



## Robert Koch-Institut Deutschlands Public Health Institut

### Doktorandin/Doktorand (mehrere Stellen) Künstliche Intelligenz & Public Health (m/w/d)



**Arbeitsbeginn**  
01.10.2023



**Bewerbungsfrist**  
20.04.2023



**Befristung**  
3 Jahre



**Vergütung**  
E 13 TVöD



**Standort**  
Wildau



**Referenz**  
935773| 64/23

Wir suchen befristet für 3 Jahre Verstärkung in der Abteilung ZKI-PH „Zentrum für künstliche Intelligenz - Public Health“ an unserem Standort Wildau. Werden Sie Teil von einer der größten deutschen Ressortforschungseinrichtungen. Wir zählen auf Sie!

[Hier bewerben](#)

Hierbei handelt es sich um eine Teilzeitposition mit 25,35 Wochenstunden (65 % einer Vollbeschäftigung), die es Ihnen unabhängig von Ihrer Arbeitszeit ermöglicht, Ihre Doktorarbeit anzufertigen.

Wir bieten Ihnen die Möglichkeit der Mitarbeit an spannenden Forschungsprojekten zum Wohle der Öffentlichen Gesundheit mit dem Qualifizierungsziel der Promotion nach 3 Jahren sowie die Integration in eines unserer interdisziplinären, innovativen und internationalen Teams. Sie erhalten umfangreiche Unterstützung beim Aufbau Ihres professionellen Netzwerkes, der persönlichen Weiterbildung und Karriereplanung.

Aus befristungsrechtlichen Gründen richtet sich die konkrete Vertragsdauer (max. 3 Jahre) nach der individuellen Verfügbarkeit von Zeiten im Sinne des § 2 (1) des Wissenschaftszeitvertragsgesetzes (WissZeitVG).

#### Ihre Aufgabe bei uns

**Wissenschaftliche Forschungstätigkeit** im Bereich **Künstliche Intelligenz für Public Health**, insbesondere:

- eigenverantwortliche Bearbeitung eines spannenden Forschungsprojektes an der Schnittstelle KI, Medizin und Bioinformatik z.B. durch die
  - Weiterentwicklung von Frühwarnsystemen, Risikoeinschätzung von Krankheitserregern; Optimierung der Berechnung von Krankheitslasten, Visualisierung komplexer Zusammenhänge, Big Data Analysen, automatisierte Analyse von Hochdurchsatzexperimenten in der Molekularbiologie, Bild- und Videoanalysen sowie Mustererkennung komplexer Public Health Daten
- Mitarbeit bei der Algorithmen-/Methodenentwicklung sowie Entwicklung entsprechender Softwaretools
- Präsentation von Forschungsergebnissen auf nationalen und internationalen Kongressen und Fachtagungen
- Erstellung von wissenschaftlichen Publikationen sowie Projekt- und Forschungsberichten

Eine komplette Liste an verfügbaren Themen befindet sich auf der [Webseite des ZKI-PH](#).

#### Das bringen Sie mit

##### Formale Voraussetzungen

- ein abgeschlossenes wissenschaftliches Hochschulstudium (Universitätsdiplom, Master oder Staatsexamen) in einer relevanten (informations-) technischen, naturwissenschaftlichen, medizinischen, Public Health Fachrichtung oder vergleichbar

#### Unser Zentrum

Das neue Zentrum für Künstliche Intelligenz in der Public Health Forschung (ZKI-PH) bietet ein einzigartiges Arbeits- und Forschungsumfeld: mit innovativen Methoden, einem internationalen Team und modernster Ausstattung setzen wir uns für die Gesundheit der Bevölkerung ein.

Bei ausländischen Bildungsqualifikationen benötigen wir einen Nachweis über die Gleichwertigkeit mit einem deutschen Abschluss.



## Kenntnisse und Erfahrungen

- in der quantitativen Datenanalyse sowie im Umgang mit komplexen Datenkörpern
- in der Anwendung min. einer Programmiersprache (z.B. Python, Java ...)

