

Sébastien Campion

Bogemanstraat, 90 • Wemmel, 1780 • België • sebastien.campion@pm.me • +32 456 02 77 09

Professional Summary

Innovative software engineering leader with over 20 years' experience in developing and managing innovative technological solutions. Recognised expert in artificial intelligence, machine learning, and large-scale systems architecture. With a diverse background in European institutions, academic research, and the private sector, I combine cutting-edge technical expertise with proven skills in managing multidisciplinary teams and steering complex projects.

My strengths include

- Strategic management of large-scale AI and big-data projects
- Designing and implementing robust and scalable software architectures
- Inspirational leadership and development of engineering talent
- Effective collaboration with stakeholders at all organisational levels
- Ability to translate business needs into innovative technology solutions

Passionate about technological innovation and its societal impact, I'm looking to lead high-calibre engineering teams to tackle the most ambitious technological challenges.

Technical Skills

Artificial Intelligence and Machine Learning

- Large Language Models (LLM): EUBERT (more than 35,558 downloads)
- Deep Learning: Classification, Constraint Optimization
- Natural Language Processing (NLP)
- Large-scale Similarity Analysis
- Fraud and Ransomware Detection using ML

Architectures and Platforms

- Large-scale Distributed Systems Design
- Hybrid Architectures (open-source and proprietary)
- Data Science Platforms
- HPC (High-Performance Computing) Infrastructures
- R&D Data Science Lab Design

Software Development

- Languages: Python, C, Java, HTML, JS, CSS.
- Open-source Development
- Software Forge Management (18,000 users, 6,300 projects)
- Agile and DevOps Methodologies

Security and Cryptography

- Secure Messaging Protocols (Signal)
- Malware Analysis

- Applied Cryptography

Data Management

- Big Data: Storage and Processing of Very Large Volumes (400TB+)
- Vector stores
- Multimedia Databases
- Metadata Aggregation and Management

Tools and Environments

- Cloud Environments: AWS, Azure.
- Containerization: Docker, Kubernetes, LXC.
- Version Control Tools: Git, Subversion, Bitbucket
- ML Ops Platforms: DataIku, JupyterLab, NVidia CUDA
- HPC Environments: EuroHPC (slurm), Meso cluster (oar)

Methodologies

- Agile Project Management
 - Design Thinking and Rapid Prototyping
 - MLOps (Machine Learning Operationalization)
-

Key Projects

Safety and Security Analytics Platform (DG TAXUD, European Commission)

Duration: Dec 2019 - June 2022

Role: AI Solution Architect

- Design and operational deployment of an AI-based security analysis platform for the Directorate-General for Taxation and Customs Union.
- Challenges overcome:
 - Integration of multiple heterogeneous data sources.
 - Design of a scalable architecture capable of processing massive volumes of data in real-time.
 - Implementation of AI models for anomaly detection and predictive analysis.
- Results:
 - Successfully deployed in production and utilized by over 80 customs officers and data scientists.
 - Designed and implemented a high-performance system capable of processing 800 messages per second in near real-time.
 - Enabled efficient data-driven decision-making and risk assessment in customs operations.
 - Significantly enhanced the ability to detect and respond to potential customs violations and security threats.

Operational Maintenance of INRIA Forge

Duration: Sep 2007 - Sep 2015 (8 years)

Role: Technical Director

- Management and evolution of the collaborative software development platform for the National Institute for Research in Computer Science and Automation.
- Challenges overcome:

- Scalability: growth from a few hundred to 18,000 users and 6,300 projects.
- Management of 2 TB of source code with high availability and security constraints.
- Continuous integration of new features while maintaining system stability.
- Results:
 - Platform availability maintained at 99.9% over 8 years.
 - Implementation of a robust backup and recovery system, with recovery time < 2 hours with virtualization.
 - 80% reduction in support ticket volume, indicating improved system stability and user experience.

