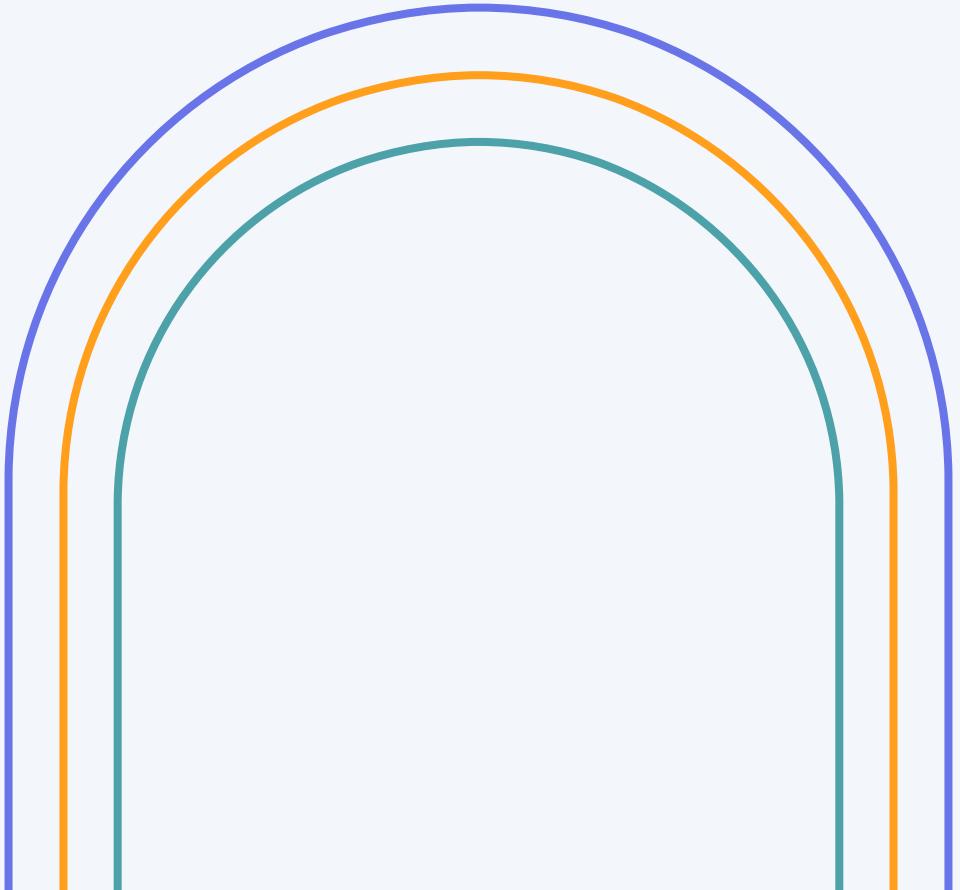
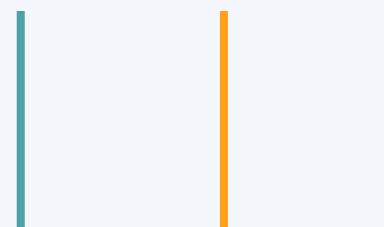


---



# Introducción a MLOps

¿Cómo gestionar tus proyectos  
de ciencia de datos?



# Sobre mi

Jose Alberto Arango Sánchez

Machine Learning Engineer at Konecta

Me interesa la ciencia de datos, ML, DL, Voice, ciencia y software.



[jose.arangos@udea.edu.co](mailto:jose.arangos@udea.edu.co)



1. ¿Qué es MLOps?
2. ¿Para qué MLOps?
3. Academia vs Industria.
4. Ciclo de vida de un proyecto de Machine Learning.
5. Etapas de adopción de MLOps.
6. Grados de automatización
7. Describe el problema.
8. Herramientas
9. Recursos de interés



