Curriculum Vitae

Last update : 10 Fev 2025

Personal Information

Andrews Cordolino Sobral (dual citizenship : French and Brazilian, 39 years old)

andrewssobral@gmail.com

andrewssobral.wixsite.com/home

www.linkedin.com/in/andrewssobral

@andrewssobral

Summary

AI Architect & Researcher — Ph.D. in Computer Vision & ML — Head of AI at ActiveEon

- 20+ years of software engineering, 10+ years in AI & Computer Vision.
- Specialized in **AI at Scale, HPC+IA, MLOps, Deep Learning, Generative AI, LLM Fine-Tuning & Optimization, and Cloud-based AI solutions**.
- Expertise in **enterprise AI workflows, distributed computing, large-scale automation, fine-tuning of LLMs, and cloud-based AI deployment**.
- Passionate about **customer-facing AI solutions, AI consulting, research, open-source contributions, and AI innovation**.

Professional Background

Head of Artificial Intelligence (AI), ActiveEon (2019 – Current) — Paris, France

Leading the AI division at ActiveEon, driving innovation in AI at Scale, HPC+IA, and MLOps. Overseeing research, development, and deployment of enterprise-grade AI workflows, optimizing AI automation and orchestration.

Key Responsibilities :

- Lead a team of 4 PhD-level AI & ML researchers.
- Develop and scale end-to-end AI workflows for enterprise applications.
- Architect parallel & distributed AI solutions, optimizing multi-node and multi-GPU execution.
- Design large-scale AutoML frameworks for hyperparameter optimization and neural architecture search.
- Spearhead AI automation, accelerating adoption of AI-driven processes.

Major Contributions :

- Developed **ProActive AI Orchestration**, a scalable AI automation platform.
- **Led the development of the MLOps Dashboard**, a centralized monitoring hub for AI model deployment, performance tracking, and resource management.
- Developed monitoring solutions for **any AI model**, including LLMs and Generative AI, ensuring real-time observability, drift detection, and infrastructure health tracking.
- Designed **custom inference tracking** to provide insights into model resource utilization, inference rates, and latency across diverse AI models.
- Created distributed AutoML workflows enabling large-scale hyperparameter optimization.
- Implemented AI solutions for **Thales Alenia Space (Satellite anomaly detection)**, **SAFRAN (Aircraft engine diagnostics)**, and **Hydro Quebec (Power network analytics)**.

Technologies & Tools :

- **Programming :** Python, C++, MATLAB.
- **AI/ML Frameworks :** PyTorch, TensorFlow, Keras, OpenCV, NVIDIA RAPIDS.
- **MLOps & Deployment :** Kubernetes, Apache Spark, Docker, HPC, Ray, Triton Inference Server .
- **Monitoring & Visualization :** Grafana, Kibana, Prometheus, TensorBoard, Streamlit.

Previous Roles at ActiveEon :

- **Senior AI Research Engineer (2017 – 2019)** : Led research in AI automation and machine & deep learning-driven optimization.

Industry Experience :

- 20+ years in software engineering, architecture, and design.
- Extensive experience in AI, ML, and Computer Vision since 2010.

Academic Background

Ph.D. in Computer Vision and Machine Learning, University of La Rochelle, France (2013 – 2017)

Research focused on advanced matrix and tensor methods for robust low-rank/sparse representation, subspace learning, and incremental learning of multidimensional streaming data. Key applications included :

- **Moving Object Detection** : Developed robust background-foreground separation techniques for multi-spectral and multi-featured video sequences.
- **Subspace Clustering** : Applied matrix/tensor decomposition for human activity recognition using 3D skeletal data.
- **Intelligent Video Analytics** : Worked on object detection, segmentation, and tracking for action recognition and behavior classification.

Embedded AI and Robotics
Beyond theoretical research, I developed real-time computer vision and image processing algorithms for embedded platforms, optimizing deep learning models for :

- **Edge AI Systems** : NVIDIA Jetson AGX Xavier, TX2, Xavier NX, Nano, Raspberry PI, and PandaBoard.
- **Autonomous Systems** : Embedded AI for small-scale self-driving cars and robotic applications.

Research and Development Timeline :

- Since 2010 – Computer Vision, Image Processing, and Machine Learning.
- Since 2013 – Matrix/Tensor Decomposition and Optimization.
- Since 2015 – Deep Learning and Neural Networks for real-time AI applications.

