

Pavol Mulinka

tel.: +34 661816178 | dirección: Barcelona, Catalunya, España | email: mulinka.pavol@gmail.com | LinkedIn: <https://www.linkedin.com/in/mulinka/> | GitHub: <http://github.com/5uperpalo> | Google Scholar: <https://scholar.google.com/citations?user=zsJ4nfoAAAAJ>

Resumen

Soy Data Scientist con más de 7 años de experiencia en el diseño e implementación de soluciones de IA/ML. Estoy especializado en sistemas de Agentic AI y en patrones de diseño basados en datos. Lidero discusiones técnicas de diseño con equipos multidisciplinares y he publicado investigación en telecomunicaciones, NLP y ML.

Habilidades

nombre	nivel	notas
Python	Avanzado	Scikit-learn, Flask, FastAPI, Pandas, SQLAlchemy, SciPy, Pydantic, etc.
JavaScript	Básico	mejora de soluciones, scripting y uso de MOA
SQL	Intermedio	consultas, análisis de datos, administración de bases de datos locales
KQL	Intermedio	consultas, análisis de datos
Prompt engineering	Avanzado	Ollama, Azure OpenAI SDK, Google GenAI SDK
Agentic AI	Intermedio	Langchain, Langsmith
Kubernetes	Intermedio	Microk8s, K3s, Helm charts, administración, proyectos FIREMAN y SUCCESS6G
Docker	Intermedio	dockerización de múltiples soluciones
Data Analysis	Avanzado	EDA, análisis estadístico, modelado de datos
Data Mining	Intermedio	ETL
Machine Learning	Avanzado	modelado predictivo, GBM, cadenas de Markov, Hidden Markov Models, etc.
Deep Learning	Intermedio	GAN, GAIN, transformers, PyTorch
Cloud	Intermedio	AWS, Azure, GCP, cloud computing, computación distribuida
Linux	Intermedio	Bash, scripting, soluciones ligeras
Networking	Intermedio	diseño, implementación, troubleshooting
etc.		

Idiomas

- **Español** – C1/C2
- **Inglés** – C2
- **Checo** – C2
- **Eslovaco** – nativo

Experiencia laboral

- **Cybersecurity Data Analyst**, autónomo, híbrido, Forescout – NL (17/09/2024 – 31/12/2025)

- **Descripción:** Análisis y clasificación de datos procedentes de diversas herramientas y sistemas de seguridad; identificación de patrones, anomalías y potenciales amenazas de seguridad; aplicación de soluciones de Agentic AI y prompt engineering para mejorar la precisión de la clasificación y automatizar tareas del Data Analyst. Desarrollo e implementación de análisis de impacto automatizados de modificaciones en soluciones core.

