

## Rob Otto

**Ingeniero de Machine Learning & Especialista en IA** 📍 Buenos Aires, Argentina

(Disponibilidad Remota/Híbrida) ✉ rob.otto@email.ficticio | 🔗

[linkedin.com/in/rob.otto.no.me.busquen](https://linkedin.com/in/rob.otto.no.me.busquen) | 🐙 [github.com/rob.otto.noexiste.perfil](https://github.com/rob.otto.noexiste.perfil)

### Perfil Profesional

Ingeniero de Software especializada en Inteligencia Artificial con más de **4 años de experiencia** en el ciclo de vida completo de modelos de Machine Learning. Experta en el desarrollo de soluciones de **Procesamiento de Lenguaje Natural (NLP)** y **Visión por Computadora**, con un enfoque reciente en la implementación de **LLMs (Large Language Models)** en entornos productivos. Apasionado por transformar la investigación académica en soluciones de negocio escalables y eficientes.

---

### Habilidades Técnicas

- **Lenguajes:** Python (Experto), SQL, C++, Bash.
- **Deep Learning & ML:** PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, Keras, XGBoost.
- **IA Generativa:** Fine-tuning de LLMs (Llama 3, Mistral), **RAG (Retrieval-Augmented Generation)**, LangChain, Prompt Engineering.
- **MLOps & Cloud:** AWS (SageMaker, Lambda), Azure ML, **Docker**, Kubernetes, MLflow, DVC (Data Version Control).
- **Bases de Datos:** PostgreSQL, MongoDB, Pinecone (Vector DB), ChromaDB.
- **Matemáticas:** Álgebra Lineal, Cálculo Multivariable, Estadística Bayesiana.

---

### Experiencia Laboral

#### Senior AI Engineer | TechNova Solutions (Fintech)

*Noviembre 2022 – Presente*

Lidero la iniciativa de automatización de atención al cliente utilizando IA Generativa.

- Diseñé e implementé un sistema de **RAG (Retrieval-Augmented Generation)** utilizando LangChain y OpenAI API, reduciendo el tiempo de respuesta de soporte en un **45%**.
- Desarrollé un pipeline de MLOps en **AWS SageMaker** para reentrenar automáticamente modelos de detección de fraude, mejorando la precisión (F1-Score) del 88% al **94%**.

- Optimicé costos de inferencia mediante la cuantización de modelos (int8) y el uso de ONNX Runtime, logrando un ahorro del **30% en costos de nube**.
- Mentora técnica de 2 ingenieros junior en buenas prácticas de código y despliegue de modelos.

## **Data Scientist | Retail Analytics Corp**

*Junio 2020 – Octubre 2022*

Encargado del desarrollo de modelos predictivos para optimización de inventario.

- Creé modelos de **Time Series (Prophet y LSTM)** para predecir la demanda de stock en 500 sucursales, reduciendo el sobre-stock en un **15%**.
- Implementé un sistema de recomendación híbrido (filtrado colaborativo + basado en contenido) que aumentó el ticket promedio de compra online en un **12%**.
- Automaticé la limpieza y preprocesamiento de datos (ETL) utilizando **Apache Airflow** y SQL, ahorrando 10 horas semanales de trabajo manual al equipo de datos.

---

## **Proyectos Destacados**

### **LegalSumm AI (Proyecto Personal)**

- Desarrollo de una herramienta SaaS para resumir contratos legales complejos.
- Fine-tuning de un modelo **Mistral-7B** utilizando LoRA (Low-Rank Adaptation) con un dataset legal en español.
- Despliegue de la API utilizando **FastAPI** y Docker en una instancia GPU de bajo costo.

### **Sistema de Detección de Defectos (Visión por Computadora)**

- Entrenamiento de una red **YOLOv8** para detectar fallas en líneas de producción manufacturera en tiempo real.
- Integración con cámaras IoT y procesamiento en el borde (Edge Computing) usando NVIDIA Jetson.

---

## **Educación**

**Máster en Inteligencia Artificial y Ciencia de Datos** Universidad de Buenos Aires (UBA) |  
*2019 – 2021*

- Tesis: "Optimización de Transformers para textos médicos en español".

**Licenciatura en Ciencias de la Computación** Instituto Tecnológico de Buenos Aires (ITBA) |  
*2015 – 2019*

- Promedio destacado.

---

## Idiomas

- **Español:** Nativo.
- **Inglés:** C1 - Avanzado (Capacidad para leer papers técnicos y trabajar en equipos internacionales).