.

- **Explainable AI (XAI):** Enhancing transparency and interpretability of model decisions in emotion recognition.

The field continues to evolve rapidly, promising more nuanced and context-aware systems for understanding the complexities of human emotion.

Questions and Feedback?

Thank you for your attention! Please share your thoughts or ask any questions.

Email
Mahfujul_cse735@nubtkhulna.ac.bd

Social Media
www.bio.link/mahfuj735

Phone
+8801726725791