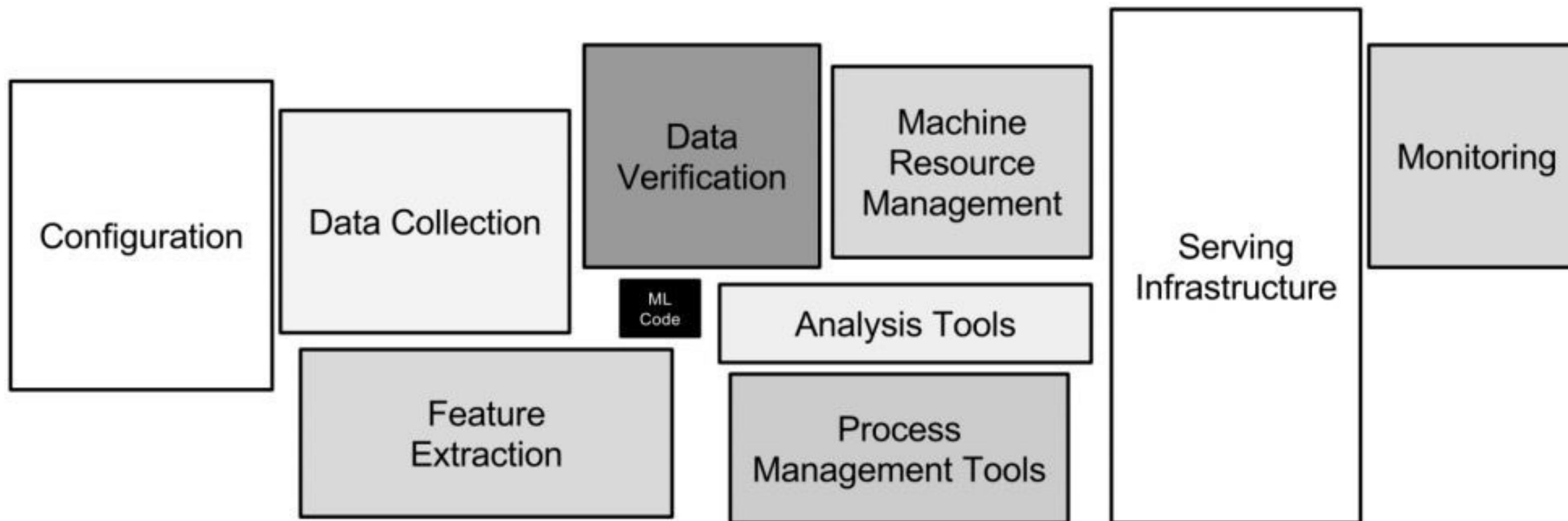
**Tabla de  
contenido**

---



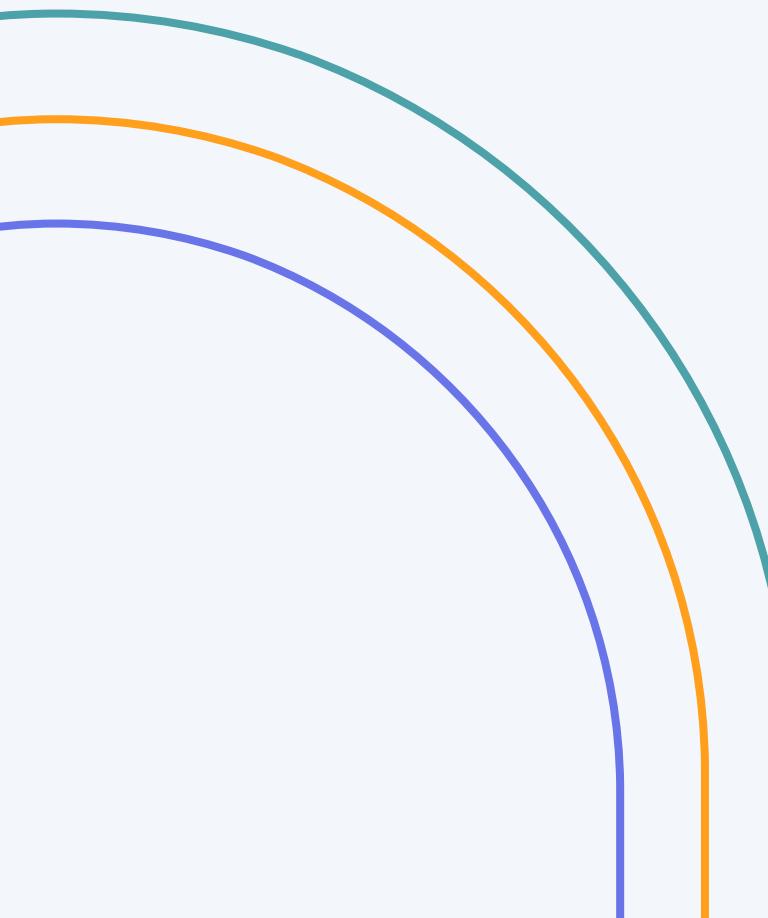
# Hidden Technical Debt in Machine Learning Systems

**D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips**  
{dsculley, gholt, dg, edavydov, toddphillips}@google.com  
Google, Inc.





