

# Daniel T. Soukup

*Data Scientist & Mathematician*

☎ +43 665 65115723  
✉ [daniel.t.soukup@gmail.com](mailto:daniel.t.soukup@gmail.com)  
📄 [danielsoukup.github.io](https://danielsoukup.github.io)  
📧 [danielsoukup](#)  
📌 [danielsoukup](#)

## Strengths

- High proficiency with Python, data structures and algorithms.
- Strong knowledge of Machine and Deep Learning.
- Professional experience with NLP techniques and packages.
- Relentless problem solver; engaging public speaker; dedicated team player.

## Work experience

### Data Scientist @ Mostly AI (2019 - present)

- Creating anonymous synthetic data sets which mimic the statistical properties of the source data while removing privacy concerns; generative deep neural networks.

### Data Scientist @ Uniqa Insurance Group (2019)

- Medical claims classification using novel word-embedding techniques, combined CNN/LSTM architectures; mixed data sources; cloud computing and large-scale model training with GPUs.
- 9% improvement of the deployed models KPI.

### Course Instructor @ U of Toronto/Calgary/Vienna (2014 - 2019)

- Calculus, linear algebra and combinatorics courses; coordinating multi-sectional lecture, 200+ students and 5+ teaching assistant.
- Cryptography outreach for high schools; student mentoring; Daniel B. deLury Teaching Award; excellent student reviews.

### Researcher @ U of Calgary/Vienna (2016 - 2019)

- Plenary talks at conferences; 20 refereed publications in top international journals.
- Major EU funding for independent research projects (MSCA, €224,933).
- Research focus: local/global analysis of large graphs and digraphs; random combinatorial objects; decomposition problems for algebraic and topological structures.

## Education



### University of Toronto (2015)

- PhD in Mathematics
- Ontario Trillium Scholarship



### Eötvös Loránd University (2011)

- MSc in Mathematics
- BSc in Mathematics

### MOOCs (2018 - present)

- Databases and SQL for Data Science [IBM Db2 Warehouse]; Applied Text Mining.
- Object-Oriented Software Design and Architecture, Design Patterns and Service-Oriented Architectures.

### data4good hackathon (2019)

- Text analysis for the geo-distribution of urban greening projects; project website and github repo.
- Cleaning and processing unstructured data; topic modelling and NER; visualization.

## Skills

- Python for DS
- pandas, numpy
- scikit-learn
- Keras, PyTorch
- NLTK, gensim
- spaCy
- GloVe, w2v, BPEmb
- seaborn, folium
- networkx
- Java
- SQL
- git
- MS Azure
- Google Cloud
- graph theory
- logic
- teaching
- research
- communication