Myasthenia gravis is a chronic life-threatening autoimmune neuromuscular disorder characterized by varying degrees of skeletal muscle weakness. Voice-based algorithms could serve as a more accessible, cost-effective, and noninvasive screening tool for gMG crisis prediction and patient monitoring.

gMG crisis prediction and disease assessment through voice AI model on a wearable device.

Voice digital biomarker for gMG crisis prediction and patient monitoring.

Javier Díez Pérez

gMG crisis prediction and disease assessment through voice AI model on a wearable device

Author: Javier Díez Pérez

Business Model Proposal: Artificial Intelligence for Crisis Diagnosis in Myasthenia Gravis

1. Introduction:

Myasthenia gravis is a chronic life-threatening autoimmune neuromuscular disorder characterized by varying degrees of skeletal muscle weakness. The disease frequently affects muscles that control eye and eyelid movement, facial expression, and swallowing [1].

Symptom fluctuations can complicate the assessment of a patient's overall clinical status. This variability necessitates close and personalized monitoring to respond swiftly in case of deterioration. It's important to implement a patient-centered care approach with regular evaluations and adaptable treatment plans to effectively manage these fluctuations and enhance patient outcomes.

A myasthenic crisis occurs when these muscles weaken to the point that ventilation is inadequate. Most myasthenia **gravis crisis (MCs)** don't occur suddenly, allowing for early detection and management of triggers and contributing factors. This window of opportunity is crucial because respiratory failure, which is a hallmark of MC, can be life-threatening. Early and accurate diagnosis of these crises can substantially impact patient outcomes, enabling timely interventions.

Additionally, patients with **generalized myasthenia gravis** (gMG) often feel overwhelmed by the burden of their condition, leading to lowered mood and increased stress, both of which can adversely impact disease progression and prognosis. Implementing a passive assessment system could help diminish their perception of disease burden, decrease stress (a key trigger for gMG crises), and enable them to lead a more independent life.

Assessing respiratory function in patients with myasthenia gravis (MG) presents challenges [2], particularly in predicting or detecting the **imminence of a myasthenic crisis** (ICM), which lacks clear diagnostic markers and is currently reliant on the clinician's judgment [3, 7, 10].

1. **Assessment Challenges for ICM**:
   - Existing scales like MG-ADL and QMG have limitations in detecting ICM.
   - Dyspnea, evaluated in MG-ADL, does not always precede a myasthenic crisis.
   - Forced vital capacity (FVC) used in QMG requires specific instruments and time.
2. **Tools and Tests**:
   - **Inspiratory Pressures (PIM, PEM)**: Not commonly used and unclear interpretation in MG.

- **Peak Expiratory Flow**: Used in asthmatic patients; may not indicate ICM due to factors like poor lip closure.

3. **Simple Methods**:
    - A practical, apparatus-free method is counting numbers out loud after a deep inspiration. Being unable to count to 20 may indicate a poor respiratory prognosis, as documented in other neuromuscular diseases like Guillain-Barré syndrome.

Voice synthesis is a complex process that relies on the combined effects of the respiratory system, the nervous system, and the larynx. Anything that affects these systems can influence the voice, whether it is perceptible audibly or detectable through computer analysis.

Breathing/Speech AI can play a significant role in predicting myasthenia gravis (MG) crises by monitoring and analyzing changes in voice patterns that might indicate a decline in neuromuscular performance.

Recent studies have investigated the use of voice/speech as an accessible, straightforward and reasonably affordable, modality for detecting and predicting the risk of various conditions, extending beyond respiratory or infectious diseases to include conditions like hypertension and diabetes [5, 6, 14, 16]. AI can automatically detect subtle changes in sound signals that reflect physio-pathological changes in the vocal cords, which in turn affect voice characteristics. International voice research consortiums like Bridge2AI [17] and Colive Voice [16] aims to integrate the use of voice as biomarker of health in clinical care to assist in screening, diagnosis, and treatment of a broad range of diseases.

