

Jose M. Rivera-Rubio

Curriculum Vitae

+44(0) 7922783534
josemrivera@gmail.com
joserivera.org
jriverarubio
jmrr
Skype: josem.rivera

Summary

I am a Data Science professional that can deliver solid engineering solutions across many areas of the Data and AI spectrum. I can work as a Data Scientist, Data Engineer or Machine Learning Engineer, but more importantly, I bring experience from each of these disciplines to be able to accomplish end-to-end Data Science solutions. From developing Machine Learning algorithms to creating complex, Big Data-ready ETL/ELT pipelines to deploying Machine Learning as a Service, I can take a project from idea to service and deploy this to the cloud.

Experience

September 2020–Present **Lead Machine Learning Engineer, OutMatch HCM, London, UK.**

LaunchPad Recruits was acquired by OutMatch in September 2020. The Machine Learning and Data Science offering that our Data Team built at LaunchPad between 2016 and 2020 played a key role in the acquisition. Once integrated in the new company, my mission is to grow the team and OutMatch's Data products portfolio so we can provide a closed-loop or *Continuous Intelligence* across the company to become an *Intelligent Enterprise*. To achieve that, the projects I'm working on are:

Projects:

- **Predict.** Expansion of *Predict* to build **generalised models** that apply to different customers recruiting for the same profiles. This will allow us to apply our models beyond volume customers and work on improving our feature set and models' performance by leveraging larger datasets. *Technologies:*
 - *Machine Learning:* Python scientific stack: scikit-learn, Numpy, Pandas, NLTK, spaCy, gensim (NLP), librosa, Deep Learning frameworks: Keras, TensorFlow, PyTorch (experimental prototypes), ...
 - *Application Layer:* Flask, PostgreSQL, microservice architecture, serverless (AWS Lambda, API Gateway, SQS queues, DynamoDB), S3, structlogs, AWS CloudWatch.
 - *Data Science/Data Engineering tooling:* MLflow, AWS Batch, Apache Airflow, AWS Sagemaker.
 - *Devops/DataOps, CI:* Docker, AWS CloudFormation and Terraform infrastructure-as-code, GitHub, Circle CI.
- **BI Analytics and Data Engineering platform** Analytics unification (ETLs and BI solutions) of each of OutMatch's integrating companies. I am leading the efforts to consolidate a single data pipeline and BI solution across the enterprise that can be incorporated into our Machine Learning Platform. *Technologies:*
 - Data Pipelines: Change Data Capture, data streaming, AWS Kinesis,
 - Data Lake/Warehouse: CSV, Apache Parquet, S3, Snowflake, AWS Redshift
 - BI layer: Looker, Yellowfin, Sisense
- **Machine Learning Lifecycle** At LaunchPad I set up MLflow for our Machine Learning Lifecycle platform and built an abstract *ML lifecycle service* within our app to log all our simulations' data to produce rich performance and tracking analytics and achieve reproducibility of our results. My aim is to close the loop and link our performance and drift reports with our model build and analysis engine and build a continuous learning service that can reduce the manual intervention so we can focus on analysis and supervision whilst delivering value to our clients.

September 2016– **Machine Learning Engineer and Architect**, *LaunchPad Recruits (now OutMatch HCM)*, London, UK.

September 2020 At LaunchPad Recruits I have the role of being a hybrid between a Data Engineer and a Data Scientist, writing code to establish flexible data pipelines that can accommodate from datasets of few GB to Big Data scales and make predictions with state-of-the-art algorithms. I also work on prototyping and productionising Machine Learning and Pattern Recognition algorithms for multiple data signals of all nature: text, audio and video. I use cutting-edge technology from systems and devops to production code of complex algorithms in the cloud. As a perfectionist, I love using best practices in code management and delivery and Agile project management.

- **Projects:**

- **Predict**, a cloud-based predictive analytics platform written in Python that combines machine learning and human decision making to ensure a consistent and quicker candidate sifting whilst reducing both the risk of missing hidden gems & hiring unsuitable candidates.
- **Verify**, a recruitment analytics platform written in Node.js. It is a tool that provides insights on reviewer scoring consistency and bias aiming to ensure fairer, more consistent hiring decisions. *Technologies:*
 - Application Layer: Node.js (ES6), Express
 - Data Layer: PostgreSQL, Sequelize ORM, ironMQ.
 - Testing, Devops, CI: Mocha, testem, Docker, GitHub, Circle CI
- **Project management**, I have led the implementation of a modified version of the Scrum agile process within a Data Science team. Given its particularities, and the added difficulties of having very specialised team members and a distributed team, it has taken some iterations to polish but my team has achieved faster shipping timelines, less frustrations and a more lean and iterative software delivery.

