

## EMPAIA

Ökosystem für Pathologie-Diagnostik mit KI-Assistenz

|                                 |   |                        |               |
|---------------------------------|---|------------------------|---------------|
| <b>Programm / Ausschreibung</b> | IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - Vorbereitung Horizon Europe | <b>Status</b>          | abgeschlossen |
| <b>Projektstart</b>             | 01.07.2020  | <b>Projektende</b>     | 30.06.2023    |
| <b>Zeitraum</b>                 | 2020 - 2023   | <b>Projektlaufzeit</b> | 36 Monate     |
| <b>Keywords</b>                 | KI, Pathologie, Erklärbarkeit   |                        |               |

### Projektbeschreibung

Unser Ziel ist es, maschinelle Entscheidungen transparent, rückverfolgbar und damit für einen medizinischen Experten interpretierbar machen. Künstliche Intelligenz (KI) und Maschinelles Lernen (ML) sind bereits sehr erfolgreich. Einige Methoden übertreffen heute schon die menschliche Leistungsfähigkeit - auch in der Medizin. Wir werden PathologInnen in die Lage versetzen, den Kontext verstehen zu können und bei Bedarf, eine maschinelle Entscheidung nach dem "Warum" zu hinterfragen. Unsere Forschung wird in Erweiterung von statistischen und modell-freien Klassifizierungen, auch kausale Modelle einschließen. Wir werden konkret:

1. Untersuchen, wie medizinische Expertinnen und Experten ihre Entscheidungen erklären, indem wir ihre Strategien analysieren, dies mit den zugrunde liegenden Erklärungsfaktoren der Daten vergleichen, um so ein strukturelles Kausalmodell der menschlichen Entscheidungsfindung zu formalisieren und Merkmale in diesen auf KI/ML-Ansätze anzuwenden.
2. Methoden entwickeln, um die Qualität von Erklärungen zu messen, und untersuchen wann dies ausreichend ist; wir werden Methoden entwickeln "Kausalität" zu messen, d.h. das Ausmaß, in dem eine Erklärung einer Aussage gegenüber einem menschlichen Experten ein bestimmtes Maß an kausalem Verständnis mit Effektivität, Effizienz und Zufriedenheit in einem bestimmten Nutzungskontext erreicht.
3. Die gewonnenen Erkenntnisse aus 1) und 2) nutzen, um neue Human-AI Interfaces zu entwickeln, zu testen und zu bewerten, die von medizinischen Experten trainiert werden können, um ihnen die zugrunde liegenden Prinzipien verständlich zu machen. Dies wird die Zuverlässigkeit, Verantwortlichkeit, Fairness und das Vertrauen in KI erhöhen und eine ethisch verantwortungsbewusste KI fördern.

### Abstract

We will make machine decisions transparent, re-traceable, hence interpretable for a medical expert. AI/Machine Learning is very successful and certain methods exceed human performance - even in medicine. Our goal is to enable pathologists to

understand the context, thereby enabling them to re-enact on demand. Our research will include structural causal models, because current AI/ML work on either a statistical or model-free mode. We hypothesize that in medicine we need the guidance of a model of reality, like the ones used in causal inference tasks. We will

1) investigate how medical experts explain their decisions by looking at their strategies, as they can explore the underlying explanatory factors of the data, so to formalize a structural causal model of human decision making and mapping features in these to AI/ML approaches. For example in digital pathology such mechanistic models can be used to analyze and predict the response of a functional network to features in histology slides, molecular data and family history.

2) develop methods to measure the quality of explanations and when it is enough; we will pioneer in solutions to measure "causability" as the extent to which an explanation of a statement to a human expert achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use.

3) use the gained insights from 1) and 2) to develop, test and evaluate novel interface techniques, trainable by medical experts to make underlying principles understandable to them. This will enhance reliability, accountability, fairness, trust in AI methods and foster ethical responsible ML. Our integrating elements are stochastic and-or graphs, scalable Bayesian rule lists, and we will build on and extend stochastic computation in order to produce interpretable models which include structural variabilities and geometric variabilities (what and where).

## **Projektpartner**

- Medizinische Universität Graz