On the other hand, different advances to build sound and multimodal foundational models currently allows training, fine-tuning for different use cases. Namely, Google has recently released an open foundational model, Health Acoustics Representations (HeAR) [4], that generates embeddings from health acoustic data. These embeddings facilitate the efficient development of AI models for health-related acoustic tasks, such as identifying disease status from cough sounds or assessing lung function using exhalation sounds from spirometry. This approach requires less data and computational resources compared to training a model from scratch without these embeddings or a pretrained model. HeAR has been trained on over 300 million two-second audio clips, capturing five types of non-speech health acoustic events: coughing, breathing, throat clearing, laughing, and speaking.

These capabilities could provide healthcare professionals with valuable insights and enable early interventions to manage or prevent myasthenic crises, thereby improving patient outcomes. Integrating these voice AI systems into regular patient monitoring routines [13] could serve as a non-invasive and accessible tool for the ongoing management of myasthenia gravis (MG) [7, 8, 9].

2. Objective:

The objective is to develop an innovative, non-invasive artificial intelligence (AI) model that can efficiently analyze patient voice data to identify early signs of myasthenic crises. This approach aims to complement existing diagnostic methods, providing a quicker and potentially more accessible solution for gMG patients.

The edge AI model will be integrated into a wearable device, enabling passive and continuous monitoring of the status of gMG patients. This approach offers patients greater autonomy while reducing the stress associated with frequent clinical visits and the uncertainty they can bring. Such a solution can be later used or applied in various contexts; clinical trials, clinical routine, crisis management etc. …

3. Methodology:

Study Preparation: We will engage in ethical collaborations with hospitals and medical research institutions to collect diverse datasets of voice recordings from myasthenia gravis patients. It's crucial to obtain data both during crisis events and stable phases to train a robust model. The goal of the study preparation phase is to build a rigorous Study Protocol, which is composed of Administrative Information section, containing all relevant administrative details and necessary information, a Methods Section which provides a comprehensive description of the procedures used to set up the voice database, and Ethics and Dissemination section, where the plans for obtaining ethic approval and the dissemination strategy. The second step of the Study Preparation phase will detail the Recruitment of Subjects and the Collection of Informed Consent from each of them. [See appendix on HIPAA regulations of biometrics data - See B2AI-Voice project learnings and developments.]

Data Collection: During this step, each participant undegoes a detailed clinical evaluation, which included collecting demographic and medical history information and, for gMG patients, some medical tests to assess the severity of the disease (detail the type of assessment to be used). Demographic data such as gender and date of birth for each participant is recorded, and, only for healthy controls, we will collect medical history data such as the presence of any diseases potentially affecting speech.The outcome will be a multimodal clinical knowledge base. Standardisation and interoperability isn't mature in this field, what also give us ground to lead innovation. [Citation and appendix].

Data Analysis: Utilizing state-of-the-art machine learning algorithms, particularly deep learning models capable of handling complex data such as deep neural networks, will allow us to pinpoint anomalies in voice patterns linked to muscle weakness.

Model Development: The AI model will be iteratively trained and validated to ensure high accuracy and generalizability across different patient demographics. This process will involve training, testing, and tuning the model parameters based on performance metrics such as accuracy, sensitivity, and specificity.

Device development:

- Wearable device features:
    - Wearable: would allow for data collection (voice)[1].
    - AI model (smallest size for desired performance).
    - Chip: speed and energy efficiency.
    - Battery
    - Microphone
    - GPS
    - Connections: data transfer + alert call.
- Home dock station features: would be used to display information, and allow external data sharing.
    - Requires data annonymization automatic and encrypted transmision of data. Standard compliance with healthcare regulations will be guarranteed (HIPPA, ISO).
    - User friendly GUI.
    - Long-life battery.
    - Different connection options for data transfer and alert.
    - Compatible with reference hospitals.

4. Technological Innovation:

The proposed AI solution for gMG crisis early prediction through speech and breathing as digital biomarkers, offers a new level of precision in diagnostics, like innovations in other medical fields where AI has been used effectively. For example, AI models have been used for early detection of Parkinson's disease through voice analysis, illustrating the potential of this technology in neurology. These precedents underscore the feasibility and transformative potential of our proposed model and device.