EUBERT: Language Models for the European Parliament

Duration: Oct 2023 - Present

Role: Project Leader and Lead Maintainer

- Development and maintenance of EUBERT language models for automatic language processing in the European legislative context.
 - Challenges overcome:
 - Creation of multilingual models covering the 24 official EU languages.
 - Management of sensitive data and compliance with strict privacy regulations.
 - Performance optimization for real-time use in parliamentary processes.
 - Results:
 - Successful deployment of models capable of processing over 1 million documents per day.
 - 80% reduction in time required for analysis and categorization of legislative documents.
 - Establishment of an open-source platform, fostering collaboration between European institutions.
-

Education

ESIR - ÉCOLE SUPÉRIEURE D'INGÉNIEURS DE RENNES

“Diplôme d’Ingénieur”

Master's degree in Engineering of Computer Science and Telecommunications

Specialization in language and computer system

Rennes, Brittany, France

Sep 2002

UNIVERSITY OF ROUEN

Diplôme d’Études Universitaires Générales - DEUG Physics and Chemistry

Specialization in computer science

Rouen, Normandy, France

Jun 1999



Experience

EUROPEAN PARLIAMENT - DG ITEC / INNOV

Brussels, Belgium

AI Project Leader - Temporary Agent

Oct 2023 - Now

- Support the development and implementation of AI strategic initiatives at the European Parliament, including the AI Governance Framework
- Architect and design enterprise-level AI solutions tailored to the European Parliament's unique needs and objectives
- Collaborate in creating comprehensive staff guidelines for the responsible use of generative AI within the European Parliament.
- Conduct in-depth user requirement analyses and feasibility studies, translating findings into innovative AI-driven solutions
- Manage and mentor a team of data scientists, overseeing the delivery of high-impact studies and proof-of-concept projects across various European Parliament directorates
- Pioneer the development and maintenance of EUBERT Large Language Models (LLMs) and other cutting-edge AI models, showcased on <https://huggingface.co/EuropeanParliament>
- Drive cross-functional collaboration to ensure seamless integration of AI technologies within existing parliamentary processes and systems

INDEPENDENT - EUROPEAN PARLIAMENT - DG ITEC / INNOVATION

Brussels, Belgium

Independent IT Expert Consultant

Oct 2022 - Oct 2023

- System Architecture and Operating Model design for the future AI Center Of Excellence.
- Machine Learning Operationalisation for Legal Tech.
- Providing a "Knowledge as a Service" platform to deploy and share business knowledge rapidly. translated into machine learning models.
- Design thinking & prototyping with deep learning classifier, constraint optimizer, and legal data models.

INDEPENDENT - EUROPEAN AGENCY - OLAF

Brussels, Belgium

Independent IT Expert Consultant

Sep 2022 - Oct 2022

- Design of an R&D Data Science laboratory for new fraud detection technologies.

INDEPENDENT - EUROPEAN COMMISSION - DG TAXUD

Brussels, Belgium

AI Solution Architect

Oct 2021 - Jun 2022

- Production and deployment of the Safety and Secure Analytics platform.

PWC

Luxembourg

Senior Manager / Consultant & Subject Matter Expert

Sep 2019 - Sep 2021

- Manage, design, and implement data science platforms dedicated to large-scale real-time analysis.
- Analysis, comprehension, and technical translation of operational needs.
- Tailored integration of a self-service solution enabling cross-functional organizational model and data governance guarantee.
- Hybrid architecture between free and open-source software and proprietary software to preserve IT sovereignty with horizontal scaling capability.

- Use case implementations on machine and deep learning projects for geocoding prediction and a large-scale similarity search engine.
- Manage and support end-users' training.
- Define an operating model.
- Team leader.

INRIA - French Academic Research Institute
Research Engineer / Technical Director of the High-Security Laboratory

Rennes, France
Sep 2017 - Sep 2019

Research topics were ransomware detection, malware analysis using machine learning techniques, and cryptographic protocols (Signal).

- Contributing to scientific experimentations.
- Design infrastructures and software.
- Requirements analysis and bootstrap projects.
- Establishing roadmap.
- Supervising projects and team-leading.
- Recruiting.
- Leading a team.

