

## PhD Position in Causal Data Science

Jheronimus Academy of Data Science (JADS), Den Bosch (4 year project)

*JADS, a data science institute founded jointly by the University of Tilburg and the Technical University of Eindhoven, is looking for a talented PhD student working on “Causal Data Science”, starting as early as November 2017. The project is funded by JADS internally.*

### Project description

Data science has been tremendously successful in many fields: our recent ability to learn very flexible functions between input features,  $x$ , and output  $y$  using (extremely) large datasets has profoundly altered our ability to (e.g.,) create predictive models and realize computer vision and speech. However, often, the models we create are used to **plan interventions**; in healthcare for example a model mapping features of a person and a treatment might be used to predict survival rates (this is just another model we can learn based on data), and that model is subsequently used to choose the best treatment.

Regretfully, models trained on observational data often do not properly predict outcomes when changes are made to the underlying process. For example; when training a model using the treatment (chemotherapy, radiotherapy, and / or surgery) as input to predict survival rates of breast cancer patients on observational data, one finds that refraining from chemotherapy *improves* survival. However, this is largely illusory; a common cause -- namely the severity of the tumor -- causes both the decision by the specialist to administer chemotherapy *and* influences the survival rate. **The naïve model trained on observational data is erroneous when considering the causal effect of the treatment.** This remains true irrespective of the amount of data available.

In recent years many advances have been made in the field of estimating causal effects based on observational data. These methods however rely on a large number of assumptions, some of which are unreasonable for many data science applications. In this PhD project we would like the PhD student to a) study and compare different contemporary methods of dealing with causality, b) relax a number of the current assumptions and develop a Bayesian approach to causality, and c) develop -- in combination with clinical practitioners -- data science methods that allow for better estimation of causal effects based on a combination of observational and interventional data.

### Work environment and supervision

The position will be supervised by prof. dr. Maurits Kaptein (TiU); principal investigator of the **computational personalization lab** at JADS (see [www.nth-iteration.com](http://www.nth-iteration.com)). The work of the lab finds applications both in healthcare and in (online) marketing. The lab consists of the PI, 5 PhD students, and one post-doc.

The computational personalization lab is part of the **“Data & Humans” theme** (headed by dr. Martijn Willemsen & prof. dr. Maurits Kaptein). Within the theme prof. dr. Eric Postma (TiU), prof. dr. Chris Snijders (TU/e), prof. dr. Wijnand IJsselsteijn (TU/e), and prof. dr. Jack van Wijk (TU/e) will contribute to the

project and supervision. Other research labs within the theme focus on understanding **recommender systems**, **data visualization**, and **modeling networked sensor data**.

The PhD student will be part of the JADS eco-system: both the Data & Humans theme as well as JADS as a whole offer a strong PhD program, frequent scientific seminars by internal and external speakers, and the ability to follow graduate level courses at both TiU and TU/e. JADS offers the unique opportunity to be embedded in a starting research institute, with its associated freedom, while being firmly grounded and supported by experts from two top Dutch Universities.

The positions includes allowances for research expenses, including computer, participant costs, experimental materials, and travel and subsistence costs towards project-specific travel, including conferences. The positions also offers an opportunity to gain experience with teaching, subject to availability and qualifications.

### Requirements

You have or are about to finish a (research) MSc in a quantitative field (Computer Science, Mathematics, Statistics, Physics, AI, etc.). You have an interest in Big Data, Data Science, (Bayesian) Statistics, and Machine Learning, and you like to work both theoretically as well as see your work applied. Preferably you have some knowledge regarding at least one of the contemporary approaches to causal analysis (most notably **Pearl's do-calculus** or **Rubin's causal model**). The project is risky, and we are looking for an independent thinker who is willing to explore new directions.

### Conditions of employment

The position starts in November 2017 but a later date may be considered. JADS offers outstanding employee benefits, as well as an inspiring work environment and exciting campus life. We offer a full time temporary appointment for a period of 4 years. The gross monthly salary will be in accordance with the salary scales for PhD candidates at Dutch universities, i.e. ranging from €2,191 in the first year to €2,801 in the final year. The Collective Labour Agreement for Dutch Universities is applicable. Researchers from outside the Netherlands may qualify for a tax-free allowance equal to 30% of their taxable salary. The university will apply for such an allowance on their behalf. All university employees are covered by the so-called civil servants pension fund (ABP). The university offers very good fringe benefits such as an options model for terms and conditions of employment, excellent reimbursement of moving expenses, excellent technical infrastructure and sports facilities.

### Interested?

If you are interested in this position, please send a short letter motivating your interest and your resume to [m.c.kaptein@uvt.nl](mailto:m.c.kaptein@uvt.nl).

**Contract type:** Temporary, 4 years

**Employer:** JADS, Den Bosch.