Diego Garcia-Olano

4115 Avenue H, Austin, TX 78751 | 830-421-1805 | diegoolano@gmail.com
www.diegoolano.com | github.com/diegoolano | linkedin.com/in/diego-garcia-olano/ | twitter.com/dgolano

EDUCATION

The University of Texas at Austin, Austin, TX, USA. 9/2017-7/2022
Ph.D - Electrical and Computer Engineering - Decision, Information & Comms Engineering (DICE) Track
Advisor: Dr. Joydeep Ghosh. Intelligent Data Exploration & Analysis Lab. ideal.ece.utexas.edu
Dissertation: "In-process Diagnostic methods for Entity Representation Learning on Sequential data at Scale"
Committee: Alex Dimakis, Harris Vikalo, Atlas Wang & Byron Wallace (document | slides)

Universitat Politècnica de Catalunya - Barcelona, Spain. Graduated: 7/2015
Masters in Computer Science - Data Science
Advisors: Marta Arias, Josep Lluis Larriba Pey - Lab for Relational Algorithmics, Complexity & Learning, DAMA
Thesis: "Automated Construction & Analysis of Political Networks". github.com/diegoolano/whoyouelect

The University of Texas at Austin Austin, TX, USA. Graduated: 5/2004
Computer Science (B.A., 05/2004) (3.86 GPA of 4) Graduated with Honors
Government (B.A., 05/2005) (4.0 GPA) with minor in: Business
Hispanic Studies (B.A., 05/2004) Graduated with Honors

Universidad de Sevilla, College of Arts & Letters (FilologÌa) Sevilla, Spain; Attended: Fall 2002
Institut Catholique de Toulouse, Institut de Langue, Toulouse, France; Attended: 03/2012 - 06/2013

RESEARCH INTERESTS

Explainable multimodal ML focusing on evaluation and mitigation of Safety Alignment (PII/IP/memorization) for LLMs

PUBLICATIONS

- Hallucination reduction with CASAL: Contrastive Activation Steering For Amortized Learning. NeurIPS 2025 Mechanistic Interpretability Workshop. Wannan Yang, Xinchi Qiu, Lei Yu, Yuchen Zhang, Oliver Aobo Yang, Narine Kokhlikyan, Nicola Cancedda, **Diego Garcia-Olano** arxiv.org/pdf/2510.02324
- Measuring AI "Slop" in Text. Chantal Shaib, Tuhin Chakrabarty, **Diego Garcia-Olano**, Byron C. Wallace (under submission 2025) arxiv.org/pdf/2509.19163
- Improving LLM First-Token Predictions in Multiple-Choice Question Answering via Prefilling Attack. Silvia Cappelletti, Tobia Poppi, Samuele Poppi, Zheng-Xin Yong, **Diego Garcia-Olano**, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara. (under submission 2025) arxiv.org/pdf/2505.15323
- Genμ: The Generative Machine Unlearning Challenge (ICCV 2025 - U&ME workshop) Kartik Thakral, Shreyansh Pathak, Tamar Glaser, Tal Hassner, **Diego Garcia-Olano,** Iacopo Masi, Richa Singh, Mayank Vatsa
- Beyond Explainable AI (XAI): Overdue Paradigm Shift and Post-XAI Research Paths (in progress, 2025) Saleh Afroogh, **Diego Garcia-Olano**, 55+ researchers from Google Deepmind , IBM, Amazon, Stanford, etc.
- The Llama 3 Herd of Models (2024) Core Contributor - Memorization Safety Evals. arxiv.org/abs/2407.21783
- Vivek Miglani, Aobo Yang, Aram H. Markosyan, **Diego Garcia-Olano**, Narine Kokhlikyan, Using Captum to Explain Generative Language Model. EMNLP 2023 - NLP OSS workshop
- Fulton Wang, Julius Adebayo, Sarah Tan, **Diego Garcia-Olano**, Narine Kokhlikyan. "Error Discovery by Clustering Influence Embeddings". NeurIPS 2023 Main Track / ICLR 2023. Oral - Trustworthy ML workshop,
- **Diego Garcia-Olano,** Yasumasa Onoe, Joydeep Ghosh, Byron C. Wallace "Intermediate Entity-based Sparse Interpretable Representation Learning". EMNLP 2022. Blackbox NLP workshop
- **Diego Garcia-Olano**, Yasumasa Onoe and Joydeep Ghosh. "Improving and Diagnosing Knowledge-Based Visual Question Answering via Entity Enhanced Knowledge Injection". ACM WWW 2022. MUWs Workshop
- Detección de sesgos en razón del género en decisiones judiciales utilizando PLN (Jornadas Argentinas de Informática 2022) Mariana Brocca, Antonela Tommasel, **Diego Garcia-Olano,** Andrés Pace, Ezequiel Valicenti
- **Diego Garcia-Olano**, Yasumasa Onoe, Ioana Baldini, Joydeep Ghosh, Byron Wallace and Kush Varzney. "Biomedical Interpretable Entity Representations". ACL-IJCNLP 2021.

