POLITECNICO DI MILANO

PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 40th cycle

Research Area n. 1 - Computer Science and Engineering

PNRR 630 Research Field: SPEECH-BASED EMOTION RECOGNITION AND CONVERSATIONAL INTERFACES FOR MENTAL HEALTH AND WELL-BEING

Monthly net income of PhDscholarship (max 36 months)
€ 1500.0 In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	Speech Emotion Recognition (SER) is a branch of Emotional Artificial Intelligence that involves the use of speech corpora, algorithms, and machine learning models to identify and analyze the emotional states of speakers based on their vocal expressions. By examining various features of speech, ranging from prosodic features such as rhythm and intonation to spectral and voice quality features, SER systems detect emotions like happiness, sadness, anger, surprise,and more. This technology, sometimes combined with text and facial emotion recognition, is increasingly used in conversational interfaces and interactive applications to provide more personalized and empathetic responses. While academic and industrial interest in this field is growing, SER remains an open research area. The objectives ofthis PhD research are: i) to address some critical challenges in the current state of SER, including the quality of speech corpora, the effectiveness of audio features for emotion representation, and the emotion classification models; ii) toapply the scientific results for mental health monitoring of different populations (children and adults). The potential of SER in this application area is grounded in research in psychiatry and neuroscience, which indicates that emotionalstates and mental health conditions are linked through biological mechanisms (e.g.,

	neurotransmitter imbalances) andemotional states over time can be indicators of various mental health conditions. For example, persistent sadness orlow mood is a hallmark of depression, and difficulty in emotion regulation is a potential indicator of borderline personality disorder or neuro-developmental disorders such as autism.
Methods and techniques that will be developed and used to carry out the research	The research will develop methods and technological tools for accurately gathering and pre-processing (noisecleaning and annotation) speech data. This will lead to the creation of a new large speech corpus in Italian to enable the exploration and application SER in this under-represented language (currently, only three small emotional datasetsare available in Italian, and they have many limitations). Additionally, the research will identify new (combinations of) speech features, develop analysis techniques to extract them, and test their capability to effectively describe every emotion. There is currently no commonly accepted agreement on which features represent emotions most accurately. Furthermore, the research will compare numerous existing models (both traditional and deep learning-based) for classifying emotions from the extracted features. These models will be tested on the Italian dataset also using different(combinations of) features, to identify the strengths and weaknesses of the various models and improve/refine the most promising ones. To apply the scientific results for mental health monitoring, the research will develop innovative conversational applications for various ecological contexts (school, home, workplace) that engage users in age-appropriate tasks.These features will be integrated with new tools to analyze the evolution of emotional states over time and to relate them to subjects' mental health and well-being. The research will benefit from multidisciplinary collaboration with various stakeholders. Data gathering for speech corpus creation will be performed in cooperation with the company (specialized in e-health) hosting the candidate's six-month internship, supporting the recruitment of subjects among their healthcare service users. A network of local schools and care centers

	involved in ongoing educational projects with Polimi will also contribute to the recruitment of speakers and users. Neuroscientists from the University of Milan Bicocca and the University of Trento, along with mental health specialists from the Istituto Neurologico Besta (Milan), will contribute to the design and validation of the analytical tools for the clinical interpretation of the emotional data.
Educational objectives	The PhD candidate will develop a multidisciplinary profile with robust methodological and technological competencies in various areas: Speech-Based Emotion Recognition (SER) and its core techniques (e.g., dataset creation and pre-processing, audio-signal analysis, and machine learning), conversational interfaces, and emotional AI. The candidate will develop research skills in these fields and the ability to use these technologies to address real-life problems from multiple perspectives (technical, business, and psychological). Additionally, he/she will gain domain-specific experience in human emotional, affective, and cognitive functions, as well as mental health treatment. Finally, the candidate will acquire the ability to exploit this knowledge and skills to design new audio-analysis and classification techniques and create personalized emotion-driven interactive applications.
Job opportunities	The research interest in SER, emotional AI, and conversational interaction is continuously growing, as evidenced by the increasing number of conferences and publications in the area. For this reason, at the end of their PhD, the candidate will have attractive opportunities to work in various research contexts – academic, public, and private. The increasing integration of AI in everyday applications creates a high demand in industry for skilled professionals who can design, implement, and manage these advanced systems. Countless job opportunities will open up in companies across a wide range of sectors, including e-health and beyond. The specific competence in SER, conversational interaction, and emotional AI will make the candidate attractive for highly qualified positions in any ICT enterprise interested in leveraging AI to create

	speech-based, personalized, and emotionally resonant user experiences, thereby gaining a competitive advantage in the market.
Composition of the research group	0 Full Professors 1 Associated Professors 2 Assistant Professors 3 PhD Students
Name of the research directors	Franca Garzotto

Contacts
franca.garzotto@polimi.it; phone (office): +39-02 23993505 https://i3lab.polimi.it/

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	750.0 €
By number of months	6

National Operational Program for Research and Innovation	
Company where the candidate will attend the stage (name and brief description)	MEDISPA S.r.l.
By number of months at the company	6
Institution or company where the candidate will spend the period abroad (name and brief description)	MIT – MASSACHUSETTS INSTITUTE OF TECHNOLOGY/Mc Govern Institute for Brain Research (USA)
By number of months abroad	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information
EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student. TEACHING ASSISTANTSHIP: availability of funding in recognition of supporting teaching activities by the PhD student. There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations.

COMPUTER AVAILABILITY:
1st year: Yes
2nd year: Yes
3rd year: Yes