5. Expected Impact:

- **Improved Diagnosis:** By offering a real-time, accessible tool for crisis prediction, healthcare providers can adopt a more proactive approach to managing myasthenia gravis, reducing hospital stays and associated healthcare costs.
- **Cost Reduction:** AI-driven diagnostics can significantly reduce reliance on expensive and time-consuming tests currently necessary for managing the condition.
- **Accessibility:** The solution will be designed for deployment on common digital platforms, including mobile and web applications, ensuring broad accessibility for patients and clinicians.

[1] Potential wearables include necklace-shape microphone (similar use case in cough monitoring), wrist-worn devices with embedded microphones, similar.

6. Implementation Plan:

- **Phase 1:** Initial research and development will focus on data collection and model training (fine-tuning), anticipated to last 12-18 months. This phase will also involve securing ethical approvals and partnerships.
- **Phase 2:** Integration and pilot testing will occur in selected healthcare centers. Feedback from clinicians and patients during this phase will be critical to refine the user interface and improve model accuracy.
- **Phase 3:** Full-scale commercialization will target both public and private healthcare sectors. Initial focus will be on developed regions with high healthcare technology adoption rates, with expansion plans following in emerging markets.

7. Monetization Strategy:

- **Licensing Sales:** Licensing the AI technology to hospital networks and specialty clinics will provide a steady revenue stream. Models like IBM Watson for Oncology, which sells licenses to medical institutions, serve as a successful precedent. [11, 12]
- **Subscription Model:** Offering a subscription service for continual access and updates can ensure a recurring revenue model. Similar strategies are employed by companies like MedTech firms using SaaS platforms for diagnostics. Suscription could be coupled to drug administration / follow up.
- **Consulting Services:** Providing tailored consulting and customization options for large healthcare providers can expand revenue opportunities, akin to how AI diagnostics companies often tailor algorithms for different medical conditions and/or their use in clinical trials / RWE studies.

8. Conclusion:

This proposal envisions leveraging cutting-edge AI technology to create a first-of-its-kind diagnostic tool for myasthenia gravis crises. By addressing a clear gap in the current healthcare landscape, this tool can significantly enhance patient care and optimize resource allocation in neurology departments worldwide.

Voice may offer a proxy for detailed time-series data only found in high-resource areas, while simultaneously providing voice, speech, and respiratory data to compliment patient-reported information. Ultimately, AI models trained on voice may be used in the clinic and home, supporting patients in hospital "deserts" where healthcare is not readily accessible.

We also envision the impact of the developments required to successfully implement such device will have in improving patient's quality of life in other rare disorders (MS, neuromuscular, mitochondrial disorders, …) key therapeutic areas, such immunology, rheumatology, respiratory, cardiovascular, oncology, pediatrics or surgery.

9. References:

[1] P. Narayanaswami *et al.*, 'International Consensus Guidance for Management of Myasthenia Gravis: 2020 Update', *Neurology*, vol. 96, no. 3, pp. 114–122, Jan. 2021, doi: 10.1212/WNL.0000000000011124.

[2] M. Zimmer, B. Kleiser, J. Marquetand, and F. Ates, 'Characterization of Muscle Weakness Due to Myasthenia Gravis Using Shear Wave Elastography'.

[3] H. Wiendl *et al.*, 'Guideline for the management of myasthenic syndromes', vol. 16.

[4] S. Baur *et al.*, 'HeAR -- Health Acoustic Representations', Mar. 04, 2024, *arXiv*: arXiv:2403.02522. doi: 10.48550/arXiv.2403.02522.

[5] J. M. Kaufman, 'Acoustic Analysis and Prediction of Type 2 Diabetes Mellitus Using Smartphone-Recorded Voice Segments', *MAYO CLINIC PROCEEDINGS*.

[6] B. Taghibeyglou, J. M. Kaufman, and Y. Fossat, 'Machine Learning-Enabled Hypertension Screening Through Acoustical Speech Analysis: Model Development and Validation', *Machine Learning*.

[7] D. E. C. Vicente, 'Miastenia Gravis: Pautas de Actuación y Seguimiento'.

[8] P. Narayanaswami et al., 'International Consensus Guidance for Management of Myasthenia Gravis: 2020 Update', Neurology, vol. 96, no. 3, pp. 114–122, Jan. 2021, doi: 10.1212/WNL.0000000000011124.