2015–2016 **Data Engineer Lead**, *Intern Avenue*, London, UK.

I was the Data Engineer Lead at Intern Avenue, where I was in charge of multiple Big Data and machine learning projects within the company: from recommender systems to NLP, personality insights and data visualisation. At Intern Avenue I dedicated my efforts at developing data analysis, feature engineering and machine learning algorithms that helped provide a better matching between candidates looking for an internship or a graduate job and top employers.

- **Technologies:** recommender system using Apache PredictionIO (Spark, Spark MLlib, HBase), Python for data analysis, Apache Zepellin using PySpark, SparkSQL, MySQL for Business Intelligence. Devops using Ansible, RPM packaging, Docker and deployments in SoftLayer (IBM Cloud) bare metal servers.

2014 **Summer Intern**, *THE MATHWORKS*, Cambridge, UK.

Development/Application Engineering team. Developed MATLAB and Simulink models that interfaced with a variety of sensors and actuators connected to target hardware (Arduino and Raspberry Pi) using I2C as communication protocol and automatic C code generation for deployment on target hardware:

- Contributed the C code to interface the I2C Linux kernel tools allowing for code generation and deployed use on Raspberry Pi.
- Added MATLAB and Simulink support to 6 different sensors and actuators.
- Developed MATLAB OOP classes and examples for individual sensor/actuator operation.
- Created Simulink sensor and actuator integration models: 10-DOF inertial measurement unit (IMU) sensor fusion model with 3D visualization, pan and tilt unit using microsensors.
- Developed prototypes of hardware camera stabilisation using microsensors and IMU readings, and object tracking camera using object tracking algorithms and I2C microsensors.

I also expanded two current examples from the Computer Vision toolbox providing a more robust object tracking technique and adding face tracking to the face recognition demo.

- October 2011 **PhD researcher**, *Imperial College London*, PI: Dr Anil A. Bharath.
– May 2015 During my PhD I have been involved in the following projects:
- Hippocampal models for localization: Proposed a biologically inspired method that uses neural networks to improve the visual recall of already visited places.
 - Appearance-based indoor localization from wearable cameras: Developed a state-of-the-art pipeline for appearance-based localization in indoor spaces. This included the development of new descriptors based on filtering techniques that improved the performance on very ambiguous localization data, i.e. indoor corridors.
 - Sparse coding for localization: Adapted successful techniques in image compression and denoising to create “dictionaries of places” for learning a representation of different routes inside a building.
 - The RSM dataset of “visual paths” for benchmarking visual localization algorithms. It contains more than 1.8 km of video sequences captured with mobile and wearable devices along 6 indoor locations. I developed a benchmark to evaluate different methods using C++ and MATLAB.
 - A house-hold products dataset (the SHORT-100 dataset). SHORT-100 contains more than 150,000 images capturing usage particularities of blind and partially sighted people.
 - Picture This... project. Developed an Android app and C++ (OpenCV) backend to provide user localization based on image-matching against a previously acquired dataset.
- October 2008 **Research Assistant**, *The Minerva Project: Vodafone Spain, regional Government and University of Seville R&D project*, PI: Dr Alejandro Carballar.
– February 2010 This initiative funded my final year project. I researched on the IMS protocols and worked on R&D project monitoring and management. I was able to work in a multidisciplinary environment, studying the applications of my main research topic in other areas of interest such as Bioengineering or Social Services.
- March – July 2010 **IT Coordinator (Erasmus intern)**, *Auto ID Services Ltd.*, St. Helens, UK.
Internet and VoIP network and systems administration, mobile computing, IT security. Training and induction of new staff, documentation and IT strategy. Complete development of the new website.

Education

- 2011 – 2016 **Ph.D. in Computer Vision**, *Imperial College London*, Awarded 01/07/2016.
My thesis work focused on the development of Computer Vision and Pattern Recognition algorithms for appearance-based visual localization. Particularly, my main interest was in biologically inspired methods that could also work in other fields such as object recognition. My research also involved extensive use of Machine Learning algorithms both for supervised and unsupervised learning (e.g. neural networks, SVMs, multivariate regression, LDA, clustering methods, etc.).
Additionally, I have extensive experience in developing benchmarks for appearance-based methods. I have contributed the benchmark pipeline code and novel datasets for hand-held object recognition (SHORT dataset) and visual localization from wearable cameras.
- Supervisor:* Dr Anil A. Bharath
- 2012 **International Computer Vision Summer School**, *University of Catania, Italy*.