Research Interests

My research spans across multiple domains at the intersection of AI, mathematics, and real-world applications, with a focus on :

- **Computer Vision & Image Processing** – Object detection, segmentation, tracking, and action recognition in video sequences.
- **Machine Learning & Deep Learning** – Neural networks, generative AI, and automated model optimization.
- **Matrix & Tensor Decomposition** – Low-rank and sparse modeling, dimensionality reduction, and subspace learning for high-dimensional data.
- **Applied Mathematics & Optimization** – Algorithmic improvements for scalable AI, hyperparameter optimization, and neural architecture search.
- **Edge AI & High-Performance Computing (HPC)** – AI at scale, distributed training, multi-GPU deep learning, and embedded AI.
- **Enterprise AI & AI Solutions Architecture** – Designing scalable AI pipelines, optimizing AI performance, and cloud-based deployments.
- **Generative AI & LLMs** – Fine-tuning and optimizing large-scale models for real-world AI applications.

Experience

With over two decades of experience in software development and AI, my expertise spans multiple domains :

- **Software Engineering & Architecture (Since 2001)** – Design, development, and optimization of desktop and web applications, database management, and enterprise solutions.
- **Computer Vision & Machine Learning (Since 2010)** – Image/video processing, object detection, segmentation, and behavior analysis.
- **Matrix & Tensor Methods (Since 2013)** – Low-rank and sparse decomposition, subspace learning, and multidimensional data processing.
- **Deep Learning & Neural Networks (Since 2015)** – CNNs, generative AI, MLOps, and large-scale distributed AI training.

Programming Skills

Programming Languages : Python, C++, MATLAB, Java, PHP
AI & Scientific Computing : PyTorch, TensorFlow, Keras, OpenCV, scikit-learn, H2O, Caffe, NVIDIA RAPIDS, Tensor Toolbox
MLOps & Deployment : Docker, Kubernetes, Apache Spark, Ray, Dask, FastAPI, Flask, Triton Inference Server
LLM Optimization & Serving : LoRA, AutoAWQ, FlashAttention-2, vLLM, Text Generation Inference (TGI), Hugging Face Transformers
Big Data & Databases : SQL, PostgreSQL, MongoDB, Redis, Elasticsearch
Cloud AI & Scalable Deployment : AWS (SageMaker, Lambda), Azure AI, Google Cloud AI, Kubernetes, Apache Spark, Ray, FastAPI

Languages

Portuguese : Native proficiency.
French : Fluent (professional and academic).
English : Advanced (professional and technical communication).

Academic Projects

A collection of my academic projects is available on my GitHub and Behance profiles :
github.com/andrewssobral
Bē be.net/andrewssobral

I have also developed and maintain two open-source libraries for background/foreground separation in videos, widely used in computer vision research and applications :

- **BGSLibrary** (C/C++) : A collection of background subtraction algorithms for moving object detection.
- **LRSLibrary** (MATLAB) : A framework for robust low-rank and sparse decomposition techniques applied to background modeling.

Journal Reviewer

I serve as a peer reviewer for **10+ high-impact AI journals**, including :
Elsevier : CVIU, IVC, PRL, NC, Information Fusion
Springer : NCA, CC, FITEE, JIVP
IEEE : TIP, TNNLS, TCSVT
Other Notable Journals : MDPI Sensors, JEI, JOSA-A, PLOS ONE
For a complete list, visit my Publons profile : publons.com/author/619460/andrews-sobral

Google Scholar

My research contributions have been widely cited in the fields of computer vision, machine learning, and artificial intelligence. My research has been cited in **top-tier AI and Computer Vision conferences/journals**, reflecting my contributions to the field.
Google Scholar Metrics
Citations : 2757 — **h-index :** 17 — **i10-index :** 18
For a full list of my publications, visit : scholar.google.fr/citations?user=0Nm0uHcAAAAJ For a complete list of my published papers, please refer to Page 3 of this document.

Education

Sep, 2013 - May, 2017 Ph.D. on Computer Vision and Pattern Recognition, University of La Rochelle, France (European Doctorate label). **Thesis** : "Robust Low-Rank and Sparse Decomposition for Moving Object Detection : From Matrices to Tensors".

Jun, 2014 - Aug, 2014 and Jun, 2015 - Aug, 2015 Ph.D. research stage at Computer Vision Center (CVC), Barcelona, Spain. **Research focus** : 1) Feature Extraction and Selection for Video Action Recognition, and 2) Robust Subspace Clustering of Human Activities from 3D Skeletal Data.