[9] 'Breathing disorders during sleep in myasthenia gravis'. 1992 by the American Neurological Association

[10] CADTH COMMON DRUG REVIEW Clinical Review Report ECULIZUMAB (SOLIRIS) Alexion Pharma Canada Corporation

[11] I. Kulkov, 'Next-generation business models for artificial intelligence start-ups in the healthcare industry'.

[12] T. Jiang, 'Crucial Business Model Elements for Medical Device Startup Companies in Emerging Markets'.

[13] V. Despotovic, A. Elbeji, K. Funfgeld, M. Pizzimenti, H. Ayadi, and P. V. Nazarov, 'Digital Voice-Based Biomarker for Monitoring Respiratory Quality of Life: Findings from the Colive Voice Study'.

[14] A. Elbéji et al., 'A voice-based algorithm can predict type 2 diabetes status in USA adults: Findings from the Colive Voice study'.

[15] G. Fagherazzi and Y. Bensoussan, 'The Imperative of Voice Data Collection in Clinical Trials'.

[16] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, 'Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice'.

[17] Yael Bensoussan 'Bridge2AI-Voice: An ethically-sourced, diverse voice dataset linked to health information'.

Supplementary material: Business Model Canvas

Business Model Canvas: Business Model Canvas

1. **Customer Segments**:
    - Patients at risk or diagnosed with gMG.
    - Other potential customer segments (not targeted initially).
        1. Neurologists and healthcare professionals specializing in neuromuscular disorders.
        2. Hospitals and clinics providing neurological care.
        3. Research institutions and universities.
        4. Health insurance companies.
2. **Value Propositions**:
    - Early detection of gMG, leading to improved patient outcomes and treatment plans.
    - Non-invasive and easy-to-use digital biomarker analysis.
    - Cost-effective alternative compared to traditional diagnostic methods.
    - Enhances precision in diagnostic practices and reduces diagnostic errors.
    - Reduction of uncertainty and stress: better patient quality of life and autonomy.
3. **Channels**:
    - Partnerships with hospitals and clinics for direct inclusion in diagnostic procedures.
    - Collaborations with research institutions for further validation and promotion.
    - Online platforms and telemedicine services.
    - Medical conferences and professional healthcare networks.
4. **Customer Relationships**:
    - Develop trust through validation studies and clinical trials.
    - Provide training and support for healthcare professionals.
    - Offer premium, personalized diagnostic solutions.
    - Community engagement and patient advocacy groups to increase awareness.
5. **Revenue Streams**:
    - Subscription-based model for healthcare institutions.
    - One-time purchase/license for devices and software.
    - Consultation and training services.
    - Data and analytics services for research purposes.
6. **Key Resources**:
    - Expertise in AI and machine learning.
    - Partnerships with healthcare providers and researchers.
    - Accurate speech and breathing analysis algorithms.
    - Regulatory and compliance expertise.

7. **Key Activities**:
   - Research and development for improving AI algorithms.
   - Clinical trials and validation studies.
   - Building partnerships with healthcare providers.
   - Continuous software updates and customer support.
8. **Key Partnerships**:
   - Patient associations and advocacy groups.
   - Collaborations with neurology experts and medical institutions.
   - Alliances with biotechnology and medical device companies.
   - Partnerships with regulatory bodies for compliance.
   - Health IT platforms and telehealth providers.
9. **Cost Structure**:
   - R&D costs for algorithm and device development.
   - Marketing and partnership development expenses.
   - Regulatory compliance and approval processes.
   - Customer support and training services.

Value Proposition Canvas:

1. **Customer Jobs**:
   - Timely and accurate diagnosis of gMG.
   - Reduce the burden of invasive traditional diagnostic processes.
   - Enhance monitoring and treatment planning for patients.
2. **Pains**:
   - Diagnostic delays and inaccuracies in current practices.
   - High costs of traditional diagnostic methods.
   - Patients' discomfort with current diagnostic procedures.
3. **Gains**:
   - Early detection enabling proactive treatment strategies.
   - Non-invasive, user-friendly diagnostic process.
   - Improved patient satisfaction and health outcomes.
4. **Products & Services**:
   - AI-based diagnostic tool for speech and breathing analysis.
   - Companion mobile apps for easy data collection and analysis.
   - Integration with existing Electronic Health Records (EHR) systems.
5. **Pain Relievers**:
   - Provides a more accurate, non-invasive diagnostic method.
   - Reduces costs associated with prolonged or incorrect treatments.
   - Increases the speed and efficiency of the diagnostic process.
6. **Gain Creators**:
   - Enables early intervention strategies.
   - Supports healthcare professionals with advanced diagnostic tools.
   - Elevates the standard of care for neuromuscular conditions.