# ¿Qué es MLOps?





# ¿Qué es el MLOps?

MLOps es un conjunto de prácticas y herramientas que permiten automatizar el ciclo de vida de un modelo de Machine Learning.

# ¿Qué es el MLOps?

---



Datos

Código



Lifecycle

ML model

# MLOps y herramientas



Herramientas



Mentalidad



Trabajo en equipo

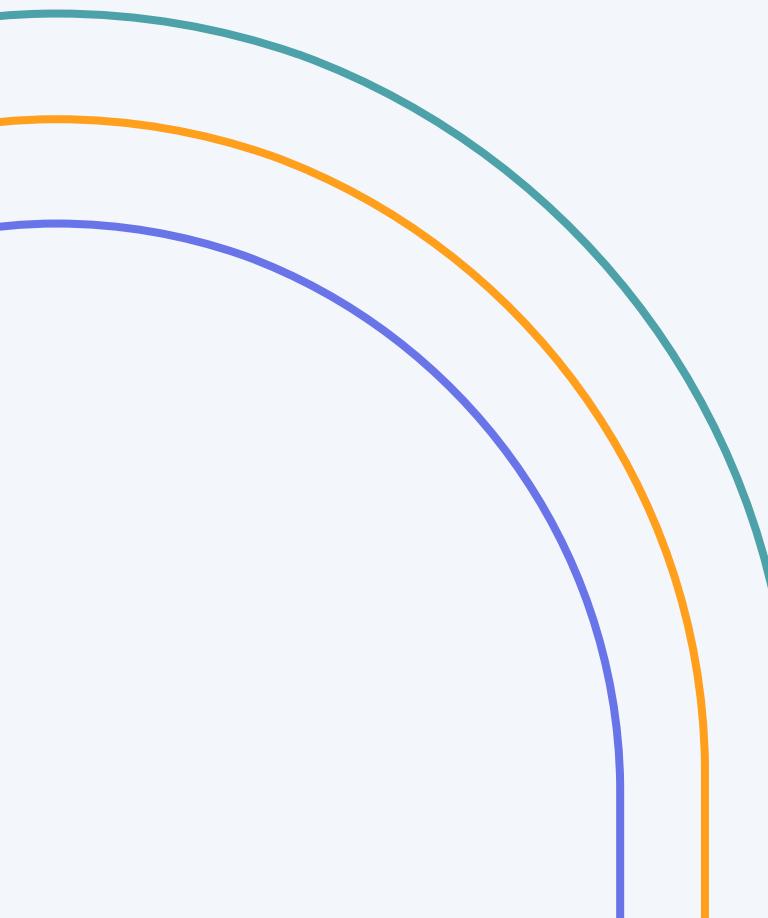


Cultura





# ¿Para qué MLOps?



# ¿Has hecho esto?

- Entrenar y usar un modelo en Jupyter
- Poner manualmente un modelo en Docker
- Enviar tu modelo por correo electrónico