## Zusätzlich wünschenswert

- Erfahrungen in der Umsetzung von Projekten im Bereich Künstliche Intelligenz, Maschinelles Lernen und/oder Deep Learning
- erste Erfahrungen in Bezug auf wissenschaftliches Publizieren und Präsentieren von Forschungsergebnissen
- erste Kenntnisse oder Erfahrungen in der Public Health Forschung

## Weitere Voraussetzungen

- Sprachkenntnisse (CEFR-Niveau): Englisch mindestens C1, Deutsch mindestens A2 bzw. die Bereitschaft, sich weitere Deutschkenntnisse schnellstmöglich anzueignen, um effektiv im Arbeitsumfeld kommunizieren zu können (Deutschkurse werden angeboten)
- Bereitschaft zur Teilnahme an einer erweiterten Sicherheitsüberprüfung nach § 9 Sicherheitsüberprüfungsgesetz (SÜG) sowie deren positiver Abschluss (nur für Projektthema ZKI-PH\_PhD2023\_03 und ZKI-PH\_PhD2023\_05)

## Damit überzeugen Sie uns

- Selbstständigkeit, eigenverantwortliche Arbeitsweise und Verantwortungsbewusstsein
- Kooperations- und Teamfähigkeit im Einsatz für gemeinsame Ergebnisse
- Zuverlässigkeit und eine termingerechte, qualitativ hochwertige Arbeitsweise

## Wir freuen uns auf Ihre Bewerbung

unter [www.interamt.de](http://www.interamt.de)

zur StellenID 935773

bis 20.04.2023.

## Das zeichnet uns aus

### GELEBTE VERANTWORTUNG

Wir bieten einen Job, mit dem Sie einen Unterschied machen. Als Deutschlands wichtigstes Public Health Institut können wir viel bewegen.

### TEAMWORK UND ZUSAMMENHALT

Unsere Ziele erreichen wir nur gemeinsam. Daher legen wir großen Wert auf respektvolles Miteinander und eine offene Feedback-Kultur.

### FLEXIBILITÄT UND ARBEITSZEIT

Wir gehen mit der Zeit: wir bieten individuelle, flexible Arbeitszeitmodelle, 30 Urlaubstage im Jahr und die Möglichkeit für längere Auszeiten.

### MOBILITÄT

Mobile Arbeit gehört für uns zum beruflichen Alltag. Für die Fahrt ins Büro gibt es ein Jobticket und für kurze Wege auch Dienstfahrräder.

### ARBEIT, FAMILIE UND MEHR

Familienservice, Kitaplätze, Gleitzeit: Das Leben ist bunt und unsere Angebote sind es auch. So divers wie unsere Beschäftigten selbst!

### FREIZEITANGEBOTE UND EVENTS

Ob Teamsport, Entspannungskurs oder Cinema-Club, bei uns kommt auch die Freizeit nicht zu kurz. Gemeinsame Events? Auch die gibt es regelmäßig!

### FORT- UND WEITERBILDUNG

Wir können nur gemeinsam wachsen. Daher unterstützen wir Ihre berufliche, akademische und persönliche Weiterentwicklung.

### TOP AUSSTATTUNG

Bereichernde Arbeitserfahrungen und modernes Equipment für vielfältige Forschung und Administration.

## Sie haben Lust auf eine neue Herausforderung?

Werden Sie Teil der RKI-DNA! Wir freuen uns auf Sie.

