William Berrios

wberriosr@uni.pe, wiliam@contextual.ai
Github Google Scholar

EDUCATION

National University of Engineering, Peru — Jan 14 - Aug 19
B.S. Mechatronics Engineering (Robotics)
Summa Cum Laude. Rank 1/46

Hochschule Furtwangen University, Germany. — Sep 18 - Feb 19
International exchange program.
Courses: Deep Learning, Robotics, IoT, Control Systems

PROFESSIONAL EXPERIENCE

Founding Research Engineer — Apr 23 - Oct 25
Contextual AI,USA — Supervisor: Douwe Kiela & Amanpreet Singh

- Built an automated stress-testing framework for production RAG 2.0 (retrieval-augmented generation) models; validated that Contextual Language Models (CLMs) outperformed GPT-4-based and Mixtral baselines across domains and reliability axes (grounding, retrieval, reasoning, citation). .
- Technical lead for GLM: designed evaluation (automated + human), built the training stack (instruction-tuning → offline RL → online RL), and led red-teaming; achieved SOTA factuality on FACTS (Google Research/DeepMind) and deployed to production enterprise APIs.
- Authored LENS ("Language Models Enhanced to See") and shipped a document parser that infers document hierarchy (tables/figures/sections) for retrieval and reasoning; delivered a +14.8 percentage-point end-to-end accuracy lift in the RAG 2.0 pipeline and powers document ingestion at scale.
- Authored LMUnit ("Fine-grained Evaluation with Natural-Language Unit Tests"): created datasets, trained underlying models, and integrated claim-/task-level scoring into internal automatic evaluation across customer workloads.
- Built a multi-agent orchestration framework for telecom log analysis: coordinated specialized agents with separate context windows to process massive logs beyond a single model's limits, enabling automated root-cause diagnosis and rigorous evaluation.

Junior Research Scientist — Jan 23 - Mar 23
Artificio, USA - Remote — Supervisor: Arturo Deza

- Developed a brain-aligned transformer model, serving as the core search engine for identifying similar artworks
- Contributed to the Machine Learning pipeline as a key team member in the development of the Artwork Search Engine product, ensuring seamless data processing and model integration.

Data Scientist — May 22 - Dec 22
Rimac Seguros, Peru

- Designed and implemented a recommendation system model to enhance cross-selling of insurance products by suggesting the most suitable options to customers.
- Designed and deployed a machine-learning solution to accurately estimate insurance product prices, optimizing pricing strategies for enhanced competitiveness and profitability.

Neuro AI Researcher — Jan 22 - Aug 22
Center for Brains, Minds and Machines, MIT — Supervisor: Arturo Deza

- Designed and trained a dual-stream Vision Transformer that jointly enforces rotational invariance and adversarial robustness; improved out-of-distribution reliability and increased alignment to primate visual areas (V1/V4/IT).
- Achieved state-of-the-art on Brain-Score's V4 and ranked #2 overall on the Brain-Score competition at submission (March 2022).
- Authored "Joint Rotational Invariance and Adversarial Training of a Dual-Stream Transformer Yields State-of-the-Art Brain-Score for Area V4"; presented at NeurIPS 2022 workshops (LXAI oral; SVHRM poster).

Computer Vision Researcher — Jan 21 - Apr 21
Electronic Visualization Laboratory, Univesity of Illinois at Chicago — Supervisor: Elisabeta Marai & Juan Trelles

- Implemented a Cost-Effective Active Learning (CEAL) pipeline for biocuration: triaged predictions by confidence, auto-pseudo-labeled high-confidence cases, and routed low-confidence cases to biocurators using entropy/margin uncertainty; reduced manual labeling and improved label quality.

- Developed hierarchical CNN classifiers aligned to biomedical ontologies/taxonomies (e.g., category → subcategory), enabling learning from partially labeled records and improving data efficiency and probabilistic calibration under class imbalance and label noise.
- Integrated ontology-aware training with a human-in-the-loop review workflow so biocurators could target super- and sub-classes despite incomplete ground truth, enabling scale across large unlabeled biomedical datasets

Data Scientist Jan 20 - Dec 20
Banco Pichincha, Peru

- Increased productivity of business areas by implementing machine learning models for loan default prediction, credit card customer behavior, and debt collection management.
- Trained 5 co-workers from the Advanced Analytics team in Python, Machine Learning, and MLops.

Undergraduate Researcher in Robotics & AI Jan 18 - Aug 18
Intelligent Systems Lab, Lima, Peru Supervisor: Alberto Coronado

- Developed a comparison of traditional and machine learning methods for evaluating the health condition of bearings presented in mechanical systems.
- Implemented a prototype of an autonomous mobile robot for parking surveillance using path planning and an object detection algorithm for recognizing license plates.