- Daniel Gillick, Sayali Kulkarni, Alessandro Presta, Jason Baldridge, Eugene Le and **Diego Garcia-Olano.** "Learning Dense Representations for Entity Retrieval". CoNLL 2019
- **Diego Garcia-Olano**, Alan Gee, Joydeep Ghosh and David Paydarfar. "Deep Classification of Time-Series Data with Learned Prototype Explanations". ICML 2019 Time Series Workshop (4 pages) bit.ly/2IJIuEK & IJCAI 2019 Workshop on Knowledge Discovery in Healthcare Data (8 page) arxiv.org/abs/1904.08935
- Kris Sankaran, **Diego Garcia-Olano**, et.al, "Applying Machine Learning methods to Enhance the Distribution of Social Services in Mexico", Data Science For Social Good Datafest, 08/2016 and arxiv update 2017, Chicago. Paper: arxiv.org/abs/1709.05551 , talk at UChicago DataFest: goo.gl/AG66bg , poster: goo.gl/Bc4HAf
- **Diego Garcia-Olano**, Marta Arias and Josep L Larriba-Pey: "Automated Construction and Analysis of Political Networks via open government and media sources", ECML SoGood workshop 2016, Italy. paper: goo.gl/xgCU81
- Ivan Paz-Ortiz, **Diego Garcia-Olano**, Carlos Gay-García. "TF-IDF Assessment of Similarity in Climate Change Programs in Mexico", Simultech 2015 France. diegoolano.com/ivanpaz

COMMUNITY INVOLVEMENT
- Primary Co-organizer of Unlearning and Model Editing workshop (ECCV 2024, ICCV 2025) - workshop link
- Committee Member for PhD 2025 - Samuelle Poppi (Università di Pisa) - Responsible AI in Vision & Language
- NeurIPS 2025 (Main/Mech Intrep workshop), 2024, 2023 (Main/Regulatable ML workshop), 2022, 2021 reviewer
- ACL 2023 reviewer, EMNLP 2023, ACL-IJCNLP 2021 reviewer, ACL Rolling Reviewer
- ICLR 2025, 2024 (Data-centric Machine Learning Research workshop), 2023, 2022, 2021, 2020 reviewer (Top Reviewer Award), ICLR AI for Social Good 2019 reviewer
- ICML 2025, 2021, 2020 reviewer, Texas Linguistics Summit Conference 2019 reviewer
- Data Science For Social Good - Project scoping and Candidate interviewer: 2017, 2018