- **Palabras clave:** applied research, data analysis, data science, machine learning, statistics, text processing, investigation, analytical thinking, collaboration, problem solving, proactivity, Agile
- **Tecnologías:** Python, SQL, KQL, NLP, LLM, AI, Agentic AI, Langchain, Langsmith, Docker, Terraform, Streamlit, AWS, GCP, Azure, Azure AI, OpenAI, Gemini, Google GenAI SDK, JavaScript, Grafana, GitHub, RAG
- **Data Scientist**, autónomo, híbrido, Assetario – SK (01/03/2021 – 26/01/2024), (*Contrato autónomo en paralelo*)
 - **Descripción:** Diseño, análisis e implementación de modelos de machine learning para la predicción del customer lifetime value en aplicaciones móviles y recomendaciones de in-app purchases. Mejora del procesamiento de datos y feature engineering para aumentar las tasas de conversión y la personalización.
 - **Palabras clave:** data analysis, data science, machine learning, deep learning, communication, recommendation system, In-App-Purchases, mobile phones, problem solving, Predicted Lifetime Value (pLTV), programming, statistics, Agile
 - **Tecnologías:** Python, Athena, SQL, AWS, LLM, Hugging Face, Transformers, GBM, Weights & Biases, H2O, DBSCAN, OPTICS, Scikit-learn, PyTorch, MLflow, GitHub
- **F5 Load Balancer Specialist**, autónomo, remoto, Oksystems – CZ (01/12/2021 – 01/10/2022), (*Proyecto autónomo en paralelo*)
 - **Descripción:** Migración de una solución existente de Apache XML firewall y load balancer a F5 para un proyecto del Ministerio de Agricultura.
 - **Palabras clave:** migration, scripting, load balancer, firewall
 - **Tecnologías:** F5, XML, Bash, Postman
- **Data Scientist**, jornada completa, presencial, CTTC – ES (01/10/2020 – 30/09/2024)
 - **Descripción:** Diseño y análisis de enfoques de machine learning para la detección de patrones y anomalías en datasets reales y sintéticos. Liderazgo de proyectos con foco en computación distribuida y análisis de datos a gran escala.
 - **Palabras clave:** research, applied research, data analysis, data science, machine learning, deep learning, project leading, pattern detection, anomaly detection, programming, statistics, analytical thinking, collaboration, research paper writing, investigation, distributed machine learning, federated learning, V2X, 5G, 6G, problem solving
 - **Tecnologías:** Python, SQL, Flask, Celery, Docker, Helm, Kubernetes, Redis, H2O, YOLO, Transformers, GBM, LLM, Hugging Face, Weights & Biases, DBSCAN, OPTICS, Scikit-learn, PyTorch, Istio, Microk8s, K3s, Raspberry Pi, Flower, Kepler, Prometheus, InfluxDB, MinIO, KServe, Knative, MLflow, GitHub, GAN, GAIN, river, deep-river, MOA, Kafka, Airflow, KSQL, Faust, Kubeflow, Zero-to-JupyterHub, MySQL, RAG
- **Python Developer**, autónomo, remoto, Centrales Eléctricas Eslovacas – SK (01/06/2020 – 01/10/2020)
 - **Descripción:** Diseño y desarrollo de una interfaz de comunicación entre el modelo de optimización hidroeléctrica de Slovak Electricity y una GUI de usuario.
 - **Palabras clave:** containerization, in-memory database, web, multiprocessing, unit testing, programación procedural, programación orientada a objetos, refactorización de código
 - **Tecnologías:** Docker, Redis, Flask, Celery, Python, GitHub
- **Data Scientist**, contrato, presencial, NII Tokyo – JP (08/03/2019 – 03/09/2019)
 - **Descripción:** Aplicación de machine learning no supervisado sobre trazas de red para detectar e interpretar patrones desconocidos. Mejora de clustering jerárquico basado en densidad para una mejor interpretación de medidas de red. Refactorización de código para entornos de computación distribuida.

- **Palabras clave:** MAWI, Darknet, anomaly detection, big data, machine learning, deep learning
- **Tecnologías:** PySpark, Python
- **Data Scientist**, contrato, presencial, O2 Telefónica – ES (21/11/2018 – 20/02/2019)
 - **Descripción:** Análisis de la relación entre el nivel socioeconómico y el rendimiento de red. Investigación de posibles discriminaciones en el despliegue de red. Correlación de datos públicos con mediciones de red mediante análisis geoespacial.
 - **Palabras clave:** Lower-layer Super Output Areas (LSOA), data analysis, machine learning, geospatial machine learning
 - **Tecnologías:** QGIS, ArcGIS, GeoPandas, PySpark, Python
- **Data Scientist**, contrato, presencial, AIT Vienna – AT (01/03/2018 – 31/08/2018)
 - **Descripción:** Análisis de ciberseguridad y rendimiento de red. Desarrollo de sistemas de detección y diagnóstico de anomalías. Integración de técnicas de machine learning en plataformas de big data, proyecto [BIG-DAMA project](#).
 - **Palabras clave:** stream-based machine learning, supervised machine learning, unsupervised machine learning, MAWI, Cloud latency, network performance analysis, anomaly detection, big data
 - **Tecnologías:** Cloudera, PySpark, Apache Pig, Hive, Kafka, Elasticsearch, Python
- **Network Engineer**, jornada completa, presencial, CZ y SK (01/08/2007 – 01/03/2018)
 - **Puestos (en orden descendente):**
 - Network Engineer (VSHosting, Praga, CZ)
 - Network Consulting Engineer (Verizon, Praga, CZ)
 - Senior System Engineer (AT&T, Bratislava, SK)
 - HP Radia Specialist (Soitron, Bratislava, SK)
 - HP Monitoring Support Specialist (Soitron, Bratislava, SK)
 - IT VoIP Support Specialist (Soitron, Bratislava, SK)
 - **Descripción:** Diseño, implementación, soporte y documentación de infraestructuras de red, con un fuerte enfoque en seguridad, troubleshooting y rendimiento.
 - **Palabras clave:** networking, VoIP, wireless, VPN, security, firewall, design, troubleshooting, implementation
 - **Tecnologías:** Bash, IOS

