FONDAZIONE BRUNO KESSLER
MAGAZINE

Artificial intelligence for emergency medicine: Pietro Ferrazzi's research

August 6, 2026

Transforming clinical documentation into structured data to advance research

Transforming millions of clinical documents into structured information that can support research and improve the work of emergency room physicians. This is the goal of the work of **Pietro Ferrazzi**, a doctoral student in the Natural Language Processing (NLP) unit at FBK's Center for Augmented Intelligence and a PhD candidate in the Brain, Mind and Computer Science doctoral program at the University of Padua. At the heart of his research is one of the still-open challenges in natural language processing for medicine: developing models capable of working with real clinical documentation, not just datasets derived from the scientific literature.

After moving to Trentino to pursue his PhD, Ferrazzi has built an interdisciplinary career that combines statistics, big data, and artificial intelligence. After earning a bachelor's degree in Statistics and a master's degree in Big Data at the University of Padua—enriched by two study experiences at the Université Libre de Bruxelles and the Universitat Politècnica de Catalunya (Barcelona)—he chose to continue his career at Fondazione Bruno Kessler, where he works on natural language processing applied to medicine.

"My research focuses on post-training methods for natural language processing models in the medical field. These models are generally developed and evaluated using datasets derived from the scientific literature. My goal is to understand how to improve their performance when they are applied to emergency room clinical documentation, which has very different characteristics."

One of the case studies through which Ferrazzi tests these methodologies is Case Report Form Filling, developed as part of the European eCREAM project, which aims to optimize **emergency medicine** research by making better use of the information contained in **clinical documentation**.

Every day, healthcare professionals collect a large amount of patient information in textual form. To make this information usable for **research**, however, it must be manually transferred into a structured format. One useful tool for this purpose is the Case Report Form, the form used to collect clinical data in a standardized way. It is an essential activity, but one that is lengthy and repetitive.

*“The goal is to automate this process. Starting from the clinical documentation, we want to automatically identify information such as previous diseases, ongoing therapies, or clinical parameters and transform it into structured data,”*Pietro Ferrazzi explains.

To develop and evaluate these systems, the research team used a particularly valuable dataset consisting of **approximately two million clinical documents** collected over three years at San Giovanni Bosco Hospital in Turin and manually annotated by clinicians.

“Having access to notes written by physicians is already unusual. Even rarer is having data directly from a hospital that can be published, which requires an enormous investment of time and expertise.”

Building on this dataset, Ferrazzi and researchers from FBK’s Natural Language Processing unit organized the **Case Report Form Filling Shared Task**, hosted as part of the CL4Health workshop at the LREC 2026 international conference, with the goal of bridging the gap between the needs of emergency medicine and the natural language processing community.

The first edition of the competition involved twelve international research groups, which developed and compared artificial intelligence systems using the dataset made available by FBK. In addition to presenting the results during the conference, the initiative fostered the creation of new scientific collaborations. For the next edition, the group is working to involve new European partners with the aim of further expanding the research community focused on this challenge.

Case Report Form Filling is also an important part of Ferrazzi’s doctoral research. His **dissertation** focuses on **post-training methodologies** for natural language processing applied to medicine, and this case study allows him to evaluate how these methodologies perform when moving from datasets derived from the scientific literature to real clinical documentation.

*“It is an important testing ground because clinical data have very different characteristics from traditional datasets. In addition, there are still very few resources of this type available for Italian, whereas the availability for English is significantly greater,”*Ferrazzi adds.

The ultimate goal is to develop tools capable of reducing the workload of healthcare professionals while at the same time unlocking the enormous wealth of information stored in hospital archives. Making this information accessible means not only simplifying activities that are currently performed manually, but also providing structured data for research that can contribute to the development of emergency medicine increasingly supported by artificial intelligence.

TAGS

- #artificialintelligence
- #augmentedintelligence
- #big data
- #dataset
- #dati clinici
- #FBKPhDProgram
- #language
- #medicina
- #modelli
- #NLP
- #statistica

AUTHORS

- Michela Antino