



Politecnico
di Torino

FONDAZIONE
links
PASSION FOR INNOVATION

Multimodal Emotion Recognition Using Large Audio-Language Models

ASHKAN SHAFIEI

S342583

PARASTOO ALAVI

S340942

AMIR MASOUD ALMASI

S337006

PART OF A BIGGER PICTURE



“While AI systems can be used to enhance social skills, they could also affect people’s physical and mental wellbeing.”

European Commission

REAL-WORLD IMPACT & APPLICATIONS



Human-Computer Interaction

Empathetic virtual assistants and intelligent tutoring systems that adapt to user emotional states, providing personalized support and natural conversations.



Healthcare Monitoring

Continuous tracking of patient emotional states enables personalized mental health care, early intervention for depression or anxiety, and improved therapeutic outcomes.



Social Media Analysis

Detecting sentiment shifts and mental health signals from multimodal posts—combining text, voice notes, and images—for community wellbeing and crisis prevention.

PROJECT GOALS

1

Design a multimodal AI system to detect emotion

2

Promote human-centered and ethical ai systems



VALUE PROPOSITIONS



IMPROVED
EMOTIONAL WELL-
BEING



DATA-DRIVEN
HEALTH INSIGHTS



ENHANCED
PATIENT SUPPORT
SYSTEMS

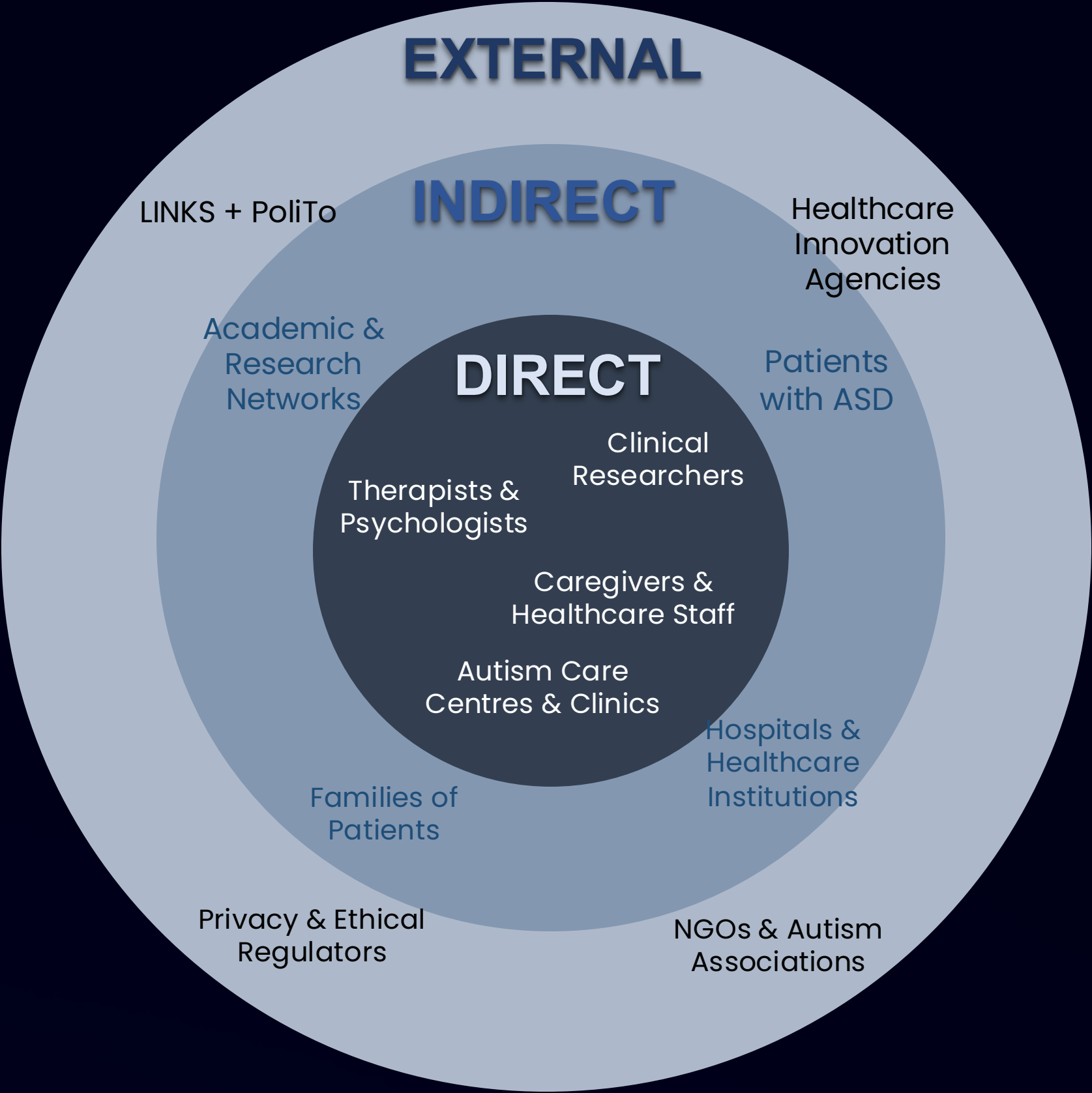


RESPONSIBLE &
HUMAN-CENTERED
AI

SUSTAINABLE DEVELOPMENT GOALS



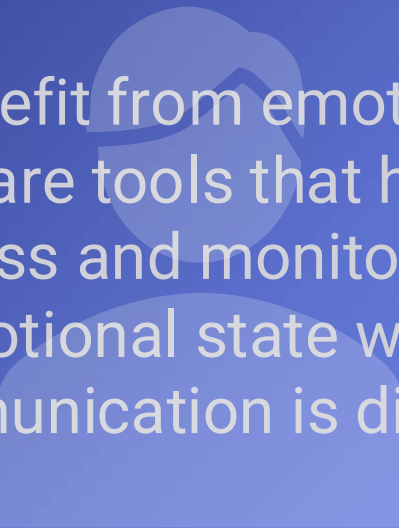
STAKEHOLDERS MAP



PERSONAS

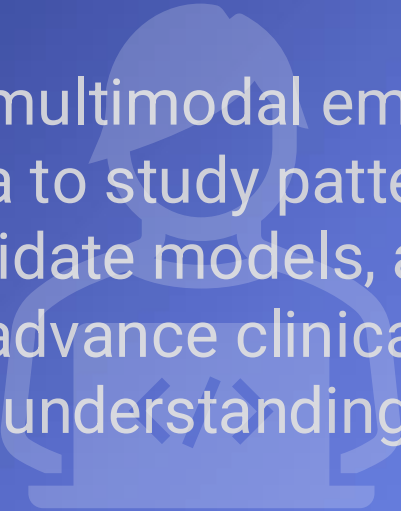
PATIENTS

Benefit from emotion-aware tools that help express and monitor their emotional state when communication is difficult



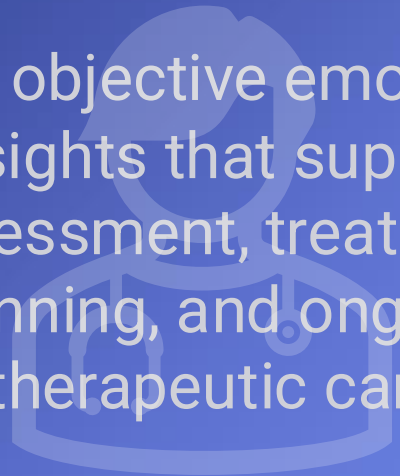
RESEARCHERS

Use multimodal emotion data to study patterns, validate models, and advance clinical understanding



THERAPISTS / PSYCHOLOGISTS

Gain objective emotional insights that support assessment, treatment planning, and ongoing therapeutic care



MANAGEMENT

1 DESIGN

1. Define project goals
2. Identify stakeholders
3. Outline value proposition
4. Specify functional requirements
5. Design presentation theme

2 MANAGE

1. Create Gantt timeline
2. Develop risk assessment table
3. Assign team roles and tasks
4. Setup GitHub repository and policy