Experiencia docente

- **Mentor de “Sakura Science Plan”, NII, Tokio – JP (2019)**
- **Asistente de profesor de “Network Operating Systems”, FEE, CTU, Praga – CZ (invierno 2015, 2016)**
- **Asistente de profesor de “Digital Engineering”, FEE, CTU, Praga – CZ (invierno 2014)**
- **Asistente de profesor de “Communication Processes Control”, FEE, CTU, Praga – CZ (verano 2014 y 2015)**

Proyectos

Investigación

- **SUCCESS-6G: Towards robust, secure and computationally efficient vehicular services in 6G**, [github](#), Co-investigador, CTTC – ES (01/03/2023 - 31/12/2024)
 - **Descripción:** Investigación y desarrollo de sistemas seguros y eficientes de monitorización en tiempo real del estado de vehículos y provisión de fallos para redes 6G usando ML e IoT en el contexto de V2X.
- **FIREMAN (Framework for the Identification of Rare Events via MAchine learning and IoT Networks)**, [github](#), Co-investigador, CTTC – ES (01/02/2020 - 30/04/2023)
 - **Descripción:** Diseño, desarrollo e implementación de un framework basado en big data que cubre todas las fases desde la sensorización y adquisición de datos hasta el análisis estadístico y la toma de decisiones operativas, para identificar, detectar, predecir y prevenir eventos raros en procesos físicos industriales.
- **Practical Privacy-Preserving Data Collection and Utilization using Provable Cryptographic Tools**, Investigador principal, FEE, CTU, Praga – CZ (2019)
 - **Descripción:** Investigación e implementación de métodos de recogida de datos con preservación de la privacidad mediante herramientas criptográficas.
- **Privacy Protection and Machine Learning Utilization of IoT Data in Cloud**, Investigador principal, FEE, CTU, Praga – CZ (2018)
 - **Descripción:** Ingeniería de pipelines de análisis de datos IoT en la nube con foco en privacidad y machine learning.
- **Smart-home IoT and Cloud Telemetry Datamining**, Investigador principal, FEE, CTU, Praga – CZ (2017)
 - **Descripción:** Desarrollo de herramientas de data mining para dispositivos IoT de smart home y sistemas de telemetría en la nube.
- **Cloud Performance Analysis and Improvement**, Investigador principal, FEE, CTU, Praga – CZ (2015–2016)
 - **Descripción:** Análisis de métricas de rendimiento en la nube y propuesta de estrategias de optimización para entornos de cloud computing.
- **Methods Enhancing Work with Cloud Data**, Investigador principal, FEE, CTU, Praga – CZ (2014)
 - **Descripción:** Diseño de metodologías para la gestión y el análisis eficiente de datasets en la nube.
- **Metrics for Automated Detection of Cloud Anomalous Behavior**, Investigador principal, Cisco Systems – CZ (2013)
 - **Descripción:** Desarrollo de métricas y sistemas de detección automatizados para identificar anomalías en entornos cloud.

Open Source

- **pytorch-widedeep**, Colaborador (2021–2024)
 - **Principales contribuciones:** [Deep Imbalanced Regression](#), [New loss functions](#), [Custom Imbalanced DataLoader](#)
 - **Descripción:** Implementación de métodos y funciones de pérdida para el procesamiento de datasets desbalanceados.
- **Wikimedia Scoring Platform Team**, Colaborador externo (2020–2021)

- **Principales contribuciones:** Pavol86: Compress Gensim models, python-mwtext, Tokenization of "word" things for CJK, deltas, revscoring, editquality
- **Descripción:** Lideré investigación e implementación para tokenización de lenguas CJK. Desarrollo de benchmarks comparativos de segmentadores estado del arte y contribuciones a estrategias de compresión de modelos en Gensim.

Educación

- **PhD en Telecommunicaciones**, Facultad de Ingeniería Eléctrica, Czech Technical University, Praga – CZ (2013–2021)
 - **Tesis:** “Hierarchical density-based clustering and interpretation for network measurements”
 - **Descripción:** Desarrollo de pipelines de machine learning de clustering jerárquico basado en densidad para el análisis de datos de red y la detección de patrones desconocidos.
- **Máster (MSc) en Telecommunicaciones**, Facultad de Ingeniería Eléctrica, Slovak University of Technology, Bratislava – SK (2007–2009)
 - **Tesis:** “Classifiers for identification of the speaker”
 - **Descripción:** Diseño e implementación de modelos de clasificación de locutor mediante técnicas de machine learning.
- **Grado (Bc) en Telecommunicaciones**, Facultad de Ingeniería Eléctrica, Slovak University of Technology, Bratislava – SK (2004–2007)
 - **Tesis:** “Measurement of glottal period of human voice”
 - **Descripción:** Investigación detallada sobre la medición e identificación del período glotal para el análisis de señales de voz.