*Wir leben Chancengleichheit und begrüßen alle Menschen in allen Dimensionen von Diversität. Deshalb gewährleisten wir die berufliche Gleichstellung. Schwerbehinderte Menschen werden bei gleicher Qualifikation und Eignung bevorzugt berücksichtigt.*

*Das Bundesministerium für Gesundheit kann im Rahmen seiner aufsichtsrechtlichen Befugnisse im Einzelfall Einblick in Ihre Bewerbungsunterlagen nehmen. Ihre Daten werden nach Abschluss des Bewerbungsverfahrens gelöscht.*

## Sie haben noch weitere Fragen?

### Ihr Kontakt

Anne Lettau

Telefon: +49 30 18754-3551

E-Mail: [LettauA@rki.de](mailto:LettauA@rki.de)



Robert Koch-Institute, Germany's Public Health Institute

## Center for Artificial Intelligence in Public Health Research (ZKI-PH)

ZKI-PH PhD Programme (2023-2026)



# ZKI-PH PhD Programme (2023-2026)

## **ZKI-PH\_PhD2023\_01 (ZKI-PH4 & ZIG2)**

Identifying climatic and non-climatic factors driving outbreaks of malaria in Madagascar using artificial intelligence

## **ZKI-PH\_PhD2023\_02 (ZKI-PH3 & FG16)**

AI-based spatial mapping of multiple microbial pathogens in complex, polymicrobial communities

## **ZKI-PH\_PhD2023\_03 (ZKI-PH5 & ZBS6)**

AI-supported proteomics analysis for an effective antimicrobial therapy decision

## **ZKI-PH\_PhD2023\_04 (ZKI-PH5 & FG32)**

Unsupervised learning for surveillance indicators

## **ZKI-PH\_PhD2023\_05 (ZKI-PH2 & ZBS5)**

Characterizing the evolution and epidemiology of Lassa virus across hosts and reservoirs using AI-related methods

## **ZKI-PH\_PhD2023\_06 (ZKI-PH4 & FG24)**

Applying causal inference in AI-based spatio-temporal modelling of smoking behavior and tobacco control measures

## **ZKI-PH\_PhD2023\_07 (ZKI-PH5 & MF1)**

Effective 3D representations of protein structures to identify antimicrobial resistance at protease-associated sites

## **ZKI-PH\_PhD2023\_08 (ZKI-PH5 & FG26)**

Analyzing public discourse in information media with AI for deeper insights into concurrent developments in public mental health

## **ZKI-PH\_PhD2023\_09 (ZKI-PH2 & FG13)**

Analysis of transmission dynamics of hospital-associated *S. aureus* from whole genome sequencing data



# ZKI-PH\_PhD2023\_01 (ZKI-PH4 & ZIG2)

## Identifying climatic and non-climatic factors driving outbreaks of malaria in Madagascar using artificial intelligence

### Background:

Climate change is one of the greatest and most immediate threats to human and planetary health, and Madagascar is one of the countries most affected by climate change in the world. Vector-borne diseases are particularly sensitive to changes in climate. Altered temperature, precipitation, and extreme weather events can all influence the prevalence and distribution of arthropod vectors. Malaria, transmitted by *Anopheles* mosquitoes, is widespread throughout tropical and subtropical regions. In 2021, there were 247 million cases of malaria worldwide, with 95% of all cases occurring in the WHO Africa Region (WHO 2022). Malaria remains one of the leading causes of death in low-income countries; and is particularly climate-sensitive, with the WHO predicting that climate change will lead to a 15% annual global increase in deaths from malaria between 2030 and 2050 (WHO 2020).

### Aim/s:

The aim of this project is to identify and map the climatic and non-climatic factors associated with outbreaks of malaria in Madagascar using artificial intelligence, to support the development of early warning systems for local public health departments.

### AI methods:

The use of AI in this project would enable an automated approach to reviewing several data sources and combining these findings to prospectively identify regions that are likely to experience outbreaks of malaria. Relevant techniques span from machine learning for time series analysis, geospatial data analysis, and causal inference.

