Lorenzzo Santhiago Lima

AI Trainer & LLM Evaluator

Florianópolis, Brazil | lorensanth@gmail.com

Summary

AI trainer with 14 months of freelance experience at RWS reviewing prompts, checking claims in model responses, and evaluating how models use earlier messages in a conversation.

Experience

Freelance AI Trainer & LLM Evaluator, RWS

Jul 2025 – Sep 2026 (approx.)

- Ranked prompts by how clearly they described the task and its requirements.
- Checked factual claims in model responses and flagged errors or claims without enough evidence.
- Checked whether model responses used the right details from earlier messages, including corrections from the user.
- Transcribed material for AI training.

Skills

Prompt ranking, factual verification, instruction following, multi-turn context review, transcription

Languages

Brazilian Portuguese — Native
English — Fluent

Prompts and factual accuracy

Portfolio | Fictional examples for demonstration

Choosing a prompt for a billing handoff

The billing team needs a short summary it can act on. Both prompts receive the same support transcript.

Prompt A

“Resuma a conversa de atendimento e explique o que aconteceu.”

Prompt B

“Resuma a conversa em até três tópicos para a equipe de cobrança. Inclua o problema, o que já foi verificado e a ação pendente. Se faltar algum dado, escreva ‘não informado’.”

I would choose B. If the transcript ends before support checks the charge, B tells the model to mark that gap. A leaves it open whether the summary should include a pending action at all.

Checking a refund answer against policy

A customer asks whether they can request a refund 12 days after purchase and when they might receive it. The policy allows requests within 14 calendar days of purchase. Approved refunds are usually processed within five business days.

Reply under review

“Sim, o reembolso tá garantido e cai na conta em cinco dias corridos.”

The customer is still within the request window. Approval is pending, and the five-day estimate refers to processing in business days. The policy gives no date for the money to reach the account.

Suggested reply

“Sim, ainda dá pra pedir o reembolso. Se for aprovado, o processamento costuma levar até cinco dias úteis.”

Conversation and transcription

Portfolio | Fictional examples for demonstration

A meeting changes after the first message

The user asks for a short, informal message to Marina confirming Thursday at 2 p.m. and asks for no emoji. Then the user sends an update.

User update

“Corrigindo: mudou pra sexta às 10h. Mantém o resto.”

Model reply

“Oi, Marina! Confirmando nossa reunião de quinta às 14h. Até lá!”

The draft still sounds right for Marina, but it confirms the old appointment. Change the day and time to Friday at 10 a.m. Keep the rest of the message as written.

A self-correction in a transcript

This exercise uses a mock speech reference. The transcription brief calls for preserving self-corrections and marking unclear speech as [inaudível].

Mock speech reference

“Eu mandei na terça... quer dizer, na quarta, pro [inaudível].”

Transcript under review

“Eu mandei na terça pro financeiro.”

The speaker corrected Tuesday to Wednesday. The recipient is unclear, so “financeiro” needs to come out. Restore the spoken correction and keep the uncertainty marker where the recipient would appear.

When reviewers disagree

Portfolio | Fictional calibration exercise

A fluent answer makes an unsupported promise

Two reviewers are ranking replies to a customer who asks whether an order will arrive tomorrow. The order record shows a shipping label was created. It contains no carrier scan or delivery estimate.

Reply A

“Sim, chega amanhã. O código de rastreio já foi gerado.”

Reply B

“A etiqueta de envio já foi gerada, mas ainda não temos uma previsão de entrega.”

One reviewer prefers A because it directly answers the question. The other prefers B because the order record does not support a delivery date. If the rubric gives factual accuracy priority, B ranks higher. The exact error in A is the promise of delivery tomorrow.

Calibration note to the project lead

“I agree that a direct answer is useful when we have an estimate. Here the record only confirms the label. Can we use B as the reference for this example and clarify what counts as evidence for a delivery date? I’d check the rule on a second item where the carrier does provide an estimate, so we don’t teach reviewers to reject supported dates.”

How the decision would be recorded

Preferred reply: B. Evidence: label created, no carrier estimate. Error in A: unsupported delivery date. If the rubric does not say how to resolve the disagreement, flag the item for the lead before using it as a reference for other reviewers.

lorensanth@gmail.com