

# **Stefan Seman**

ML Engineer, Data Scientist, PhD  
Researcher

Born on July 31<sup>st</sup>, 1988, Belgrade, Serbia  
Contact: Mob: +38162262131

Email: [sseman@outlook.com](mailto:sseman@outlook.com)

LinkedIn: [Link](#)

ResearchGate: [Link](#)

Publications: [Link](#)



## **ABOUT ME:**

Experienced researcher and data scientist with history in higher education and published research. Strong educational background. Skilled in research, data science, machine learning, AI, statistics, software development, biomedicine and sports science. Excellent communication skills. Analytical thinker with pragmatic problem-solving approach.

## **PROFESSIONAL EXPERIENCE:**

### **404 SOLUTIONS**

**SENIOR MACHINE LEARNING ENGINEER AND SOFTWARE ENGINEER**  
**(Oct 2024-Present)**

*Stack: Python, SQL, AWS (Lambdas, ECS, S3, Step Functions...), DynamoDB, PyTorch, NLP, LangChain, LlamaIndex, Data Extraction, GenerativeAI, RAG*

- **AI-Powered Document Summarizer with RAG and LLMs:** Development of a robust document summarizer leveraging Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), and Natural Language Processing (NLP) techniques.
- Large-scale data ingestion, transformation pipelines, ETL/ELT
- Delivered a scalable and high-accuracy solution for extracting insights from unstructured text data.

### **PUBLICIS SAPIENT**

**SENIOR DATA SCIENTIST AND SOFTWARE ENGINEER IN AI, R&D**  
**(Oct 2023-Sept 2025)**

*Stack: Python, SQL, FastAPI, LangChain, Transformers, PySpark, PyTorch, Docker, Azure, KubeFlow, Git, Grafana, Redis, Milvus, RAG, AI Agents, GenerativeAI, AzureAISeach, AzureOpenAI*

- **Document Chatbot:** Working on Chatbot application. Refactoring codebase, improving and optimization of data ingestion pipelines, document retrieval. Evaluation and Fine-Tuning LLMs. Developing new features and improving existing functionalities.
- **AI-Code Copilot:** Development of application from scratch for integration with existing company infrastructure. Involved in designing the architecture and project structure. Implementing and experimenting with different AI models and techniques. Implementing RAG techniques and Fine-Tuning LLMs to improve on model knowledge. Data collection, extraction, transformation and analysis. Creating PoC's for potential external client-based projects and proposition

**VALCON NETHERLANDS****DATA SCIENTIST AND DEVELOPER - SENIOR SPECIALIST****(Jan 2023 - Oct 2023)**

*Stack: Python, SQL, Azure DevOps, Azure Data Factory, Synapse, Data Lake, Docker, Team Leading, Team Management*

- Development of data extraction and document classification services from scratch. Creating solution workflow and architecture. Creating classification models and deep learning algorithms to extract important data from unstructured text. Managing project tasks and priorities, fixing bugs, and developing new features. Mentoring and leading a team of six people.

**DATA SCIENTIST - DEVELOPER****(May 2022 - Jan 2023)**

*Stack: Python, SQL, NLP, Text Summarization, Azure Cognitive Services, Git, Azure LUIS, Azure Data Factory, Data Lake, Front-End (JS, Vue)*

- Development of chatbot summarization service proof of concept. Codebase optimization and refactoring. Fixing bugs and creating Front-End user interface feature. Extracting entities and other important data by Using Azure Cognitive Services and Azure LUIS.

**OPTIMAL SYSTEMS GMBH****DATA SCIENTIST****(Oct 2021 - Apr 2022)**

*Stack: Python, Docker, Machine Learning, Git, Data Analysis, Azure*

- Working on in-house document classification service. Evaluating hyperparameters and model performance, testing, analyzing data working on day-to-day tasks in development.

