



# Stefan Seman

ML Engineer | Data Scientist | PhD Researcher

**Phone:** +38162262131

**Website:** [manse404.github.io/webpage](https://manse404.github.io/webpage)

**GitHub:** [github.com/ManSe404](https://github.com/ManSe404)

**ResearchGate:** [ResearchGate](#)

**Email:** [stefan.seman404@gmail.com](mailto:stefan.seman404@gmail.com)

**LinkedIn:** [linkedin.com/in/sseman](https://linkedin.com/in/sseman)

**Scholar:** [Google Scholar](#)

## PROFESSIONAL SUMMARY

ML Engineer, Data Scientist, and PhD Researcher with a strong background in higher education and published research. Specialize in machine learning, AI, generative AI, and software development, with unique expertise bridging data science and biomedicine. Analytical thinker with a pragmatic problem-solving approach and excellent communication skills.

**Core Focus:** LLMs, RAG Systems, Software Development, Document Intelligence

## TECHNICAL SKILLS

<b>Languages</b> Python, SQL, Bash, LabVIEW	<b>Frameworks</b> FastAPI, Flask, PyTorch, PySpark, LangChain, LlamaIndex, Transformers
<b>AWS Cloud</b> Lambda, S3, ECS, EC2, Step Functions, Bedrock	<b>Azure Cloud</b> Data Factory, Synapse, DevOps, Functions, Azure Search, Azure OpenAI, Data Lake, Databricks
<b>Databases</b> DynamoDB, PostgreSQL, Redis, Pinecone, Milvus, PgVector	<b>Tools</b> Docker, Git, Grafana, Jenkins

## PROFESSIONAL EXPERIENCE

<b>Senior ML/AI Engineer &amp; Data Scientist</b> <b>Brainiac</b> <ul style="list-style-type: none"><li>Implemented RAG pipelines and AI agent frameworks for knowledge automation and decision support</li><li>Built scalable data ingestion, preprocessing, and inference pipelines for production environments</li><li>Worked on cloud-native infrastructure on AWS to support secure and scalable deployments</li><li>Collaborated with cross-functional teams and stakeholders to translate business requirements into production AI systems</li><li><b>Stack:</b> Python, AWS, ML, RAG, AI Agents</li></ul>	Mar 2023 - Present
<b>Senior Machine Learning Engineer &amp; Software Engineer</b> <b>404 Solutions</b> <ul style="list-style-type: none"><li>Designed and operated multi-tenant AI systems for enterprise clients</li><li>Delivered document intelligence solutions for automated extraction, classification, and summarization</li><li>Built scalable cloud workflows on AWS (S3, Step Functions, Bedrock, APIs) supporting multiple organizations</li><li>Deployed, monitored, and maintained production systems across distributed client environments</li><li><b>Stack:</b> Python, AI Agents, FastAPI, RAG</li></ul>	Oct 2024 - Present

## Senior Data Scientist & Software Engineer in AI, R&D

Oct 2023 - Sept 2024

### Publicis Sapient

- Led development from POC to fully production-grade systems
- Built enterprise RAG systems with retrieval optimization
- Delivered client-facing solutions and technical propositions
- **Stack:** Python, FastAPI, LangChain, Transformers, PyTorch, Docker, Azure, KubeFlow, Redis, Milvus, AI Agents, Azure OpenAI

## Data Scientist & Developer -> Senior Specialist

May 2022 - Oct 2023

### Valcon Netherlands

- Led and mentored a team of 6 engineers
- Owned design of document processing microservices
- Built data extraction and document classification services from scratch
- **Stack:** Python, Azure DevOps, Azure Data Factory, Synapse, Data Lake, Docker

## Data Scientist

Oct 2021 - Apr 2022

### Optimal Systems GmbH

- Systematic model evaluation and hyperparameter tuning for ML classification
- Data-driven insights and testing/validation of ML systems
- **Stack:** Python, Docker, Machine Learning, Azure

## Researcher

Mar 2020 - Mar 2022

### BioIRC - Bioengineering R&D Center

- Built components for EU Horizon 2020 computational platform (SILICOFCM)
- ECG, CPET, and biomedical data processing; peer-reviewed publications
- **Stack:** Python, Statistical Analysis, SPSS, ECG, CPET, Biomedicine

