

Profile  
(1.0 FTE)  
Chair: Artificial Intelligence and Data Theory

Mathematical Institute, Faculty of Sciences, and  
Section Methodology and Statistics, Institute of Psychology, Faculty of Social and  
Behavioural Sciences

Leiden University

## **1. Description of the Research Area of the Chair**

Artificial Intelligence and Machine Learning find meaning in data. Data Theory is an international area of research that focuses on the general theory of data and develops models for data analysis. It can be described as the theory of Data Science and is closely linked to statistics and its applications, in particular in the social and behavioural sciences. Leiden University has a long tradition in Data Theory. Through this chair this will be extended to distance driven machine intelligence, in particular machine learning techniques for prediction and classification.

### **Challenges in the Research Area**

There are many new sources of data in most scientific areas, including psychology, sociology, political sciences, pedagogy, criminology, epidemiology, health sciences and biology. Besides traditional data from questionnaires, population censuses, physiological measurements, psychological tests or reaction times, nowadays there are many new types of data, often high-dimensional, for instance: imaging (EEG, fMRI, PET), omics (genomics, proteomics, metabolomics), eye-tracking, mobile phones and other wearables, diaries, internet (including facebook and twitter). The data-deluge provides attractive panoramas, but also obstacles.

The panoramas include personalized medicine, personalized learning and preventive health care. Many routes of the Nationale Wetenschapsagenda are connected to these new possibilities. Obstacles are formed by the diversity of the data sources (necessitating “data fusion”): high-dimensionality, noisiness, incompleteness, diffuse measurement levels, observational versus experimental settings. Data Theory focuses on developing a comprehensive strategy to analyse such complex data, with a view to prediction (“supervised learning”) and clustering of data points (“unsupervised learning”). The concept of “distance”, both in the high-dimensional data space and in low-dimensional representations of this space, plays a crucial role, as well as transformation of attributes. Understanding and visualization of the relationships between data points (e.g. patients, genes, brain areas, social organisations) and their attributes are important aims. The chair aims to study these and other machine learning methods in the context of typical Artificial Intelligence applications of the SAILS program. New methods will often be computer-intensive and expressed in algorithms.

Data Theory will apply in great generality, no matter the nature of the data points or their attributes. However, the research of the chair will be linked in particular to the social and behavioural sciences.

Typical important research topics are:

- Description of relationships in high-dimensional data using distances and geometry
- Dimension reduction based on geometry
- Development of advanced machine learning methods
- Optimal transformation of data
- Optimization by minimizing loss functions through dedicated algorithms
- Statistical methods for stability, generalization and model selection
- Efficient algorithms for analysis and decision making

### **Goal (“leeropdracht”)**

The professor in Artificial Intelligence and Data Theory will develop general theories of data and machine learning techniques to analyse data, in particular for categorical data. The professor makes the new techniques available through publicly accessible software (e.g. SPSS, R or Python). Visualisation will be an essential ingredient of the new techniques.

## **2. Profile of the Full Professor Data Theory**

The professor shall have:

- excellent research qualities in the field of data science, in particular in the social and behavioural sciences,
- an internationally renowned reputation in this field,
- affinity with interdisciplinary research,
- the ability to attract prestigious national and international personal grants,
- the ability to initiate international research programs and consortia,
- experience in establishing research programs and supervising and coordinating scientific research,
- the ability to supervise and inspire undergraduate, graduate, and PhD students and researchers,
- excellent communication and social skills,
- significant teaching experience at the undergraduate and graduate level.

## **3. Embedding**

At the level of Leiden University, the chair will contribute to the recently started initiative on Social Science, Humanities, Law, AI and Life Sciences (“SAILS”), in which multidisciplinary research across different faculties, with AI as a core technology, is conducted. SAILS provides ample opportunities for multidisciplinary collaborations.

The professor will have a joint appointment in the Mathematical Institute, Faculty of Sciences, and in the Section Methodology and Statistics, Institute of Psychology, Faculty of Social and Behavioural Sciences (50 % each). The professor is expected to contribute to the administrative tasks in both institutes, in particular regarding statistical science, which is embedded in the Leiden University Center of Statistical Science (LUXs), data science, which is embedded in the Leiden Center of Data Science (LCDS), and artificial intelligence in the SAILS program.

The professor will contribute to the educational programs of the Mathematical Institute and the Institute of Psychology

Voor de advertentietekst:

**Full Professor in Artificial Intelligence and Data Theory**

The professor will develop general theories of data and machine learning techniques to analyse data, in particular for complex data for prediction and classification in the social and behavioural sciences. The professor will make the new techniques available through algorithms and implement these in publicly accessible software (e.g. SPSS, R or Python). Visualisation will be an essential ingredient of the new techniques. Typical important research topics are:

- Description of relationships in high-dimensional data using distances and geometry
- Dimension reduction based on geometry
- Development of advanced machine learning methods
- Optimal transformation of data
- Optimization by minimizing loss functions through dedicated algorithms
- Statistical methods for stability, generalization and model selection
- Efficient algorithms for distance driven methods for prediction and classification.

The professor will have a joint appointment in the Mathematical Institute, Faculty of Sciences, and in the Section Methodology and Statistics, Institute of Psychology, Faculty of Social and Behavioural Sciences (50 % each).