Cursos y certificaciones

- **Cisco Certified Network Associate (CCNA, 640-802)**, (640-553) Implementing Cisco IOS Network Security, (640-460) Implementing Cisco IOS Unified Communications, (640-721) Implementing Cisco Unified Wireless Network Essentials
- **Cisco Certified Design Associate (CCDA)**; (640-863) Designing for Cisco Internetwork Solutions
- **Cisco Certified Network Professional (CCNP)**; (642-901) Building Scalable Cisco Internetworks, (642-812) Building Cisco Multilayer Switched Networks, (642-825) Implementing Secure Converged Wide Area Networks, (642-845) Optimizing Converged Cisco Networks
- **Cisco Certified Internetwork Professional (CCIP)**; (642-642) Quality of Service, (642-611) Multi-protocol Label Switching, (642-661) Border Gateway Protocol
- **Cisco Certified Design Professional (CCDP)**; (300-31) Designing Cisco Network Service, (642-873) Designing Cisco Network Service Architectures
- **Conducting Cisco Unified Wireless Site Survey (CUWSS, 642-731)**
- **Implementing Cisco Edge Network Security Solutions (SENSS, 300-206)**
- **F5 Certified Product Consultant for LTM**; F5-PCL, F50-531
- **F5 Certified Administrator**; (101) Application Delivery Fundamentals, (201) TMOS Administration
- **Juniper Networks Certified Internet Associate EX (JNCIA-EX, JN0-400)**
- **Information Technology Infrastructure Library Foundation in IT Service Management (ITILv3, Foundation)**

- **The Open Group Architecture Framework (TOGAF 9)**
- **ArchiMate 3**
- **Permiso de conducir A+B**
- **Licencia de buceo OWD**

Publicaciones

- Pavol Mulinka and Christou, Ioannis T. and Subham Sahoo and Charalampos Kalalas and Nardell, Pedro H.J.. (Oct 2025). “Towards High-Fidelity and Trustworthy Digital Twins for Fault Diagnosis in Grid Connected Inverters”. In: IEEE Transactions on Dependable and Secure Computing. doi: 10.1109/TDSC.2025.3626846.
- Pavlidis, Nikolaos and Perifanis, Vasileios and Yilmaz, Selim F. and Wilhelm, Francesc and Miozzo, Marco and Efraimidis, Pavlos S. and Koutsiamanis, Remous-Aris and Mulinka, Pavol and Dini, Paolo. (May 2025). “Federated Learning in Mobile Networks: A Comprehensive Case Study on Traffic Forecasting”. In: IEEE Transactions on Sustainable Computing. pp. 576-587. doi: 10.1109/TSUSC.2024.3504242.
- Charalampos Kalalas and Pavol Mulinka and Guillermo Candela Belmonte and Miguel Fornell and Michail Dalgitsis and Francisco Paredes Vera and Javier Santaella Sánchez and Carmen Vicente Villares and Roshan Sedar and Eftychia Datsika and Angelos Antonopoulos and Antonio Fernández Ojea and Miquel Payaro. (2025). “AI-Driven Vehicle Condition Monitoring with Cell-Aware Edge Service Migration”. url: <https://arxiv.org/abs/2506.02785>.
- Zaurin, Javier Rodriguez and Mulinka, Pavol. (Jun 2023). “pytorch-widedeep: A flexible package for multimodal deep learning”. In: Journal of Open Source Software. pp. 5027. doi: 10.21105/joss.05027. url: <https://joss.theoj.org/papers/10.21105/joss.05027>.
- Beattie, Alexander and Mulink, Pavol and Sahoo, Subham and Christou, Ioannis T. and Kalalas, Charalampos and Gutierrez-Rojas, Daniel and Nardelli, Pedro H. J.. (2022). “A Robust and Explainable Data-Driven Anomaly Detection Approach For Power Electronics”. In: 2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). pp. 296-301. doi: 10.1109/SmartGridComm52983.2022.9961002.
- Mulinka, Pavol and Sahoo, Subham and Kalalas, Charalampos and Nardelli, Pedro H. J.. (2022). “Optimizing a Digital Twin for Fault Diagnosis in Grid Connected Inverters - A Bayesian Approach”. In: 2022 IEEE Energy Conversion Congress and Exposition (ECCE). pp. 1-6. doi: 10.1109/ECCE50734.2022.9947986.
- Park, Souneil and Mulinka, Pavol and Perino, Diego. (2022). “A Large-Scale Examination of ”Socioeconomic” Fairness in Mobile Networks”. In: Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies. pp. 248–256. doi: 10.1145/3530190.3534809. url: <https://doi.org/10.1145/3530190.3534809>.
- Mulinka, Pavol and Kalalas, Charalampos and Dzaferagic, Merim and Macaluso, Irene and Rojas, Daniel Gutierrez and Nardelli, Pedro Juliano and Marchetti, Nicola. (2021). “Information processing and data visualization in networked industrial systems”. In: 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). pp. 1–6.
- Casas, Pedro and Mulinka, Pavol and Vanerio, Juan. (2021). “NetSEC at High-Speed: Distributed Stream Learning for Security in Big Networking Data”. In: Data Science–Analytics and Applications. pp. 97–104.
- Wassermann, Sarah and Cuvelier, Thibaut and Mulinka, Pavol and Casas, Pedro. (2020). “Adaptive and Reinforcement Learning Approaches for Online Network Monitoring and Analysis”. In: IEEE Transactions on Network and Service Management.