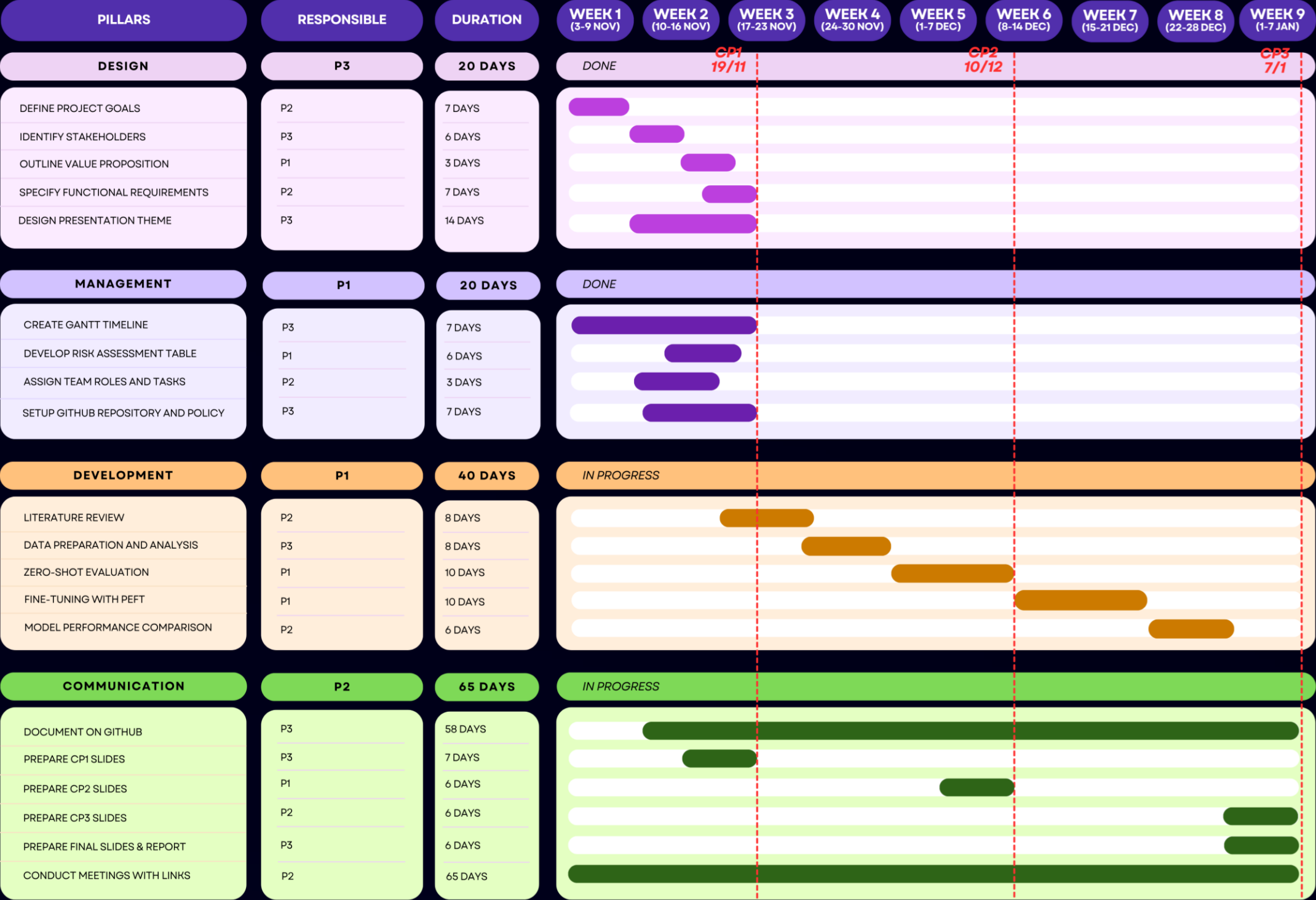
3 DEVELOP

1. Literature review
2. Data preparation & analysis
3. Zero-shot evaluation
4. Fine-tuning with PEFT
5. Model performance & comparison

4 COMMUNICATE

1. Document on GitHub
2. Prepare CP1 slides
3. Prepare CP2 slides
4. Prepare CP3 slides
5. Prepare final slides and report
6. Conduct meetings with LINKS