- 2010 – 2011 **MSc Biomedical Engineering**, *Imperial College London*, Overall mark of 71% awarded the top mark (A) in 7/11 subjects.
Dissertation: Features for the visual biopsy of polyps. Designed image processing algorithms to extract features of endoscopy images and trained machine learning algorithms to provide classification.
Awarded 'A' mark. The project was accomplished in collaboration with medical image processing company Medicsight and NHS expert colonoscopists. This thesis applied image processing, computer vision and machine learning techniques to endoscopic images, with the purpose of improving the characterisation and classification of large bowel polyps in real-time during colonoscopy. I improved the classification accuracy on two datasets by a 2 and 5% respectively (up to 86.44% and 94.64%) by adding four new features to the existing five. I also designed and implemented different approaches to narrow the margin of classification error by producing measurements of feature stability such as feature variability across the surface of polyps.
Supervisor: Dr Anil A. Bharath
- 2003 – 2009 **MEng Electrical and Electronic Engineering**, *University of Seville*, Spain, 1st class. Specialised in Computing, Signal Processing and Radiocommunications.
Dissertation: IP Multimedia Subsystem (IMS), the network convergence enabler
Awarded with 1st class with honours (10/10) in July 2009. Developed and tested IMS service deployment. Conducted a study of IMS future viability. Project accomplished whilst working within the Minerva Project. Built foundations for three scientific articles.
Supervisor: Dr Alejandro Carballar *Examiners*: Dr Antonio Estepa, Dr Rafael Bachiller
- 2009 **Mobile Communications Expert Course — Technologies and Mobile Applications: GPRS/UMTS**, University of Seville and Vodafone Spain Foundation, Researched on IP Multimedia Subsystem. Developed and simulated Java code for IMS service deployment on Ericsson SDS simulator.
- 1999 – 2003 **Secondary studies and Spanish Bachillerato**, IES Albergo College, Physics, Chemistry, Maths, Biology, plus History, Spanish, English, French and Graphic design at the A2 equivalent level. Grade: 9.2/10.

Publications

Journals and peer reviewed conference papers

- J. Rivera-Rubio**, I. Alexiou, and A. A. Bharath, "Appearance-based indoor localization: A comparison of patch descriptor performance," *Pattern Recognition Letters*, c. 10 pages, North-Holland, 2015.
- J. Rivera-Rubio**, I. Alexiou, A. Bharath, R. Secoli, L. Dickens, and E. C. Lupu, "Associating locations from wearable cameras," in *British Machine Vision Conference (BMVC)*, c. 13 pages, Nottingham (UK), 2014.
- J. Rivera-Rubio**, S. Idrees, I. Alexiou, L. Hadjilucas, and A. A. Bharath, "A dataset for hand-held object recognition," in *IEEE International Conference on Image Processing (ICIP)*, pp. 5881–5885, Paris, 2014.
- J. Rivera-Rubio**, S. Idrees, I. Alexiou, L. Hadjilucas, and A. A. Bharath, "Small Hand-held Object Recognition Test (SHORT)," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 524–531, Steamboat Springs (CO, USA), 2014.
- J. Rivera-Rubio**, S. Idrees, I. Alexiou, L. Hadjilucas, and A. Bharath, "Mobile Visual Assistive Apps: Benchmarks of Vision Algorithm Performance," in *New Trends in Image Analysis and Processing – ICIAP 2013 SE - 4*, vol. 8158 of *Lecture Notes in Computer Science*, pp. 30–40, Springer Berlin Heidelberg, 2013.

J. Rivera-Rubio, A. Madera, and A. Carballar, "Incógnitas de IMS y claves para su lanzamiento," in *Telecom I+D*, c. 4 pages, Madrid, 2009.

A. Madera, **J. Rivera-Rubio**, and A. Carballar, "DVB-H, realidad o ficción," in *Telecom I+D*, c. 4 pages, Madrid, 2009.

Other Publications

J. Rivera-Rubio (translator), *The peculiar memories of Thomas Penman*. By Bruce Robinson. Barcelona: Cabaret Voltaire, 1 ed., 2010.

J. Rivera-Rubio, A. Madera, and A. Carballar, "Hacia el Todo-IP en movilidad," *BIT*, no. 179, pp. 59–62, 2010.

Teaching Experience

Teaching assistant – study groups

2011 – 2014 **Signals and Systems**, *BE2-HSAS*.

Students learn how to approach and solve signal processing problems. I led 3 study groups of 20 students each. Prepared and presented the background and material for each problem, encouraged the students to solve them and choose the best approach. I helped maintain student interest by providing a detailed step by step explanation of the problems.

Teaching assistant – lab demonstrator

2011 – 2013 **Signals and Systems**, *BE2-HSAS*.

2013 **Image Processing**, *BE3-HIPR*.

2012 **Statistics and Data Analysis**, *BE9-MSTDA*.