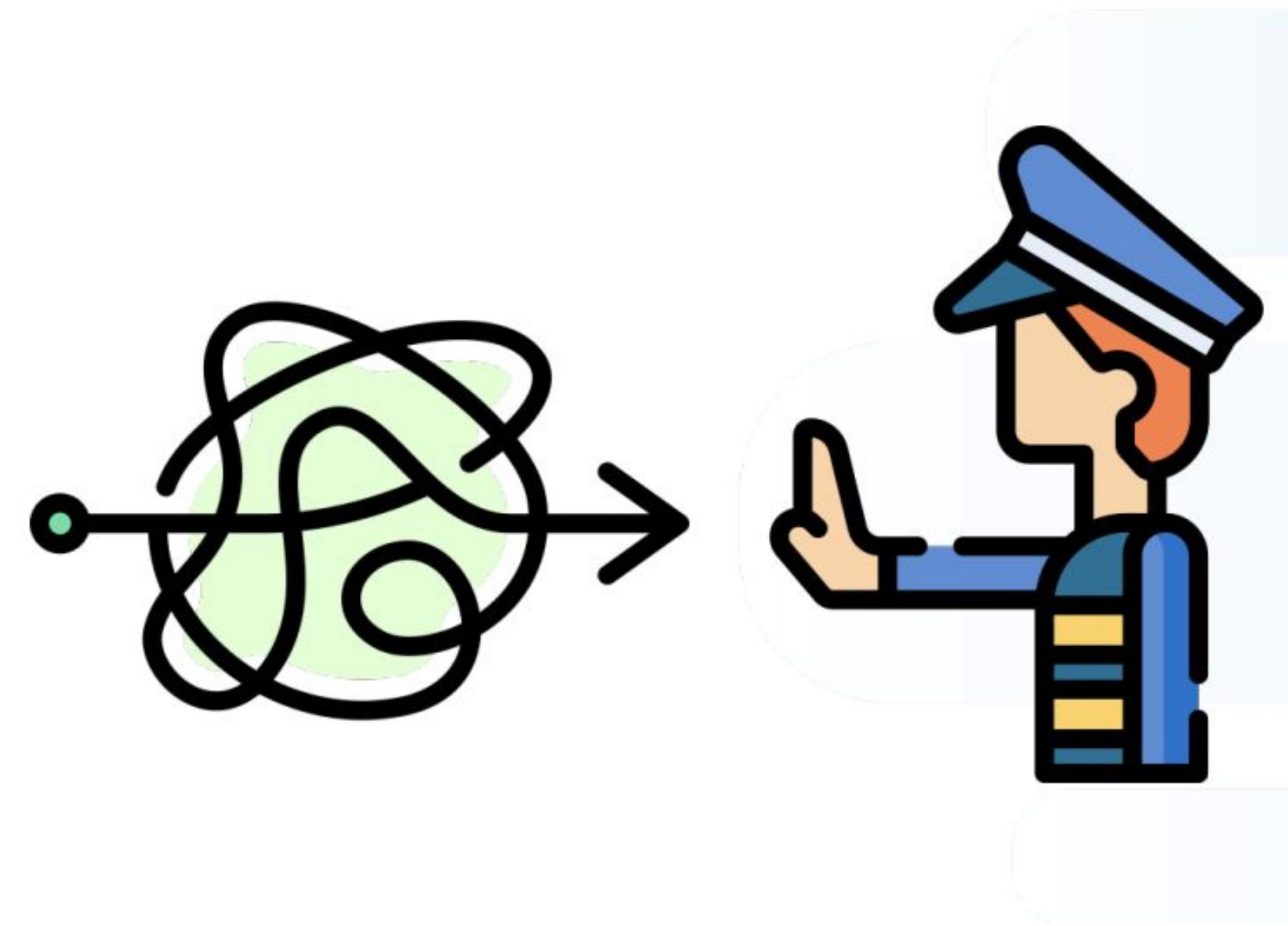
# ¿Has hecho esto?

- Entrenar y usar un modelo en Jupyter
- Poner manualmente un modelo en Docker
- Enviar tu modelo por correo electrónico

