

JUAN PABLO ZAPATA CORDOBA

SENIOR AI TRAINER & LLM EVALUATOR | RLHF, SFT & AI CODING | PEER REVIEWER

Soledad, Atlantico, Colombia | +57 300 3283386 | juanpazapata2013@gmail.com | jpzcsolutions.com

PROFESSIONAL SUMMARY

Senior AI Trainer and LLM Evaluator with 1.5+ years of hands-on AI post-training experience across RLHF, supervised fine-tuning (SFT), preference ranking, pairwise comparison, rubric-based evaluation, reference-response authoring, red teaming, and AI code evaluation, backed by 5+ years of professional experience across AI and software engineering. Specialized in instruction-following and coding evaluations, producing high-quality human feedback, identifying model failure modes, and maintaining training-data quality through peer review, adjudication, and reviewer calibration. Strong technical foundation in JavaScript, TypeScript, Python, debugging, testing, and production software systems.

CORE AI TRAINING SKILLS

Post-Training & Human Feedback: RLHF | Supervised Fine-Tuning (SFT) | Preference Ranking | Pairwise Comparison | Multi-Criteria Scoring | Response Rewriting | Golden / Reference Responses | Preference Rationales

LLM Evaluation: Rubric-Based Evaluation | Instruction Following (IF) | Correctness | Relevance | Factuality | Completeness | Style | Safety | Response Verification | Failure-Mode Analysis | Edge-Case Analysis

AI Coding Evaluation: Code Execution | Debugging | Test Authoring & Execution | Bug Reproduction | Refactoring | Code Generation | Code Review | Security Review | Prompt-to-Code Verification | Repository / Terminal Workflows

Evaluation Design & QA: Rubric Authoring | Evaluation Criteria Design | Peer Review | Reviewer Calibration | Adjudication | Quality Assurance | Guideline Development | Discrepancy Resolution | Corrective Feedback

Technical: JavaScript | TypeScript | Python | React.js | Next.js | Node.js | Express.js | REST APIs | MySQL | PHP | Git | Docker

AI TRAINING & EVALUATION EXPERIENCE

INVISIBLE TECHNOLOGIES | Freelance AI Trainer & Peer Reviewer - RLHF / LLM Evaluation | 03/2026 - Present

Remote | Primarily English-language AI training environment

- Perform RLHF-oriented model evaluation through pairwise comparison, preference ranking, multi-dimensional scoring, and detailed written rationales against evolving project guidelines.
- Advanced to Peer Reviewer and audited 325+ production AI training tasks for evaluation accuracy, instruction-following assessment, label quality, rubric compliance, and contributor consistency.
- Author and review golden/reference responses, assessing model behavior across instruction following, correctness, relevance, completeness, style, safety, and other task-specific dimensions.
- Approve, reject, correct, and return contributor submissions while providing structured feedback and resolving scoring, labeling, and guideline-interpretation errors.
- Perform reviewer calibration and adjudication on ambiguous cases, aligning evaluator judgment with project standards and maintaining consistent rubric application.
- Identify recurring discrepancies and guideline gaps, escalate edge cases, and contribute clarifications that improve evaluation consistency and downstream training-data quality.
- Conduct adversarial evaluation and red teaming on selected tasks to surface subtle model weaknesses, unexpected behaviors, and instruction-following failure modes.

SCALE AI / OUTLIER | Technical Squad Lead & Senior Generative AI Engineer - AI Training / Coding Evaluation | 02/2025 - 01/2026

Remote | Coding, generalist, multimodal, and quality-review workflows

- Produced RLHF and SFT training data through preference ranking, pairwise comparisons, response scoring, response rewriting, and creation of ideal model responses across coding and general AI tasks.
- Authored and reviewed complex evaluation rubrics with independent, verifiable criteria for instruction following, correctness, functional behavior, code quality, edge-case handling, and task-specific requirements.
- Created and reviewed golden/reference responses from scratch and improved existing responses to provide high-quality supervised demonstrations and evaluation targets.
- Evaluated AI-generated code across JavaScript, TypeScript, Python, frontend, backend, APIs, refactoring, debugging, and code-generation tasks, with strongest depth in frontend and refactoring workflows.
- Executed generated code, reproduced bugs, wrote and ran tests, validated outputs, inspected repositories, and used IDE/terminal workflows to determine functional correctness beyond surface-level code review.
- Compared competing model solutions and wrote evidence-based preference rationales covering correctness, instruction following, implementation quality, security, efficiency, and maintainability.
- Designed prompts and adversarial tasks to expose reasoning errors, implementation failures, edge cases, and instruction-following weaknesses; also performed external research and factual verification for generalist evaluations.
- Led a technical squad of 5+ engineers, oversaw 120+ frontend AI training projects, and contributed to 500+ AI training/evaluation tasks while standardizing review practices and improving team efficiency by approximately 30%.
- Developed 60+ functional frontend applications in JavaScript and TypeScript as high-quality training/evaluation artifacts and evaluated multimodal model behavior across vision and audio tasks.
- Actively improved guidelines and rubrics by identifying recurring ambiguities, resolving evaluator discrepancies, and proposing clarifications based on production edge cases.

ADDITIONAL SOFTWARE ENGINEERING EXPERIENCE

Production engineering background used as domain expertise for AI code evaluation, debugging, testing, architecture review, and technical reasoning.

MILLENNIUM CENTER | Full-Stack Developer | 04/2024 - 02/2025

Barranquilla, Colombia

- Architected a scalable enterprise management system using Node.js, React/Next.js, TypeScript, Docker, and MySQL; built REST APIs and microservices for shift management, payroll, employee workflows, and automation.
- Reduced operational costs by 35% and improved response times by 42% through production process automation and system redesign.

ENERGIA SOLAR HELIOS | Full-Stack Developer & Data Analyst | 02/2022 - 03/2024

Barranquilla, Colombia

- Built production platforms, APIs, database integrations, and automation systems using TypeScript, JavaScript, React/Next.js, Node.js/Express, PHP, MySQL, and AWS S3.
- Improved application-processing efficiency by 45% and reduced operational errors by 60% through workflow automation and systems integration.

TJDATA GROUP SAS | Full-Stack Developer & IT Support | 07/2021 - 01/2022

Soledad, Colombia

- Developed an e-commerce platform using React/Next.js and PHP, configured CRM systems, and improved online sales conversion by 18%.

SELECTED AI EVALUATION STRENGTHS

Coding Domains: Frontend Development | Backend & APIs | Refactoring | Code Generation | Debugging | Testing | Repository-Level Review | Security & Quality Analysis
Evaluation Methods: Human Preference Data | Pairwise Ranking | Multi-Dimensional Scoring | SFT Demonstrations | Golden Responses | Rubric Design | Adjudication | Red Teaming
Collaboration: Peer Review | Reviewer Calibration | Guideline Clarification | Technical Leadership | Asynchronous Remote Work | English-Language Production

EDUCATION & TRAINING

Associate Degree in Information Systems Analysis and Development | SENA - National Learning Service | 2022 - 2024
Oracle for Java Course | MinTIC Colombia | 2022
Mobile Device Programming | SENA | 2021

LANGUAGES

English: Professional working proficiency | **Spanish:** Native