Mar, 2010 - Dec, 2012 M.Sc. in Mechatronics Engineering with focus on Computer Vision and Pattern Recognition, Polytechnic School, Federal University of Bahia, Brazil. **Dissertation** : "Highway Traffic Congestion Classification using Holistic Properties". For more info, please see HTCC project.

Jun, 2004 - Jun, 2009 B.Sc. in Computer Engineering with focus on Mobile Robotics at AREA1 Engineering School, Brazil. **Project** : "RAVE : A robot platform for research and development in mobile robotics using r/c car". For more info, please see RAVE 1.0 and RAVE 2.0.

Professional experience

May, 2017 – Current — Paris Area, France

Head of AI at ActiveEon (from July 2019 to present).

Senior AI Research Engineer (from May 2017 to July 2019).

- Type of contract : CDI (permanent employee contract).
- Leading AI initiatives at ActiveEon, focusing on AI at Scale, HPC+IA, and MLOps. Responsible for the development and orchestration of AI workflows, including large-scale distributed computing and automation.
- Key Responsibilities :
  - Lead a team of 4 PhD-level AI & Machine Learning researchers.
  - Develop and optimize AI workflows for parallel and distributed execution.
  - Architect multi-node and multi-GPU AI solutions for enterprise applications.
  - Design and implement large-scale AutoML frameworks for hyperparameter optimization and neural architecture search.
  - Oversee AI-driven automation and orchestration for enterprise clients.
  - Developed and deployed the MLOps Dashboard, providing real-time observability, model server monitoring, and system-wide AI performance tracking.
  - Enabled monitoring and deployment across all AI model types, including LLMs, Generative AI, and traditional ML/DL models.
  - Built a real-time Drift Notification System, allowing proactive detection of data and model performance drift for AI applications.
- Key Projects & Use Cases :
  - ProActive AI Orchestration – Designed an end-to-end AI automation platform for scalable deployment.
  - MLOps Dashboard – Led the architecture and implementation of a centralized monitoring system for AI deployments at scale.
  - Model Servers and Monitoring – Developed a scalable solution for monitoring model inference performance, server health, and resource utilization.
  - Drift Detection System – Implemented real-time alerts for operational and data drift, enhancing AI reliability across all domains.
  - LLM & GenAI Integration – Optimized MLOps capabilities for modern LLMs and Generative AI models, addressing their scalability, monitoring, and infrastructure needs.
  - Thales Alenia Space (TAS) – Anomaly detection using CNNs on hybrid circuit images from satellite hardware.
  - SAFRAN Aircraft Engines – Developed AutoML-based anomaly detection models for aircraft engine sensors.
  - Hydro Quebec (IREQ) – Machine learning and data analytics for power networks.

— Desjardins Data Cup 2018 – Led feature engineering and ML model development for financial data prediction.

Technologies & Tools :

— Python, C++, MATLAB, OpenCV, PyTorch, TensorFlow, Keras, NVIDIA RAPIDS.
— Kubernetes, Apache Spark, Docker, HPC, Distributed AI frameworks.
— Grafana, Kibana, Prometheus, TensorBoard for real-time monitoring & visualization.
— Streamlit, Hugging Face, Triton Inference Server for AI model serving and monitoring.

Jan, 2024 – Current — Remote, USA

AI Consultant at Ogre.run.

— Led the development and optimization of Large Language Models (LLMs) for specialized use-cases, focusing on fine-tuning open-source models like Llama2 7B and Mistral 7B.
— Key Responsibilities :
  — Designed AI solutions for real-world enterprise applications, focusing on LLM fine-tuning, model optimization, and efficient AI deployments.
  — Provided AI consulting for customers to integrate and optimize AI workflows.
  — Developed end-to-end LLM pipeline from data collection to model serving
  — Implemented efficient fine-tuning strategies using LoRA
  — Optimized model performance using FlashAttention-2 and AutoAWQ quantization
  — Contributed to the open-source miniogre project
— Technologies & Tools :
  — LLM Development : Hugging Face Transformers, LoRA, SFT
  — Model Optimization : AutoAWQ, FlashAttention-2
  — Deployment : vLLM, Text Generation Inference (TGI), Docker

May, 2020 – Current — São Paulo Area, Brazil

Associate Researcher at Advanced Institute for Artificial Intelligence.

— The Advanced Institute for Artificial Intelligence is a consortium of outstanding researchers in different areas of AI aiming to tackle socially relevant problems..

Jun, 2015 – Apr, 2016 (11 months) — Orléans Area, France

Computer Vision and Deep Learning Consultant at CamPark Solutions.