### Keywords:

Climate change, Infectious Diseases, Malaria, Madagascar, Machine Learning



# ZKI-PH\_PhD2023\_02 (ZKI-PH3 & FG16)

## AI-based spatial mapping of multiple microbial pathogens in complex, polymicrobial communities

### Background:

Pathogenic microorganisms inhabit polymicrobial niches as biofilms in many human infections and in the environment. Behavior of these pathogens is influenced by environmental conditions and interaction with co-localized microbes. While next generation sequencing revolutionized the identification of microbes in complex communities, spatial mapping to identify co-localized, potentially interacting microbes remains challenging. So far, the most successful approach to resolve spatial organization of polymicrobial communities uses fluorescence *in situ* hybridization (FISH) with binary coded DNA probes targeting phylogenetically informative ribosomal RNA sites combined with other information using image processing. Resolving the structure of polymicrobial biofilms allowed for the generation of hypothesis on the function of individual microbes within communities. Understanding the behavior of microorganisms in these communities is important to predict their reaction to changing environments, e.g. exposure to antimicrobials or climate change, and opens up new avenues for intervention for the improvement of public health.

### Aim/s:

This project aims to improve the mapping of microbes in polymicrobial communities for application to clinical biofilms and environmental samples. It aims at an AI-supported generalizable tool that allows to study the response of such communities to changing environmental conditions including stressors such as anti-microbials.

### AI methods:

Modern image recognition is based on Deep Learning, which will be utilized in this project. You will design and train deep neural networks to identify different microbes.

### Keywords:

Microbial Pathogens, Clinical biofilms, Fluorescence microscopy, AI based spatial mapping, Deep Learning





# ZKI-PH\_PhD2023\_03 (ZKI-PH5 & ZBS6)

## AI-supported proteomics analysis for an effective antimicrobial therapy decision

### Background:

Antimicrobial resistance (AMR) poses an increasing challenge for the therapy of bacterial infections. Diagnostics-guided therapy relies on species identification and antimicrobial susceptibility testing. Antimicrobial resistance detection is usually based on phenotypic assays. As the proteome is closely related to the phenotype, the RKI developed a proteomics platform, which is capable of species identification and AMR detection. The platform already relies on deep learning and deep neural networks. The further implementation of AI will improve proteomics and transform the platform into an indispensable tool for routine clinical microbial diagnostics. For example, previously unknown virulence factors or antibiotic resistance determinants can be detected, which are important for therapeutic decisions.

### Aim/s:

The aim of the project is to provide an AI-based prediction tool for the detectable species-specific peptidome to enable fully unbiased species identification and to train an algorithm to predict antibiotic resistance classes from proteome-wide expression data.

### AI methods:

You will develop concepts for the architecture and implementation as a basis for the work. You will use deep neural networks as well as various supervised learning methods to program algorithms, train them on existing laboratory data and develop them further.

### Keywords:

Antimicrobial resistance, Proteomics, Diagnostics, Big Data Analysis, Deep Learning

(Security clearance for applicants will be required)



# ZKI-PH\_PhD2023\_04 (ZKI-PH5 & FG32)

## Unsupervised learning for surveillance indicators

### Background:

RKI is currently sourcing data streams for public health (PH) surveillance. One example is emergency department (ED) surveillance, where surveillance indicators are operationalized through so called syndrome definitions consisting of certain variable value conditions and their combinations (e.g. all cases with high temperature, cough-, cold-symptoms). For definition of those syndromes, expert knowledge on the underlying health issue and the characteristics of the data source and data structure is essential. However, size and complexity of the dataset poses a barrier to use its full potential. We already explored two supervised methods (Naïve Bayes classifier, rule-based approach) to help experts handle the data and optimize their syndromes. Unsupervised methods could even further advance that optimization process and guide PH practitioners to identify new indicators based on recent occurrences in the data.

### Aim/s:

For surveillance purposes, newly emerging and unexpected groups of cases shall be detected in time. Aside from already pre-defined PH surveillance indicators, unsupervised learning approaches can help to recognise previously unidentified emerging occurrences in health data streams, and therefore guide PH action in a more flexible and timely manner.

### AI methods:

Unsupervised learning models, that optimize pre-defined and suggest new indicators, e.g. by anomaly detection or clustering approaches, will be explored and implemented. To handle informative multidimensional categorical values, the potential of categorical variable encoding will be utilized. Missing data is a great challenge in processing routinely collected data, wherefore AI methods seem promising. For the validation of the chosen methods external data sources and expert feedback will be used.