SELECTED PROJECTS & PRESENTATIONS
- Leveraging State-of-the-Art Unsupervised Style Transfer Models for Counterfactual Text Generation. 5/2020. RL, Retrieval & Distengagled learning methods, Computational Linguistics Seminar project. bit.ly/3DLZFZx
- Explainability methods for NLP talk at Cognitive Scale 2/2020. bit.ly/3FYbeh5 , bit.ly/3l31Al6
- Link Detection in Political Networks using GCNs, 12/2018. NLP Final Project bit.ly/3CQHvnX
- Lecture on Modern Visualization, UT Austin, 11/2020. slides: goo.gl/R6PVzF
- Network Science Institute at Northeastern University, 10/2016: "Whoyouelect.com" tech talk. goo.gl/F1dNcv
- Alicante, Spain. 11/2014. "WP5: Preliminary analysis for Microbial Source Tracking". http://bit.ly/2eDp9ga
- "Understanding US counties: Clustering and Classification over voting, socio-economic and public health"
- 4/2014 Predicting Blue Islands in the US. Paper/Code/Visualization: diegoolano.com/electionmap

PROGRAMMING SKILLS
Highly skilled with Python, PyTorch, Tensorflow, Huggingface, unix, mysql/presto/postgres, git, js, d3, R,
Experience with: keras, postgis/qgis, spark, java, c++, nodejs, matlab, apache/nginx www.github.com/diegoolano

EMPLOYMENT EXPERIENCE

Meta - Gen AI /Superintelligence Lab Senior Research Scientist **07/2024 - present**
- Research on text memorization safety evaluation efforts (general, copyright, PII, code) for Llama 3.
- Sole IC lead on multi-modal Biometric ID and PII/IP/RAG pre/posttraining risk evaluation and mitigation for LLama 3 & 4 OSS models.
- Developed research plan and hosted PhD intern on safety steering to reduce hallucinations of post trained models efficiently (accepted at Neurips 2025 mechanistic interpretability workshop. undersubmission at ICLR)
- Co-hosted visiting PhD Researcher studying memorization of distillation around multi-modal models (ongoing)

Meta - Responsible AI / GenAI Research Scientist **9/2022 - 07/2024**
Research on fair, transparent, guidable AI systems for social impact. Label noise detection, understanding and mitigation using influence functions and model training dynamics for classifier and LLM improvements.

Fulbright - Azul, Argentina Fulbright Specialist **3/2022 - 4/2022**
Lead development and research on a tool for predicting gender bias in judicial proceedings using NLP.
Taught a workshop (in spanish) to faculty/researchers in the Law & Computer Science departments of UNICEN.

IBM Research- NY IBM Science for Social Good PhD Fellow **5/2020 - 8/2020**
Applied research on Biomedical Interpretable Entity Representations For Drug Repurposing using PubMed articles and links between UMLS and Wiki for learning word embeddings for tasks. Host: Kush Varshney, Co-host: Ioana Baldini

Google AI - Seattle Software Engineering Research Intern **5/2019 - 8/2019**
Conducted applied research on explaining seq2seq models using feature attribution methods on Transformer and LSTM based architectures with attention for machine translation. Host: Besim Avci, Co-host: Frederick Liu

Google AI - Mountain View Software Engineering Research Intern **5/2018 - 8/2018**
Conducted applied research on Entity Linking task, built upon internal framework and worked on reproducing findings for OpenAI Deep Type paper. Work from project published in CoNLL 19. Host: Jason Baldridge, Co-host: Daniel Gillick

University of Texas at Austin - Intelligent Data Exploration & Analysis Lab Teaching Assistant **9/2017 -7/2022**
TA for Fall 17 Advanced Predictive Modeling course on ML with Python (Business Masters of Analytics course)
TA for Spring 19 Responsible AI Graduate Engineering seminar course. Syllabus: www.github.com/ideal-ut/RAI-course/
TA for Spring 19 Manoj Saxena's Ethics in AI Design short course

Eric & Wendy Schmidt Data Science For Social Good at the University of Chicago DSSG Fellow **5/2016 - 9/2016**
Worked with SEDESOL of Mexico to improve its allocation of social service benefits for 80 million people. Developed predictive models to estimate incomes, needs, and likelihood of eligibility during recertification periods. Received training in machine learning for social projects by the Chief Scientist of Barack Obama's 2012 election campaign.