— CamPark Solutions is a French start-up company focused on computer vision solutions for smart city.
— I worked as a Computer Vision and Deep Learning consultant for the automated parking lot classification project. My main research activity was to develop a C++ algorithm based on deep convolutional neural networks for image classification. The algorithm was optimized to run in a Raspberry PI board. Main tools and libraries : Docker, OpenCV, Boost, Caffe, NVIDIA DIGITS.

May, 2013 – Oct, 2015 (2 years 6 months) — São Paulo Area, Brazil

Computer Vision Consultant at GeekSys.

— GeekSys is a Brazilian start-up company focused on computer vision and big data analytics.
— I worked as a Computer Vision consultant for the Heatmaper project.

Previous Experience

Software Consultant, XYZTEMAS, Brazil (2010 – 2013) Developed enterprise Java-based applications and mobile apps using HTML5 and Sencha framework.

Software Consultant and Systems Analyst, GaranteMed, Brazil (2009 – 2013) Designed and developed J2EE-based enterprise applications for the healthcare sector.

Software Consultant, GrupoLM, Brazil (2011 – 2012) Developed a vehicle fleet management system integrating J2EE and SAP.

Senior Software Architect Consultant, Indra/Politec, Brazil (2010 – 2011) Led software architecture for the SEIA project, an environmental information system. Defined system design constraints, researched SOA technologies, and developed software components.

Systems Analyst and Software Developer, IT-Integration, Brazil (2000 – 2008) Designed and developed web and desktop applications for telecommunications, specializing in Java, PHP, J2EE, and database systems (SQL, MySQL, PostgreSQL).

Additional Academic Information

Invited Talks

2020 "Low-rank and sparse tools for moving object detection", Amphitheatre, IPGG (Institut Pierre-Gilles de Gennes), 6 rue Jean Calvin 75005 Paris, France. Slideshare presentation.

2016 "Recent advances on low-rank and sparse decomposition for moving object detection", Workshop/Atelier : Enjeux dans la détection d'objets mobiles par soustraction de fond, Reconnaissance de Formes et Intelligence Artificielle (RFIA), Slideshare presentation.

2014 "Inside video action recognition frameworks", Lab. XLIM-SIC, Poitiers, France.

Event Participation

2015 "Programmation Python pour le calcul scientifique", 27-29 May, La Rochelle, France.

2014 "Traitement et analyse statistique des données massives (BigData)", 19-21 November, Poitiers, France.

Publications

ArXiv

2023 Caroline Pacheco do Espírito Silva, **Andrews Cordolino Sobral**, Antoine Vacavant, Thierry Bouwmans, José A. M. Felippe De Souza. "Discovering Local Binary Pattern Equation for Foreground Object Removal in Videos". 2023. (ArXiv).

2021 Caroline Pacheco do Espírito Silva, José A. M. Felippe De Souza, Antoine Vacavant, Thierry Bouwmans, **Andrews Cordolino Sobral**. "Automated Mathematical Equation Structure Discovery for Visual Analysis". 2021. (ArXiv).

Chapters

2015 **Sobral, Andrews**; Bouwmans, Thierry; Zahzah, El-hadi. "LRSLibrary : Low-Rank and Sparse tools for Background Modeling and Subtraction in Videos". Chapter on the handbook "Robust Low-Rank and Sparse Matrix Decomposition : Applications in Image and Video Processing", CRC Press, Taylor and Francis Group, 2015. (GitHub).

2014 **Sobral, Andrews**; Bouwmans, Thierry. "BGS Library : A Library Framework for Algorithm's Evaluation in Foreground/Background Segmentation". Chapter on the handbook "Background Modeling and Foreground Detection for Video Surveillance", CRC Press, Taylor and Francis Group, 2014. (GitHub).

Journals

2016 **Sobral, Andrews**; Zahzah, El-hadi. "Matrix and Tensor Completion Algorithms for Background Model Initialization : A Comparative Evaluation". Pattern Recognition Letters (PRL), Special Issue on Scene Background Modeling and Initialization. (DOI), (GitHub).

2016 Gong, Wenjuan; Zhang, Xuena; Gonzalez, Jordi; **Sobral, Andrews**; Bouwmans, Thierry; Tu, Changhe; Zahzah, El-hadi. "Human Pose Estimation from Monocular Images : A Comprehensive Survey", Sensors, 2016. (DOI).