MEDIEGO - Startup
Co-Founder and CTO

Rennes, France
Sep 2015 - Sep 2017

The solution developed enables to compute efficient and fast recommendations for emailing based on the user's web navigation.

- Product manager.
- Team leading.
- Market study.
- Pre-sales client analysis.
- Proof of concept.
- Support.
- Recruiting.

INRIA - French Academic Research Institute
Research engineer / Technical director for INRIA forge

Rennes, France
Sep 2007 - Sep 2015

Technical director for INRIA Forge with 18,000 users, 6,300 projects, and 2TB of source code.

Research engineer for efficient exploration and indexing in very large multimedia databases based on Machine Learning algorithms and HPC.

- Contributing to scientific experimentations
- Managing our HPC platform for indexing multimedia contents with 400TB of storage for 500k euros
- Coarse-grain distributed computing with 2000 cores CPU and GPU
- Developing and improving scientific prototypes

- Elaborating corpora / ground-truth for experimentation evaluations - Publishing demo and online web services
- Support.

Orange Labs / France Telecom R&D

Rennes, France

Junior Project Manager for TV/VOD metadata management networks

Sep 2002 - Sep 2007

- Project on metadata aggregation service for multimedia contents distributed on Orange's networks (IP, ADSL, Mobile)
- Mobile TV project: Specify and develop an electronic service guide for mobile terminals through experimentation handled to DVB-H (Digital Video Broadcasting standard).
- Development of TV/VOD portal access by Orange ADSL TV Set Top Boxes.

Thales Group / Thomson Grass Valley

Rennes, France

Junior Project Manager for TV/VOD metadata management networks

Jun - Sep 2001

- Develop a web monitoring tool that centralizes logs generated by various MPEG2 audiovisual appliances (gateway, encoder, multiplexer, decoder, ...)
-

Publications and Activities

French Ministry of Higher Education and Research

Rennes, France

Expert in the research tax credit application process (CIR)

IIEPI - Institut Européen Entreprise et Propriété Intellectuelle

Strasbourg, France

Expert trainer in software development for lawyers.

University of Rennes 1

Rennes, France

Lesson on open source development and Research - Master's degree Research.
Practical work for "Multimedia Machine Learning Tutorials" - Master's degree.

Publications

- **Multi-Device for Signal.** Applied Cryptography and Network Security, Roma, Italy. Keywords: cryptography, secure instant messaging, ratchet, multi-device. Aug 2020
- **Audio thumbnails for spoken content without transcription based on a maximum motif coverage criterion.** Annual Conference of the International Speech Communication Association, Singapour, Singapore. Keywords: spoken content processing, audio mining, motif discovery, summarization, thumbnailing. Sep 2014

- **MODIS: an audio motif discovery software. Show and Tell** Interspeech 2013, Lyon, France.
Keywords: word discovery software, pattern matching, unsupervised learning, dynamic programming. Aug 2013.
- **Efficient Mining of Repetitions in Large-Scale TV Streams with Product Quantization Hashing.** ECCV, Firenze, Italy. Keywords: large scale, product quantization hashing. Oct 2012
- **Texmix: an automatically generated news navigation portal** ICMR - ACM, Hong Kong, China.
Keywords: HTML5, hypervideo, topic segmentation, speech-to-text, named entities, similar images, dynamic resume, geotagging. Jun 2012
- **AVSST- an Automatic Video Stream Structuring Tool** NEM Summit, Barcelona, Spain
Keywords: TV Stream, Indexing, Structuring, Repetition detection, Classification, EPG, metadata, DTW. Oct 2010
- **Service Platform for Converged Interactive Broadband Broadcasting CellularWireless** IEEE Transactions on Broadcasting
Keywords: Multimedia, Digital video broadcasting, Multimedia systems, Cellular networks, Content management. Apr 2007.

Patents

- Procédé d'enregistrement thématique de contenus numériques à diffusion programmée fr 0300882
- Procédé d'enregistrement, avec mise à jour, de contenus numériques à diffusion programmée fr 0300883

PGP signature of this CV available at <https://scamp.fr/sigs/resume-2024-09-10.pdf.sig>