### Keywords:

Surveillance, Unsupervised learning, Anomaly detection, Clustering, Multidimensional categorical health data





# ZKI-PH\_PhD2023\_05 (ZKI-PH2 & ZBS5)

## Characterizing the evolution and epidemiology of Lassa virus across hosts and reservoirs using AI-related methods

### Background

Lassa fever, an infection currently affecting 150,000–300,000 people in West Africa and killing up to 5,000 people annually, is a significant viral haemorrhagic disease. It was first detected in Nigeria in 1969 through a case of human-to-human transmission. Several cases were reported in eastern Sierra Leone shortly after the first outbreak, allowing investigators to elucidate the virus's reservoir among common rodents. *Mastomys natalensis* was recognized as a reservoir host of Lassa Virus (LASV) in 1974. Although *M. natalensis* was believed to be the sole reservoir of LASV for 30 years, here we investigate the role of other small rodents such as *Rattus*, *Crocidura* and *Tatera* species and their effect on the evolutionary and epidemiological dynamics of LASV.

### Aim/s:

- 1) Reconstruct the evolutionary history of LASV and thereby elucidate the role of alternative rodent hosts.
- 2) Model and infer the seasonal epidemics of LASV in West Africa.

### AI Methods

The evolutionary and epidemiological dynamics of Lassa virus will be inferred using stochastic population biological processes implemented in a Bayesian Markov Chain Monte Carlo framework. Deep learning approaches (such as Artificial Neural Networks) can facilitate the analysis of temporal transmission dynamics.

### Keywords

Lassa virus, Virus reservoir, Phylodynamics, Bayesian MCMC, Artificial Neural Networks



# ZKI-PH\_PhD2023\_06 (ZKI-PH4 & FG24)

## Applying causal inference in AI-based spatio-temporal modelling of smoking behavior and tobacco control measures

### Background

Public Health Impact Analysis (PHIA) is an approach to assess the implications of public policies and societal changes on health, including the evaluation of preventive measures. The aim of PHIA is to provide valid evidence, using methods of causal inference, in order to inform policy makers. PHIA combines evidence from the literature with data analyses, preferably from various data sources.

Smoking is an important health risk and a general indicator for health behaviour. Different tobacco control policies have been implemented in Germany such as taxation, partial smoking bans (smoke-free environments), increased sales age, mass media campaigns, and partial advertising bans. Since policies have in part been implemented at different times in the federal states, the data represent a quasi-experimental design and can be analysed with causal inference methods such as controlled interrupted time series or difference-in-differences.

### Aim/s:

The aim of the project is to describe the spatio-temporal variation in smoking behaviour in Germany over the last 30 years and to critically assess the effectiveness of tobacco control policies.

### AI Methods

You will combine data from the different data sources (in-house and external) and build a spatio-temporal model of smoking behaviour. You will use machine-learning methods for changepoint analysis, interaction analysis or measurement error models. There will be the opportunity to use natural language processing for literature reviews and to apply methods of causal machine learning.

### Keywords:

Data synthesis, Public health impact analysis, Causal machine learning, Spatio-temporal modelling, Tobacco control



# ZKI-PH\_PhD2023\_07 (ZKI-PH5 & MF1)

## Effective 3D representations of protein structures to identify antimicrobial resistance at protease-associated sites

### Background:

Recent advances in computational (*in silico*) protein folding, such as AlphaFold2, revolutionized our view on the 3-Dimensional structure of proteins. Various research groups have started to use artificial intelligence to fold proteins. The European Bioinformatics Institute (EMBL-EBI) provides open access to over 200 million protein structures. Still, analyses of proteins are often done on their linear representation of amino acids. However, 3D structures represent the functionally active form of the molecule. Thus, investigating protein structure allows searching for essential motifs, such as antibody (AB) binding sites or sites contributing to AntiMicrobial Resistance (AMR).