Functional Jobs

1. **Patients**:

- **Assessment**: Allow patients to undergo non-invasive, quick assessments without needing frequent hospital visits.
- **Feedback**: Receive timely updates on their health status, aiding in better personal health management.
- **Adherence**: Facilitate better adherence to treatment through regular monitoring and feedback.

Emotional Jobs

1. **Patients**:
   - **Peace of Mind**: Achieve greater peace of mind due to early detection, knowing that their condition can be managed proactively.
   - **Empowerment**: Feel empowered to take control of their health with more information and less reliance on invasive tests.
   - **Reduced Anxiety**: Decrease anxiety around diagnostic procedures and potential health outcomes.

APPENDINX I: Myasthenic Crisis

A myasthenic crisis is defined as a worsening of myasthenia gravis (MG) weakness that requires intubation or non-invasive mechanical ventilation (NIMV). It represents the most severe category (Category V) on the Myasthenia Gravis Foundation of America (MGFA) scale and is considered a potentially life-threatening situation. It usually manifests as a generalized worsening of the patient's muscle weakness, followed by a proportional onset and/or worsening of bulbar symptoms and respiratory muscle weakness. However, in up to 5% of cases, respiratory muscle weakness is the first manifestation. Respiratory muscle weakness is responsible for the respiratory insufficiency that necessitates ventilatory support, often accompanied by significant bulbar/oropharyngeal muscle weakness, which increases the risk of airway obstruction and aspiration, making orotracheal intubation necessary. Approximately 10-20% of patients with MG will experience at least one myasthenic crisis during their lifetime, primarily in the early years following diagnosis, and it can be the initial presentation of MG (up to 20% of myasthenic crises are the first manifestation of the disease).

There are multiple precipitating factors for a myasthenic crisis: concomitant infections (the most common factor), adjustment/reduction of immunomodulatory/immunosuppressive treatments, surgeries, pregnancy, childbirth, certain medications, etc. However, a myasthenic crisis can also occur spontaneously within the natural course of the disease. At the start of treatment with anticholinesterase agents, a severe adverse effect has been described, the cholinergic crisis, which causes a paradoxical worsening of muscle weakness and respiratory failure. However, this is an uncommon side effect, especially when the dosage is gradually increased and kept below a total of 960 mg/day. Therefore, in the presence of these symptoms, a myasthenic crisis should always be suspected first, and respiratory function should be closely monitored. We must also remember that a myasthenic crisis can be the initial presentation of the disease and should therefore be included in the differential diagnosis of neuromuscular respiratory failure.

APPENDIX II: How speech AI could assist gMG patients.

1. **Voice Pattern Analysis**:
    - Speech AI can analyze changes in voice pitch, volume, and clarity. Any significant alterations might indicate muscle weakness associated with MG.
2. **Speech Rate Monitoring**:
    - Consistently slowed or irregular speech rate might suggest fatigue or muscle weakness, potentially indicating an impending crisis.
3. **Prosody and Articulation**:
    - The AI could detect changes in the rhythm, stress, and intonation of speech, providing clues about muscular control issues common in MG.
4. **Real-time Monitoring**:
    - Implementing real-time speech monitoring allows for continuous tracking of patients' speech, enabling early detection of potential exacerbations.
5. **Predictive Modeling**:
    - Using machine learning, speech AI can analyze historical speech data to identify patterns or predictors of a crisis, enhancing preventive care.
6. **Integration with Other Biomarkers**:
    - Combining speech analysis with other data, such as respiratory function or physical activity, could enhance prediction accuracy for MG crises.
7. **Patient Engagement**:
    - Simple tasks like counting aloud could be monitored by AI to assess neuromuscular function without requiring professional medical equipment.

