

Antonio de la Vega de León

antonio@delavega.ai | <https://www.linkedin.com/in/adlvdl/> | 41010, Sevilla, Spain

SENIOR PRINCIPAL SCIENTIST

AI / Machine Learning (ML) expert specializing in the development of innovative machine learning strategies to support novel drug discovery. Experienced in the application of ML to predict biological endpoints of small molecules, with prior experience on prioritizing molecular glues for phenotypic screening. Extensive experience using cheminformatics and AI/ML coding libraries. Strong publication record, extensive teaching background in data science and cheminformatics, and actively engaged in science outreach, with experience organizing public science events. Highly organized, adaptable communicator with proven leadership in managing multidisciplinary teams. Currently seeking new opportunities in the drug discovery or biomedical research field, looking to develop novel strategies that use ML to speed up the discovery of novel clinical candidates.

AI/ML Expert | Drug Discovery | Python | R | Data Science | Cheminformatics | Teaching | Leadership |

EXPERIENCE

PRINCIPAL SCIENTIST AND SENIOR PRINCIPAL SCIENTIST

01/2022 – 03/2025

Bristol Myers Squibb, Seville, Spain

Led a team of three scientist, a mixture of computational chemistry and ML engineers.

- Increased the efficiency by 20% and the hit identification output by 300% of the proteomics screening campaign through thoughtful application of novel ML models (LLM, contrastive learning).
- Designed 2 progressible molecular glue series using generative AI models.
- Implemented uncertainty estimation methods to guide our generative AI models.
- Developed and coordinated a mentoring scheme for the Seville research site, where 2 of my 3 mentees were promoted during the mentoring scheme.
- Designed and implemented machine learning models for predicting biological endpoints of small molecules, achieving a 15% improvement in prediction accuracy compared to previous methodologies.

RESEARCH SCIENTIST

03/2021 – 12/2021

Bristol Myers Squibb, Seville, Spain

- Developed methods to analyze the structural diversity of the molecular glue library, enabling better tracking of assay results and enhancing chemical synthesis decisions.
- Benchmarked molecular representations for drug-target interaction models, leading to two conference outputs (presentation at ACS Fall 2022 and poster at Sheffield Conference on Cheminformatics 2023)

LECTURER IN CHEMINFORMATICS

10/2018 – 02/2021

University of Sheffield, Sheffield, United Kingdom

- Led the Sheffield Cheminformatics Research Group, directly supervising 3 PhD students and supporting 3 additional PhD students, and 1 postdoc.
- Supervised projects on imputation of toxicity data, biomimetic search of conformational search and generative models for polypharmacology in Alzheimer's Disease.
- Developed one module on data visualization and taught on 3 additional modules on the MSc Data Science.

MARIE CURIE POSTDOC

11/2016 – 09/2018

University of Sheffield, Sheffield, United Kingdom

- Worked as a researcher in the European project D3i4AD (FP7-PEOPLE-2013-IAPP) to develop ML methods to support phenotypic screening in the Alzheimer's research area.

EDUCATION

PhD Computational Life Sciences, University of Bonn 2012 – 2016

M. Sc. Life Science Informatics, University of Bonn 2010 – 2012

Bachelor's degree in Biology, Universidad Complutense de Madrid 2004 – 2010

TECHNOLOGY

Machine learning: Random Forest, Support Vector Machine, Deep Neural Networks, Feature Nets, and Matrix Factorization for single-task, multi-task, and imputation-based prediction models. Variational Autoencoders and reinforcement learning for de novo drug design. Conformal prediction and evidential layers to assess model uncertainty.

Cheminformatics: Visualization tools to analyze SAR, such as molecular similarity networks or projections of descriptor space, docking and analysis of matched molecular pairs.

Programming: Python (NumPy, Matplotlib, TensorFlow, scikit-learn, RDKit), R (Tidyverse, ggplot2, networkx, Plotly), and Java (OpenEye, JUNG).

Software: MOE for chemical modeling, Autodock for docking, KNIME for data mining.

LANGUAGES

Spanish: Native | English: Fluent (C2) | German: Advanced (B2)

SELECTED PUBLICATIONS

- Torrisi M, Asadollahi S, de la Vega de León A, Wang K, Copeland W. Do chemical language models provide a better compound representation? bioRxiv 2023.11.07.566025.
- Walter M, Allen LN, de la Vega de León A, Webb SJ & Gillet VJ. Analysis of the benefits of imputation models over traditional QSAR models for toxicity prediction. J Cheminform 14, 32, 2022.
- Torrisi M, de la Vega de León A, Climent G, Loos R & Panjkovich A. Improving the Assessment of Deep Learning Models in the Context of Drug-Target Interaction Prediction. bioRxiv 2022.04.20.488898.
- de la Vega de León A, Chen B & Gillet VJ. Effect of missing data on multitask prediction methods. J Cheminf, 10:26, 2018.
- de la Vega de León A & Bajorath J. Prediction of compound potency changes in matched molecular pairs using support vector regression. J Chem Inf Model 54, 2654-2663, 2014.
- de la Vega de León A & Bajorath J. Compound optimization through data set-dependent chemical transformations. J Chem Inf Model 53, 1263-1271, 2013.