2016 Bouwmans, Thierry; **Sobral, Andrews**; Javed, Sajid; Ki Jung, Soon; Zahzah, El-Hadi. "Decomposition into Low-rank plus Additive Matrices for Background/Foreground Separation : A Review for a Comparative Evaluation with a Large-Scale Dataset", Computer Science Review, 2016. (ArXiv) (DOI).

2014 **Sobral, Andrews**; Vacavant, Antoine. "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos". Computer Vision and Image Understanding (CVIU), 2014. (DOI)

International Conferences and Workshops

2015 **Sobral, Andrews**; Javed, Sajid; Ki Jung, Soon; Bouwmans, Thierry; Zahzah, El-hadi. "Online Stochastic Tensor Decomposition for Background Subtraction in Multispectral Video Sequences". ICCV Workshop on Robust Subspace Learning and Computer Vision (RSL-CV), Santiago, Chile, December, 2015. (DOI), (GitHub).

2015 Javed, Sajid ; Ho Oh, Seon ; **Sobral, Andrews** ; Bouwmans, Thierry ; Ki Jung, Soon. "Background Subtraction via Superpixel-based Online Matrix Decomposition with Structured Foreground Constraints". ICCV Workshop on Robust Subspace Learning and Computer Vision (RSL-CV), Santiago, Chile, December, 2015. (DOI).

2015 **Sobral, Andrews** ; Bouwmans, Thierry ; Zahzah, El-hadi. "Comparison of Matrix Completion Algorithms for Background Initialization in Videos". Scene Background Modeling and Initialization (SBMI), Workshop in conjunction with ICIAP 2015, Genova, Italy, September, 2015. (DOI).

2015 **Sobral, Andrews** ; Bouwmans, Thierry ; Zahzah, El-hadi. "Double-constrained RPCA based on Saliency Maps for Foreground Detection in Automated Maritime Surveillance". Identification and Surveillance for Border Control (ISBC), International Workshop in conjunction with AVSS 2015, Karlsruhe, Germany, August, 2015. (DOI).

2015 Javed, Sajid ; **Sobral, Andrews** ; Bouwmans, Thierry ; Ki Jung, Soon. "OR-PCA with Dynamic Feature Selection for Robust Background Subtraction". In Proceedings of the 30th ACM/SIGAPP Symposium on Applied Computing (ACM-SAC), Salamanca, Spain, 2015. (DOI).

2014 Javed, Sajid ; Ho Oh, Seon ; **Sobral, Andrews** ; Bouwmans, Thierry ; Ki Jung, Soon. "OR-PCA with MRF for Robust Foreground Detection in Highly Dynamic Backgrounds". In the 12th Asian Conference on Computer Vision (ACCV 2014), Singapore, November, 2014. (DOI).

2014 **Sobral, Andrews** ; Baker, Christopher G. ; Bouwmans, Thierry ; Zahzah, El-hadi. "Incremental and Multi-feature Tensor Subspace Learning applied for Background Modeling and Subtraction". International Conference on Image Analysis and Recognition (ICIAR'2014), Vilamoura, Algarve, Portugal, October 2014. (DOI), (GitHub).

2013 **Sobral, Andrews** ; Oliveira, Luciano ; Schnitman, Leizer ; Souza, Felippe De. "Highway Traffic Congestion Classification using Holistic Properties". **(Best Paper Award)** In 10th IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA'13), Innsbruck, Austria, February, 2013. (DOI), (Project page).

National Conferences and Workshops

2014 Silva, Caroline ; **Sobral, Andrews** ; Vieira, Raissa Tavares. "An automatic facial expression recognition system evaluated with different classifiers". X Workshop de Visão Computacional (WVC'2014), Uberlândia, Minas Gerais, Brazil, October, 2014.

2014 Silva, Caroline ; **Sobral, Andrews** ; Vieira, Raissa Tavares. "Facial expression recognition in static images by generalized procrustes analysis". X Workshop de Visão Computacional (WVC'2014), Uberlândia, Minas Gerais, Brazil, October, 2014.

2013 **Sobral, Andrews**. "BGSLibrary : A Background Subtraction Library". IX Workshop de Visão Computacional (WVC'2013), Rio de Janeiro, Brazil, June 6, 2013. (GitHub).

2010 **Sobral, Andrews** ; Silva, Caroline ; Júnior, Carlos A.V.V. ; Oliveira, Fabrício M. "RAVE : A robot platform for research and development in mobile robotics using R/C car". CONEM 2010 - VI National Congress of Mechanical Engineering, Campina Grande, PB, Brazil, August 21, 2010. (Project page).