GANTT TIMELINE





FUNCTIONAL REQUIREMENTS

1

Multimodal data acquisition and alignment

2

Reliable multimodal alignment between audio and transcription

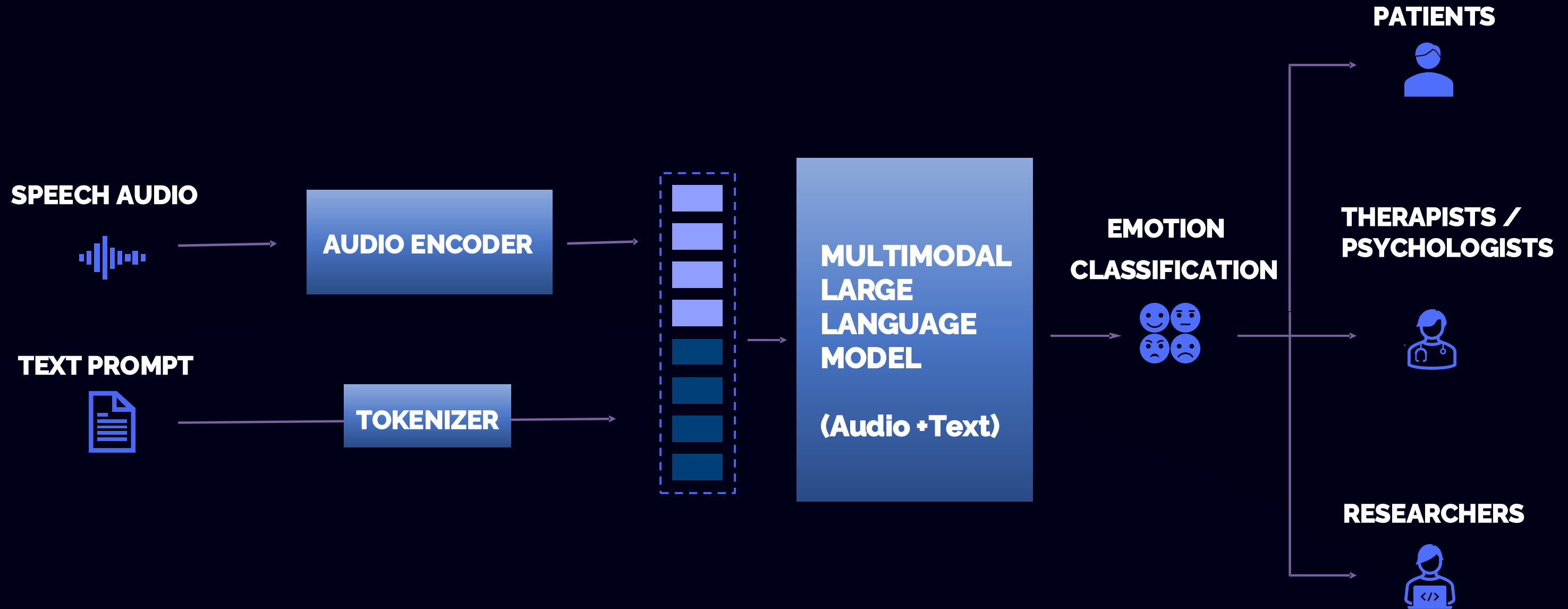
3

Generation of emotion predictions in a standardized format

4

Reporting, explainability, and clinical integration

FUNCTIONAL DIAGRAM



DESIGN RISK ANALYSIS



DATA QUALITY & BIAS



UNFAIR PREDICTIONS



REDUCE MODEL RELIABILITY



MULTIMODAL FUSION FAILURE



AUDIO & TEXT MISMATCH



WRONG CLASSIFICATION



OVERFITTING IN PEFT
FINE-TUNING



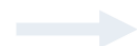
ADAPT TO TRAINING DATASET



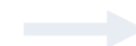
POOR GENERALIZATION



PRIVACY & ETHIC



USAGE OF SENSITIVE DATA



LOSS OF TRUST & LEGAL ISSUES

Design Failure Mode and Effects Analysis (DFMEA)					
Failure Mode	Cause	Effect	Severity	Detection	Mitigation
Biased Predictions	Imbalanced Dataset	Unfair Results	High	Medium	Dataset Balancing, Fairness Audits
Fusion Misalignment	Modal Desync	Wrong Emotions	Medium	Medium	Temporal Alignment, Validation Tests
Privacy Exposure	Raw Audio Stored	GDPR Issues	Very High	Low	Anonymization, On-device Processing
Overfitting	Small Data	Weak Generalization	Medium	High	Cross-validation, Regularization

**THANK
YOU!**