- Mulinka, Pavol and Casas, Pedro and Fukuda, Kensuke and Kencl, Lukas. (2020). “HUMAN - Hierarchical Clustering for Unsupervised Anomaly Detection & Interpretation”. In: 11th international Conference on Network of the Future.
- Mulinka, Pavol and Fukuda, Kensuke and Casas, Pedro and Kencl, Lukas. (2020). “WhatsThat? On the Usage of Hierarchical Clustering for Unsupervised Detection & Interpretation of Network Attacks”. In: The 5th International Workshop on Traffic Measurements for Cybersecurity.
- Casas, Pedro and Mulinka, Pavol and Vanerio, Juan. (2019). “Should I (re)Learn or Should I Go(on)? Stream Machine Learning for Adaptive Defense against Network Attacks”. In: The 6th ACM Workshop on Moving Target Defense (MTD 2019).
- Mulinka, Pavol and Casas, Pedro and Vanerio, Juan. (2019). “Continuous and Adaptive Learning over Big Streaming Data for Network Security”. In: IEEE International Conference on Cloud Networking CLOUDNET, 2019 International Conference on.
- Wassermann, Sarah and Cuvelier, Thibaut and Mulinka, Pavol and Casas, Pedro. (2019). “ADAM & RAL: Adaptive Memory Learning and Reinforcement Active Learning for Network Monitoring”. In: 15th International Conference on Network and Service Management.
- Mulinka, Pavol and Wassermann, Sarah and Marín, Gonzalo and Casas, Pedro. (2018). “Remember the Good, Forget the Bad, do it Fast: Continuous Learning over Streaming Data”. In: @NeurIPS 2018 Workshops, Workshop on Continual Learning.
- Mulinka, Pavol and Casas, Pedro and Kencl, Lukas. (2018). “Hi-Clust: Unsupervised Analysis of Cloud Latency Measurements through Hierarchical Clustering”. In: IEEE International Conference on Cloud Networking CLOUDNET, 2018 International Conference on.
- Mulinka, Pavol and Casas, Pedro. (2018). “Stream-based Machine Learning for Network Security and Anomaly Detection”. In: Proc. of the Workshop on Big Data Analytics and ML for Data Comm. Net., Big-DAMA@SIGCOMM.
- Mulinka, Pavol and Casas, Pedro. (2018). “Adaptive Network Security through Stream Machine Learning”. In: Proceedings of the ACM SIGCOMM ’18 Posters and Demos.
- Tomanek, Ondrej and Mulinka, Pavol and Kencl, Lukas. (2016). “Multidimensional cloud latency monitoring and evaluation”. In: Computer Networks. pp. 104–120.
- Mulinka, Pavol and Kencl, Lukas. (2015). “Learning from Cloud latency measurements”. In: Communication Workshop (ICCW), 2015 IEEE International Conference on. pp. 1895–1901.
- Kacur, Juraj and Vargic, Radoslav and Mulinka, Pavol. (2011). “Speaker identification by K-Nearest Neighbors: Application of PCA and LDA prior to KNN”. In: Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on. pp. 1–4.

Intereses

escalada, búlder, motos, cerámica, senderismo, naturaleza