

# FoundationEHR – a foundation AI model for structured electronic health records

## High potential for the use of structured electronic health records (EHRs)

Structured Electronic Health Records (EHRs) comprising diagnoses, prescriptions, and procedure codes have widespread use in many healthcare systems. They are, e.g., utilized for health-economic, pharmaco-epidemiological, pharmaco-vigilance, and marketing purposes. Furthermore, there is an increasing interest in using these data for personalized risk prediction, understanding real-world disease trajectories, constructing external control arms, or emulating entire clinical trials. In this context, artificial intelligence (AI) and, specifically, machine learning (ML) are employed more and more.

## Existing challenges

Structured EHR data is initially not ready to apply statistical modeling techniques or advanced AI / ML. Traditionally, many researchers have constructed an analysis dataset by simply indicating the presence/absence of specific diagnoses before a defined index date in a patient's history.

This approach ignores any temporal aspects of disease development and is often biased towards a priori-defined diagnosis, preventing the discovery of new insights from data.

## Our solution: Leverage the potential of EHRs via deep learning

We have developed a modern deep learning-based approach to deal with structured EHR data, which leverages the full potential of these data. FoundationEHR is a transformer-based deep neural network architecture pre-trained on almost 1 billion EHRs from ~3.5 million patients in the USA. FoundationEHR is a foundation model, i.e., an AI model, from which the customer can derive further task-specific AI/ML models in a further fine-tuning (transfer-learning) procedure. FoundationEHR distinguishes itself from other solutions through multiple aspects:

- FoundationEHR can model prescription information in addition to diagnoses

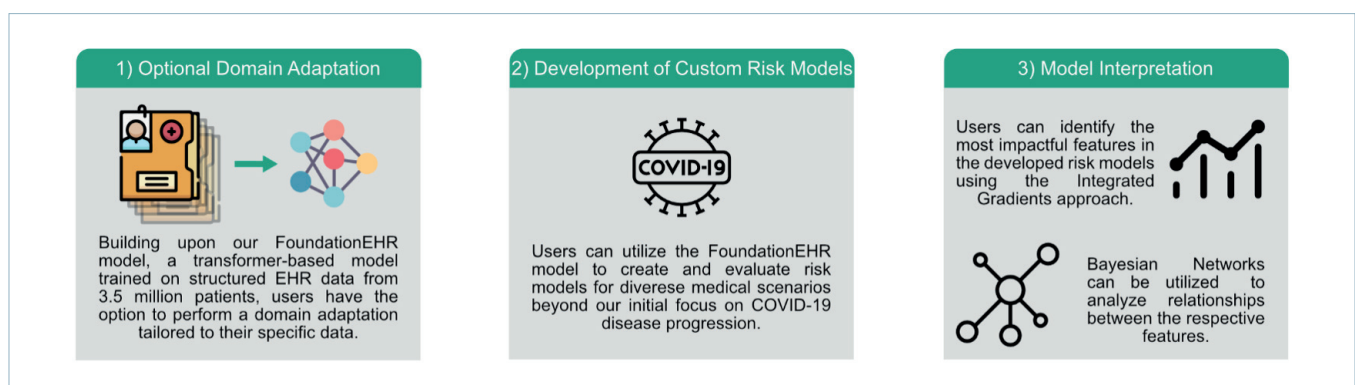


Figure 1: Example workflow for the use of FoundationEHR.

- FoundationEHR models information about a patient's age at each doctor visit. Furthermore, information about which diagnosis and prescriptions belong to the same encounter is correctly modeled.
- FoundationEHR models information about a patient's residency and gender
- FoundationEHR can integrate quantitative lab or imaging data in addition to qualitative EHRs.
- Explanation of model predictions

In our publication, we describe the technical details of the model architecture and the fine-tuning of a specific use case, namely the prediction of severe COVID-19 disease progression.

## Our offering

We offer FoundationEHR to our customers as a downloadable foundation model pre-trained via PyTorch (version 1.13.0). Customers can load FoundationEHR in their in-house Python environment and perform task-specific fine-tuning. We thus offer full flexibility regarding the intended use of FoundationEHR.

## Reference

M. Lentzen et al., "A Transformer-Based Model Trained on Large Scale Claims Data for Prediction of Severe COVID-19 Disease Progression," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 9, pp. 4548-4558, Sept. 2023, <https://doi.org/10.1109/JBHI.2023.3288768>

### Contact

Fraunhofer Institute for  
Algorithms and Scientific  
Computing SCAI  
Schloss Birlinghoven 1  
53757 Sankt Augustin  
Germany

Prof. Dr. Holger Fröhlich  
Phone +49 2241 14-4219  
[holger.froehlich@scai.fraunhofer.de](mailto:holger.froehlich@scai.fraunhofer.de)

Manuel Lentzen  
Phone +49 2241 14-4219  
[manuel.lentzen@scai.fraunhofer.de](mailto:manuel.lentzen@scai.fraunhofer.de)

[www.scai.fraunhofer.de/bio](http://www.scai.fraunhofer.de/bio)

### Distributed by

scapos AG  
Schloss Birlinghoven 1  
53757 Sankt Augustin  
Germany

Phone +49 2241 14-4400  
[info@scapos.com](mailto:info@scapos.com)

[www.scapos.com](http://www.scapos.com)