APPENDIX III: EDGE AI

Edge AI in audio processing refers to performing artificial intelligence tasks on audio directly on the device itself (the "edge") rather than sending it to a cloud server for processing. This offers advantages like real-time processing, reduced latency, enhanced privacy, and lower bandwidth usage.
Elaboration:

- **Real-time processing:**

  Edge AI allows for immediate analysis and classification of audio events, which is crucial for applications like voice assistants and emergency detection systems.

- **Reduced latency:**

  By processing data locally, the delay associated with sending and receiving data over a network is eliminated, leading to faster responses.

- **Enhanced privacy:**

Keeping audio data on the device reduces the risk of data breaches and protects sensitive information.

- **Lower bandwidth usage:**

  Processing locally minimizes the amount of data transmitted, which is beneficial in areas with limited bandwidth or unstable connectivity.

- **Operational independence:**

  Edge AI enables devices to function without a continuous internet connection, ensuring reliable performance in areas where connectivity is unreliable.

- **Efficient resource utilization:**
  Leveraging the device's processing power can lead to cost savings by reducing the need for cloud infrastructure.

APPENDIX IV: Iatrogenic MG, secondary to treatment with ICI in cancer immunotherapy.

In the context of immune-checkpoint inhibitor therapy, which is used in cancer treatment to enhance the immune system's ability to fight cancer cells, there is relevance to note:

Mechanism: Immune-checkpoint inhibitors, such as PD-1, PD-L1, and CTLA-4 inhibitors, can potentially lead to the development or exacerbation of autoimmune conditions, including myasthenia gravis. This happens because these therapies enhance immune activity, which can sometimes lead to a loss of self-tolerance and increase the risk of autoimmune side effects.

Incidence: The occurrence of myasthenia gravis as a side effect of immune-checkpoint inhibitor therapy is rare but serious. It may result in severe muscle weakness and may be accompanied by other neurological side effects.

Clinical Management: It's crucial to monitor patients undergoing immune-checkpoint inhibitor therapy for symptoms of MG. Early detection and management are vital, which may involve discontinuation of the checkpoint inhibitors and treatment with medications to manage MG symptoms, such as corticosteroids or acetylcholinesterase inhibitors.

Research and Reports: There have been various case reports and ongoing research into the exact mechanisms and management strategies for MG in the context of immune-checkpoint inhibitor therapy. Clinicians are advised to be vigilant for signs of neuromuscular side effects in patients receiving these treatments.

Overall, while myasthenia gravis is a rare side effect of immune-checkpoint inhibitors, it is important for clinicians and patients to be aware of the potential symptoms and seek prompt medical attention if they arise.

APPENDIX V: Multimodal knowledge databases (including sound).

Traditional deep-learning models have achieved amazing results but are limited by their ability to handle only a single modality. Simple architectures like convolutional neural networks (CNNs) are purpose-built for understanding particular data types and cannot capture multimodal contexts for deeper understanding.

Multimodal deep learning is the latest advancement in the AI space that allows models to work with multiple modalities. These models leverage data fusion to simultaneously process text, images and audio. This allows them to improve performance on basic machine-learning tasks like image classification and build advanced applications like text-to-image generators.

Deep learning multimodal architectures are composed of 3 distinct components. These are:

- Unimodal encoders
- Fusion network
- Classifier

The unimodal encoders are multiple standalone encoders built to process particular data types. For example, we will have two encoders for a text-to-image model, one for processing text and the other for images. The next step is to map the individual encodings onto a unified latent space using a fusion network. These networks are the backbone of multimodal processing and are often based on the transformer architectures. The last stage is a classifier, which is trained for downstream tasks, producing the final output.

Encodings are a vector representation of unstructured data like text or images. The encoding stage carries out feature extraction for the provided data. The features are stored as vectors and help the model learn data patterns.

The encoding stage involves multiple unimodal embeddings depending on the modalities involved. For example, for text, we may use popular models like Word2Vec or BERT, and for images, we may use OpenAI’s CLIP model. General models like Data2Vec can also be utilized to process text, video and audio data. Many of these models utilize attention mechanisms to generate informative data representation.

ML engineers often select embedding models based on their benchmark performance and the task at hand.