2011 **Programming II**, *BE2-HPROG2*.

Other Experience

March – July **IT Coordinator (Erasmus intern)**, *Auto ID Services Ltd.*, St. Helens, UK.

2010 Internet and VoIP network and systems administration, mobile computing, IT security. Training and induction of new staff, documentation and IT strategy. Complete development of the new website.

Programming languages and other Technology skills

Advanced PYTHON, NODE.JS, MATLAB & SIMULINK, AWS, Linux (server and desktop), SQL databases, DOCKER, LXC, \LaTeX , GIT, SVN, Office Suites, Microsoft Windows, Mac OS. Computer hardware and support,

Intermediate C, C++, SCALA, JAVA, BASH scripting, MONGODB, OPENCV, HTML, Inkscape,

Basic PHP, ROS, R, Gimp, Adobe Creative Suite.

Grants and Awards

2020 **Innovate UK**, *KTP, Knowledge Transfer Partnership*, LaunchPad Data Team was successfully awarded a KTP grant to partner with University College London to make advances in our Advanced Video Interview Automated Screening (AVIAS), I was a contributor of the grant writing and project team.

2019 **Innovate UK**, *Experimental Development Open Grant Competition*, LaunchPad Recruits was awarded close to £200,000 to create a scalable platform for taking Predict to the TB scale, I was a contributor of the grant writing and project team.

- 2017 **Innovate UK**, *Industrial Research Open Grant Competition*, Award of more than £230,000 to build a Proof of Concept of Predict, Member of the project team and contributor to the grant submission.
- 2014 **Imperial College Trust**, *Conference funding*.
- 2014 **IEEE WACV conference travel grant**, *Funding for travel to Colorado, USA, to present at the 2014 IEEE WACV conference*.
- 2013 **V&L Net Pump-Priming 2013-1 Grant**, *EPSRC Network on Vision and Language (V&L Net)*, Riccardo Secoli and **Jose Rivera-Rubio**.
Funding for the BELVIS project to construct and release a database of “visual paths” containing images and videos of different journeys together with a voice record of the position to establish the ground truth. The release of this dataset motivates research in the areas of mobile and visual localisation, voice-enabled ground truth and natural language processing of navigational information
- 2013 **Highly Commended Prize in the Poster Competition**, *Imperial College Graduate School Summer Research Symposium*, Awards given to the best 10 posters among more than 100.
- 2011 – 2014 **EPSRC PhD studentship**, *Imperial College London Department of Bioengineering*, Awarded one of three studentships among more than 120 candidates.
- 2011 **2011 International Grants**, *‘La Caixa’ Foundation: Welfare projects*, Selected for an interview on the 25th May 2011 to obtain a studentship for a PhD in Bioengineering Research at Imperial College. Shortlisted within the best 70 candidates among 19,000+ applicants to compete for 25 grants. Could not attend the interview due to Imperial College exam clash.
- 2010 **Erasmus Placement grant**, *European Union*, This studentship funded my placement at a UK company.
- 2007 – 2008 **SICUE/SENECA Scholarship**, *Spanish Education Ministry*, Scholarship and bursary to spend my MEng 5th and final year in Carlos III University of Madrid, Spanish top university in Computing.
- 2003 **Best performance award**, *Andalusian Regional Government*, Awarded with a best performance prize at college level and granted funding of first year tuition fees at university.

Communication Skills

- 2015 **Oral presentation at The National Archives**, London, UK, Title of the talk: ‘Perceptually Similar Search’.
- 2014 **Oral presentation at the Winter Conference on Applications of Computer Vision**.
- 2013 **Oral presentation at the International Workshop on Assistive Computer Vision and Robotics**, *ICIAP 2013*, Naples, Italy, Title of the talk: ‘Mobile Visual Assistive Apps: Benchmarks of Vision Algorithm Performance’.

Languages

Spanish	Mother tongue	
English	Full professional proficiency	<i>TOEFL iBT 114/120</i>
French	Intermediate proficiency	<i>Completed 4 years at the Official School of Languages</i>
Chinese (Mandarin)	Elementary proficiency	<i>Completed 2 years of the language course at University (9/10)</i>
German	Elementary proficiency	<i>Completed 1 year of the language course at University (7/10)</i>

Interests

I have a passion for literature and enjoy every bit of it: from reading books to writing and translating. I have played the piano for 15 years and I have also used technology to teach myself to play the guitar, which I now consider my first instrument. I feel that music and especially playing these instruments allows me to both relax and disconnect from my daily routine. I also love doing sports, like cycling and football, going to the cinema and watching a good TV series. One of my passions is travelling, which I would like to experience more and more with my family, friends, and to meet new people in every place.

References

References are available upon request.