Diego Garcia-Olano, **www.diegoolano.com** Freelance Developer, Data Scientist and Data Visualizer **9/2011 - 9/2017**
Principal architect/developer for clients including Glasstire, Spotify, Pitchfork, Boxee TV, Switch Energy Project, etc.
Frontend/backend/analytics (mostly python/php, d3, javascript, unix, apache/nginx, mysql/postgres/mongodb)

Freelancers Union - Brooklyn, New York, Software Developer **4/2009 - 12/2009, 5/2010 - 9/2011**
Principal developer of Political Action Committee Fundraising for State and Federal PAC's, Client Scorecard, and The Freelance Life applications used by over 120,000 members (django,jquery,postgres).

Sapling Systems - Science Technologies; Austin, Texas, Developer & Adaptive Learning Researcher **2/2005 – 2/2009**
Principal developer of the IBIS educational platform used by 35,000 students, saplinglearning.com, Provided research and working prototypes on Item Response Theory models for feedback functionality.

Growth Accelerated Partners; Buenos Aires, Argentina; Lead Software Developer **11/2007 - 11/2008**
Principal developer of Label Designer for personalwine.com. Oversaw team of developers in Costa Rica

Caritas of Austin; Austin, Texas; Lead Developer **3/2006 - 3/2008**
Developed app for a refugee employment program to digitize employment search data and facilitation of job opportunities with potential employers. Developed time-tracking system for case workers for donors transparency

Department of State, US Embassy, Caracas, Venezuela; Political Affairs Intern **9/2004 - 1/2005**
Maintained the Embassy's system for transferring cables and observed regional elections around greater Caracas area

Ford Motor Company Global Manufacturing Supply Chain; Dearborn, Michigan; Developer **5/03 - 8/03, 5/02-8/02**
Developed project to simulate daily production levels for Ford's manufacturing plant in Belgium (VB, Java)
Lead various benchmark studies comparing Ford IT capabilities with OEM's.

UT Learning Center; Austin, Texas; Tutor, **1/2002 - 7/2004**
Tutored students on a weekly basis in computer science, logic, math, spanish, accounting, and chemistry

VOLUNTEERING AND SOCIAL PROJECTS

- **Brandon Dudley campaign for Harris County Tax Assessor/Voter Registrar** **1/2016 - 4/2016**

Created digital tools to assist the campaign's volunteers to get out the vote.

- **ACLU Immigration Policy Center** Criminal Alien Program Database and Data Visualization; **1/2008 - 1/2009**
  Developed system to digitize & visualize prisoner data in Texas (2000-2008) http://goo.gl/QItf8
- **Relief Oversight** - Helped gather a team and develop a publicly generated, app to monitor the activities and effectiveness of organizations soliciting donations for relief after an earthquake in Haiti. **12/2009 - 2/2010**
- **Volunteer Healthcare Clinic**; Austin, Texas, Spanish Interpreter & Aide **5/2008 - 8/2008**
- **El Buen Samaritano**; Austin, Texas, ESL Teacher. **9/2003 - 5/2005**
  Taught English night courses for a non-profit provider of working-poor, Spanish-speaking families.

LANGUAGE PROFICIENCIES

Fluent Spanish (written & spoken)
French (Proficient - C1 level as defined by CEFR)

AWARDS & AFFILIATIONS

The Charles W. and Margaret A. Tolbert Endowed Scholarship in Electrical and Computer Engineering (2019 and 2021)
Eric & Wendy Schmidt Data Science For Social Good Fellow
HSF / Ford Corporate Scholarship Recipient & College of Natural Sciences College Scholar