### Aim/s:

In this PhD project, we will leverage the vast collection of available protein structures to learn 3D shapes that can be associated with particular functions. We will focus on AMR and protease activity, a vital function of an extensive collection of bacterial proteins. Proteases effectively allow bacteria to inactivate antibacterial drugs but can also be considered potential drug targets.

### AI methods:

We aim to develop a structural pattern classifier to make PROtein Structure AMR PredictIONS (PRESUMPTIONs) about protease sites associated with AMR. The representation of a protein will be the volume corresponding to its structure, like a 3D image. We will build on recent advances in image analysis for object recognition and apply them to the problem of recognizing structural protein motifs. The classification based on structural features will be more precise than based only on the sequence.

### Keywords:

Protein structure, Pattern recognition, AlphaFold2 folding, Antimicrobial resistance, Protease activity



# ZKI-PH\_PhD2023\_08 (ZKI-PH5 & FG26)

## Analyzing public discourse in information media with AI for deeper insights into concurrent developments in public mental health

### Background:

Media shapes public attitudes and behaviors towards health and disease, but little is known about the relationship between information media and mental health. Negative depictions of mental health in information media, such as criminality, are concerning. Two research questions can help understand this relationship: what mental health-related topics are discussed in information media and how often, and what societal events are discussed in information media and how often are mental health-related topics discussed in the same report as the events? Answering these questions can benefit public health surveillance and help interpret trends and developments after implementing public health measures. Currently, mental health core indicators show an increasing trend of psychopathological symptoms, potentially related to multiple societal events. Information media can contribute to interpreting results more objectively and less biasedly.

### Aim/s:

The aim is to understand how information media relates to population mental health. For example, mental health core indicators include an increasing number of psychopathological symptoms and are potentially related to multiple societal events. The results could benefit public health surveillance and interpretation of trends in mental health.

### AI methods:

Text mining and multivariate statistical data analyses will be used to address the research questions. This includes supervised modelling using predetermined mental health keywords and sentiment analysis, as well as unsupervised modelling to identify frequently discussed topics. Potential correlations and associations between societal events and mental health discussions will also be examined.

### Keywords:

Information media, Mental health, Public discourse, Text mining, Public health surveillance



# ZKI-PH\_PhD2023\_09 (ZKI-PH2 & FG13)

## Analysis of transmission dynamics of hospital-associated *S. aureus* from whole genome sequencing data

### Background

*Staphylococcus aureus* as leading cause of nosocomial infections can acquire resistance to antibiotics and is a high-priority pathogen for research and development of new antibiotics (WHO). The National Reference Centre for Staphylococci and Enterococci at RKI, is tasked with epidemiological surveillance of staphylococcal infections and strains with special resistance and virulence traits across Germany. It characterised staphylococcal isolates from 200 microbiological laboratories phenotypically and genotypically to elucidate putative outbreak events and resistance- or virulence-associated determinants. These analyses resulted in 4000 *S. aureus* genomic data sets. Phenotypic resistance patterns and molecular typing data are available for all investigated isolates, allowing assignment to specific clonal lineages and thus estimation of possible transmission events.

### Aim/s:

Phylogenetic analysis can be useful for microbial characterization and control of circulating bacterial clones in hospitals. This PhD project aims to identify strain collections with closely related isolates obtained over an extended period of time from hospitals within a limited geographic area, and to confirm possible “outbreak clusters”, as well as performing transmission analyses and identifying the introduction source or route of transmission for these clusters. Underlying mechanisms of differences in spreading ability of closely related clones should be discovered using advanced phylogenetic methods and inclusion of further metadata. A deeper knowledge of the genetic variability of *S. aureus* in clinical settings would lead to a better understanding of its pathogenic potential and improved strategies to contrast its virulence and resistance, in order to adjust appropriate therapies and avoid ineffective treatments for microbial infections.

### AI Methods

Bayesian inference, phylogenetic/phylogenomic approaches and supervised learning.

### Keywords

Antimicrobial resistance; Surveillance; *Staphylococcus aureus*; Nosocomial transmission