Once embeddings are created, their knowledge must be combined into a unified space. The embeddings are passed onto the fusion module, where they are combined using the defined techniques. Simpler fusion techniques involve plain concatenation or a weighted sum of the embeddings to form a single unit. However, these techniques do not capture complex cross-modality relationships and are troubled by challenges like uneven dimensions.

Some advanced techniques use attention mechanisms to assign weightage to different parts of the embeddings. Other techniques involve multilayer perceptron (MLP) networks to learn non-linear transformations from the concatenated representations. The technique selection depends on the embedding, dimensionality, modality type and task.

The fusion itself can be performed at different stages. These include:

- Early fusion: This combines the modalities immediately after embedding creation.
- Intermediate fusion: This is a middle ground where features are extracted from each modality to some extent, and then the intermediate representations are fused before further processing. This offers more flexibility and can capture some interactions between modalities.
- Late fusion: This allows each modality to be processed by an independent dedicated network, and the final outputs are combined. It allows for finer feature extraction and better individual results.

The final step of the pipeline is a classification module that uses the fused modalities to make predictions or decisions. It includes a multi-layer neural network trained for a specific task. Depending on the class labels and type of prediction, the layer can include popular functions like sigmoid or softmax. The output of this layer is the final prediction that the model requires. In the case of a classifier, the output will be the class label seen during training.

Multimodal models bring us closer to mimicking human behavior by understanding multiple inputs simultaneously. Their ability to process visual cues and textual cues in the same context allows for a holistic understanding of the world.

Multimodality is a significant leap in AI that enables machines to comprehend and respond to content in a way that resembles human perception. It improves AI robustness and performance for various practical applications.

APPENDIX VI: **HIPAA regulations for audio recording and Biometric Data handling requirements.**

HIPAA requires informed patient consent, proper documentation, and the implementation of security measures to protect recorded health information. Healthcare providers should obtain patient consent before recording any conversation and clearly explain the purpose and nature of the recording. Patients should also be warned against recording other patients without permission.

Healthcare providers should have clearly defined guidelines for audio recording as part of their privacy policies and provide regular HIPAA training to staff. These guidelines should ensure that patient consent is obtained, and protected health information is recorded, stored, and shared according to HIPAA regulations.

Ensuring HIPAA compliant audio recording

Healthcare providers can take proactive steps to ensure HIPAA compliant audio recording:

Obtain proper consent for audio recording: Healthcare providers must get patient consent before recording any conversation. Provide staff with a consent form that clearly explains privacy policies. Using informational signage can also remind patients and staff about the importance of obtaining consent and respecting privacy.

Use secure storage and encryption of audio data: HIPAA requires the use of secure storage and data encryption. Physical storage devices should be housed in secure facilities and protected by passwords and other authentication methods. Data that is digitally stored, sent, or maintained must be encrypted. Sharing an audio recording must be done via HIPAA compliant email or secure file transfer.

Implement access controls and audit trails: Access to recorded audio should be restricted to authorized individuals directly involved in the care of the patient who is the subject of the recording. Any access to audio recordings should be logged in detail through audit trails to track any potential breaches.

Business associate agreements: If using audio software or 3rd-party storage services, a business associate agreement is required.

HIPAA Rules for Biometric Data

How HIPAA Classifies Biometric Data

Under HIPAA, biometric data is considered Protected Health Information (PHI) when it is connected to an individual's health records or used in healthcare services [2]. This includes identifiers such as:

- Fingerprints
- Facial recognition patterns
- Voice patterns

Security Requirement	Implementation Details
Data Encryption	Use industry-standard encryption for data at rest and in transit.
Access Controls	Apply role-based access and multi-factor authentication.
Audit Trails	Maintain logs of all access attempts and system usage.
Secure Storage	Store data on encrypted servers with strict access protocols.

The HIPAA Privacy Rule further requires healthcare providers to:

- Obtain **informed consent** before collecting biometric data.
- Limit data collection to what is **absolutely necessary**.
- Maintain detailed documentation of data collection processes, access logs, staff training, security measures, and incident responses.

To enhance security, healthcare organizations are encouraged to encrypt biometric templates rather than storing raw biometric data. Additionally, they must offer alternative authentication methods for patients who cannot or prefer not to use biometric systems due to physical, personal, or other reasons.