PUBLICATIONS

LMUnit: Fine-grained Evaluation with Natural Language Unit Tests
Jon Saad-Falcon*, Rajan Vivek*, **William Berrios***, Matija Franklin, Bertie Vidgen, Amanpreet Singh, Douwe Kiela
Accepted @ EMNLP - 2025
[Paper] [Code] [Blog] [Tweet]

BI-LAVA: Biocuration Image Labeling through Active Learning and Visual Analysis
Juan Trelles,Andrew Wentzel, **William Berrios**, Hagit Shatkay and G.Elisabeta Marai
Accepted @ Computer Graphics Forum Journal - 2025
[Paper]

Leveraging Diffusion Perturbations for Measuring Fairness in Computer Vision
Nicholas Lui, Bryan Chia, **William Berrios**, Candace Ross, Douwe Kiela
Accepted @ AAAI - 2024
[Paper][Dataset][Tweet]

Towards Language Models That Can See: Computer Vision Through the LENS of Natural Language
William Berrios, Gautam Mittal, Tristan Thrush, Douwe Kiela, Amanpreet Singh
ArXiv 2023
[Paper][Code][Blog][Tweet]

Joint rotational invariance and adversarial training of a dual-stream Transformer yields state of the art Brain-Score for Area V4
William Berrios, Arturo Deza
Accepted @ LXAI **(oral presentation)** - Neurips 2022
[Paper][Code][Tweet]

AWARDS

1st Place Facts Grounding Leaderboard 2025
- Rank 1/34 among submissions from Big AI labs such as OpenAI, Google, Anthropic, Mistral. Organized by Google-DeepMind & Kaggle. [Tweet]

2nd Place RewardBench 2025
- Rank 2/70 among submissions from Big AI labs such as OpenAI, Google, Anthropic. Organized by AllenAI Institute. [Tweet]

1st Place BBVA Data Challenge 2023
- Rank 1/500. Organized by BBVA. Awarded $3000.

2nd Place at Brain-Score Competition 2022
- Rank 2/21. Organized by MIT Intelligence Quest and MIT IBM Watson AI Lab. Awarded $1250.

2nd Place at International Interbank Datathon 2021

· Rank 2/232. Organized by Interbank - Peru. Awarded by $6000 dollars.

1st Place at BNP Machine Learning Competition 2021
· Rank 1/100. Organized by BNP Paribas Cardif and Domino DataLab.

1st Place at International BCI Machine Learning Competition 2021
· Rank 1/400 (15+ countries). Organized by BCI Bank - Chile. Awarded by $3600.

1st Place at International BanColombia Datathon 2020
· Rank 1/80. Organized by Group BanColombia - Colombia. Awarded $2000.

Silver Medal at Ventilator Pressure Prediction 2021
· Rank 122/2605 (Top 5 %). Organized by Google Brain & Kaggle.

1st Place at WIDS Bay Area Datathon 2021
· Organized by WIDS - Stanford and Google Cloud team.

2nd Place at Brewing Data Cup 2020
· Rank 2/40. Organized by AB-InBev corporation.

Dean's List Scholar, National University of Engineering 2016-2021
· Awarded on the basis of grade point average (GPA).

MEDIA COVERAGE

Peruvian engineer from UNI revolutionizes Silicon Valley with artificial intelligence models in the United States 2025
LaRepublica Newspaper - Peru

Contextual AI's new AI model crushes GPT-4o in accuracy — here's why it matters 2025
VentureBeat

Making LLMs Multi-Modal without Fine-Tuning 2023
Chai Time Data Science YouTube channel - 15K followers & 2k views

Promoting national talent: The case of the students from the National University of Engineering, winners of the Data Science Challenge 2021
Ministry of Education of Peru

INVITED ACADEMIC REVIEWING

Neural Information Processing Systems (NeurIPS) 2022, 2023, 2025

Empirical Methods in Natural Language Processing (EMNLP) 2025

International Conference on Learning Representations (ICLR) 2024

Conference on Language Models (COLM) 2024

TALKS

Beyond text: Latin Innovations in LLMs, Vision, and Multimodal AI, Techsuyo Conference - USA 2025

Retrieval Augmented Systems in 2025. Maristas College - USA 2025

Language Models Enhanced to See. ML Pinterest Lab - USA 2023

Towards Language Models That Can See. Deza Lab @ UTEC - Peru 2023

State of the art Brain-Score for Area V4. LXAI, NeurIPS - USA 2022

Brain-Aligned Vision Transformers. Poggio Lab @ MIT - USA 2022