

MIGUEL COLLI

Bilingual AI Data Specialist (Spanish MX / English) | Junior Full-Stack Developer
Mérida, Mexico · Ph: +52 55 5967 0243 · Email: coqm86@gmail.com · [LinkedIn](#)
Portfolio: <https://mecqportfolio.vercel.app/> GitHub: <https://github.com/mecqmx>

PROFESSIONAL SUMMARY

I am a Systems Engineer focused on the AI development lifecycle, specializing in QA, multimodal annotation, model evaluation, and testing. I deeply understand that high-quality data is the prerequisite for reliable AI systems. My core focus is applying strict human evaluation criteria to training data and model outputs, which is crucial for mitigating hallucinations, improving reasoning, and fine-tuning advanced models.

I pride myself on being well-organized and consistent in my work, strictly following project guidelines and maintaining clear communication with the team. As a collaborative professional open to feedback, I am currently learning Python, Next.js, and LLM APIs to implement AI in real-world applications. My long-term goal is to grow across the full AI lifecycle.

CORE COMPETENCIES

AI Model Evaluation & Alignment	<ul style="list-style-type: none">• Multimodal RLHF across NLP, TTS, voice, and image recognition.• LLM alignment, response ranking, and red-team testing.• Supervised Fine-Tuning (SFT) and ground-truth dataset creation.• Evaluation across reasoning, creativity, safety, and contextual accuracy.
Data Collection & Quality Assurance	<ul style="list-style-type: none">• Python-based data scraping and cleaning.• Manual and automated data collection.• High-fidelity audio and video capture and content creation.• Annotation, transcription, translation, and structured review.• Dataset validation and QA workflows.
Prompt Engineering & Applied AI	<ul style="list-style-type: none">• Prompt design, testing, and optimization to enhance LLM performance.• Workflow automation prototyping through LLM API integration and Vercel deployment.
Software Development	<ul style="list-style-type: none">• Next.js (App Router) full-stack application development.• TypeScript (and JavaScript).• SQL and relational database design.• Python scripting for automation and data processing.• Tailwind CSS.• Git/GitHub version control.
Linguistics & Localization	<ul style="list-style-type: none">• Bilingual English/Spanish – MX.• Cultural adaptation and MX market localization.• Parallel corpora review.• Speech data accuracy and transcription quality control.

PROFESSIONAL EXPERIENCE

Data Annotation/QA & Model Alignment (Contractor) - Alignerr by Labelbox | Remote
April 2025 – Present

- Perform side-by-side (SxS) evaluations to ensure compliance with specific project criteria and QA standards.
- Conduct multimodal roleplay simulations (audio, video, text) to effectively stress-test AI models under real-world conditions.
- Assess AI responses across diverse criteria, including creativity, memory, and emotional awareness, providing high-quality, actionable feedback for model improvement (RLHF).
- Flag non-compliant content to ensure adherence to project guidelines and safety standards.
- Record and review natural speech samples to train and improve voice and speech recognition systems, ensuring compliance with project-specific guidelines for intonation, pronunciation, and audio quality.
- Transcribe low-quality, complex audio samples to improve advanced speech recognition systems.
- Report technical issues and collaborate with teams to ensure efficient troubleshooting and smooth project workflows.
- Maintain clear communication and share insights with team members regarding updates to project guidelines.

AI Training Data Annotator (Contractor) - Appen | Remote June 2024 – May 2025

- Followed detailed guidelines for diverse AI training tasks across speech, image, and text modalities.
- Contributed effectively to multilingual speech recognition systems by accurately reading and recording phrases with mixed accents and pronunciations.
- Evaluated and corrected AI-generated transcriptions, contributing to improved model accuracy.
- Captured and labeled images, enhancing image recognition technologies.
- Assessed the relevance and quality of search engine results to enhance performance.
- Flagged technical issues and collaborated with support teams for efficient troubleshooting.

PROJECTS & PORTFOLIO

- **Gemini AI Powered Content Generator:** An interactive application for real estate marketing content generation using Next.js and Google's Gemini AI API.
- **Next.js 16.1.1 Dashboard:** A responsive dashboard application built using the official documentation with Next.js App Router, featuring data fetching, streaming, and server actions.

EDUCATION & CERTIFICATIONS

- **Bachelor's Degree in Computer Systems Engineering.**
Instituto Tecnológico de Mérida, Mexico (2012).
- **Diploma of International Business.**
Greystone College, Sydney, Australia (2022).
- **PTE Academic English Proficiency Certification.**
Pearson Professional Centers, Sydney, Australia (2022).

ADDITIONAL INFORMATION

- Personal computer.
- Audio and video recording equipment.
- Reliable high-speed internet connection.
- Available to start immediately.