**BIOENGINEERING RESEARCH AND DEVELOPMENT CENTER - BioIRC****RESEARCHER****(Mar 2020 - Mar 2022)**

*Stack: Python, Statistical Analysis, SPSS, ECG, CPET, Data Analysis, Hypothesis Testing, Data Collection, Research, Biomedicine, Cardiology,*

- [SILICOFCM](#): Goal of the project is development of computational platform for clinical trials. It has received funding from EU Horizon 2020 research and innovation programme. Besides main goal I was also responsible for communicating with subjects and different clinical and research departments, evaluating and analyzing data, doing statistical analysis, conducting ECG and body composition analysis, conducting cardiopulmonary exercise testing, and prioritizing goals in tight-scheduled timelines. Writing scientific papers, communicating with editors and presenting findings at conferences

**UNIVERSITY OF BELGRADE****PHD RESEARCHER****(Oct 2018 - Mar 2022)**

*Stack: Python, PyTorch, SPSS, Machine Learning, Sport Science, Biostatistics, Statistical Analysis, Hypothesis Testing, Data Analysis, ECG, CPET, Exercise Physiology, Research*

- Creating project work-plan methodology, and goals. Collecting and evaluating data, doing statistical analysis and hypothesis testing. Writing papers and publishing research, while communicating with editors and presenting research at conferences. Working on several external projects while writing my Doctoral Thesis.

**UNIVERSITY OF BELGRADE****TEACHING ASSISTANT****(Academic year 2019/2020)**

*Stack: Teaching, Human Biology, Exercise Physiology*

- Teaching assistant at Faculty for Sports and Physical Education/Dept. Biomedical Sciences. Responsible for practical part of bachelor's degree course in two subjects. Exercise Physiology and Sports Medicine with Human Growth Biology. Educating and teaching students with necessary practical skills.

**SPORT AND RECREATION CENTER OBLIQUUS****SPORTS SCIENTIST****(Oct 2014 - July 2018)**

*Stack: Python, SPSS, LabVIEW, Data Analysis, Statistical Analysis, Sport Science, Team Management, Leadership Skills, Presentation Skills*

- Conducting functional diagnostics, while doing data and statistical analysis. Evaluating data about effectiveness of exercise programs and writing reports. Working with amateur and professional athletes. S&C preparation during several distinct training phases. Evaluating and analyzing severity of functional limitations while comparing with current standards and protocols.

**EDUCATION:**

- University of Belgrade, Sport Science, **PhD candidate**
- University of Novi Sad, Sports Medicine, **Master's degree**
- University of Belgrade, Sport Science, **Bachelor's degree**

**COURSES AND CERTIFICATIONS:**

- Microsoft Certified: Azure AI Associate
- Microsoft Certified: Azure Data Fundamentals
- Databricks Academy Accreditation – Databricks Lakehouse Fundamentals
- Databricks Academy Accreditations – Generative AI Fundamentals
- Certificate: Measuring and Maximizing Impact of COVID19 Contact Tracing (Coursera, John Hopkins University, 2021)

**TRAINING:****University of Belgrade · 2018 - 2019**

*LabView, Signal Processing, Signal Acquisition, Motion Analysis*

Design, optimize, build, and develop custom tools for signal processing, noise optimization for improved data gathering, data analysis and evaluation.

**University of Novi Sad · 2015 – 2016**

*Stack: Cardiopulmonary Exercise Testing, ECG, MRI, EEG*

Training for utilization of non-invasive data assessment biomedical tools. Gathering meaningful data that is utilized in biomedical and life-sciences.

**MEMBERSHIPS:**

- ACSM – American College of Sports Medicine
- SEBM – Society for Experimental Biology and Medicine
- Serbian Society for Prevention of Chronic Noncommunicable Diseases
- Serbian Mountain Rescue Service
- Montenegrin Journal of Anthropology in Sports and Physical Education – Editorial Board Member

**LANGUAGES:**

- Serbian (Native)
- English (Fluent)