## PhD Researcher

Oct 2018 - Mar 2022

### University of Belgrade

- Developed research methodology and statistical frameworks
- Applied machine learning to biomedical datasets; published papers internationally
- **Stack:** Python, PyTorch, SPSS, Machine Learning, Biostatistics

## Sports Scientist

Oct 2014 - July 2018

### Sport and Recreation Center Obliquis

- Built algorithms for signal processing and data collection
- Functional testing and performance analysis for athletes
- **Stack:** Python, SPSS, LabVIEW, Data Analysis

## KEY PROJECTS

Full list: [manse404.github.io/webpage/projects.html](https://manse404.github.io/webpage/projects.html)

### AI-Powered Document Summarizer

RAG-based summarization pipeline with LLMs for unstructured corpora  
Python, AWS Bedrock, S3, Step Functions, LangChain, LlamaIndex, RAG

### AI-Code Copilot

Enterprise AI copilot for internal infrastructure and knowledge bases  
Python, Azure OpenAI, LangChain, Milvus, Redis, Fine-Tuning, KubeFlow

### Document Chatbot

Enterprise chatbot with optimized ingestion pipelines and retrieval  
Python, FastAPI, Azure Search, Azure OpenAI, RAG, Docker

### Document Data Extractor Service

Deep learning algorithms for extracting data from unstructured text  
Python, Azure Data Factory, Synapse, PyTorch, Flask, Docker

### Large-Scale ETL Pipeline

Serverless data ingestion and transformation on AWS  
Python, AWS Lambda, S3, Step Functions, DynamoDB, ECS

### SILICOFCM - EU Horizon 2020

Computational platform for in silico clinical trials of cardiomyopathies  
silicofcm.eu  
Python, PyTorch, Biostatistics, ECG, CPET

### Document Classification Service

ML-based document categorization with containerized microservices  
Python, Docker, Azure, Flask, REST API, ML Classification

### Data Extraction Service

ETL platform with Azure Data Factory and Power BI integration  
Python, FastAPI, Azure Data Factory, Synapse, Data Lake, Power BI

### Voice-Enabled Support Chatbot

Conversational AI with speech-to-text and text-to-speech for telephony  
Python, Azure LUIS, Azure Speech Services, Azure Bot Service, Docker

## SELECTED PUBLICATIONS (TOTAL: 10+ PUBLICATIONS | 100+ CITATIONS)

### A Machine Learning Approach to Gene Expression in Hypertrophic Cardiomyopathy

Pavic, J., Zivanovic, M.N., Seman, S., et al.  
Pharmaceuticals (MDPI), 2024

### Ventilatory Efficiency Parameters Outperform Peak VO2 in Monitoring Therapy Effects in HCM

Seman, S., Tesic, M., Popovic, D., et al. (SILICOFCM Investigators)  
European Heart Journal, Vol. 45, Supplement 1, 2024

### Defining the Importance of Stress Reduction in Managing Cardiovascular Disease - The Role of Exercise

Popovic, D., Bjelobrk, M., Tesic, M., Seman, S., et al. (HL-PIVOT)  
Progress in Cardiovascular Diseases, Vol. 70, pp. 84-93, 2022

### Physical activity and exercise as an essential medical strategy for the COVID-19 pandemic and beyond

Seman, S., Srzentic Drazilov, S., et al. (HL-PIVOT Network)  
Experimental Biology and Medicine, Vol. 246, Issue 21, 2021

## EDUCATION

### PHD CANDIDATE

#### Sport Science

University of Belgrade  
2018 - Present

Focus: Machine learning applications in exercise physiology and biomedical signal processing

### MASTER'S DEGREE

#### Sports Medicine

University of Novi Sad

### BACHELOR'S DEGREE

#### Sport Science

University of Belgrade

## CERTIFICATIONS

- ✓ Microsoft Certified: Azure AI Associate
- ✓ Microsoft Certified: Azure Data Fundamentals
- ✓ Databricks Academy: Lakehouse Fundamentals
- ✓ Databricks Academy: Generative AI Fundamentals
- ✓ Johns Hopkins University: COVID-19 Contact Tracing (Coursera, 2021)

## 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

---

**Languages:** Serbian (Native)    English (Fluent)