**No son buenas prácticas en la industria!!!!**



# MLOps para reducir riesgos

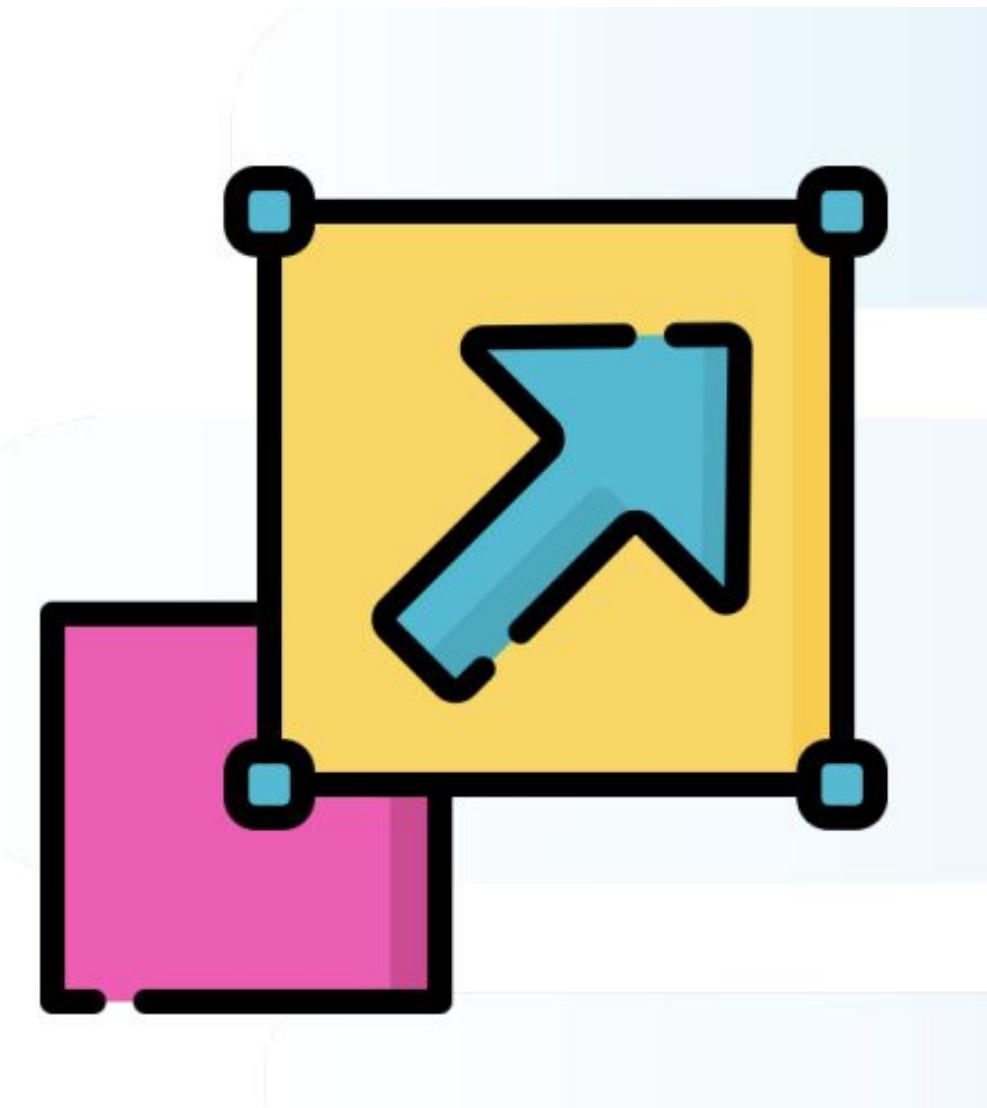


# MLOps para ahorrar tiempo

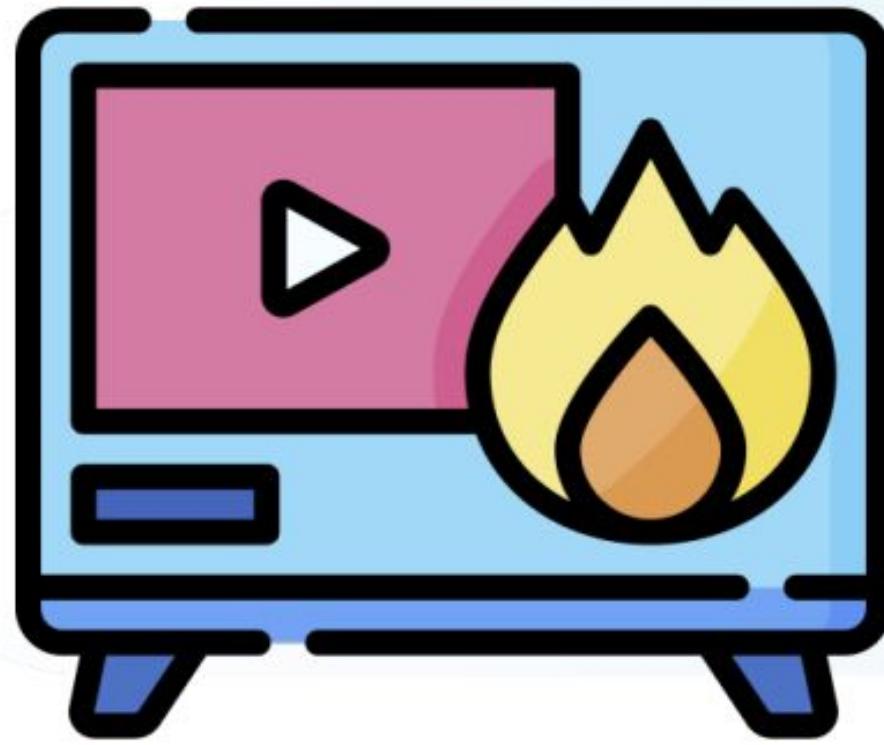
*Ship  
Faster*



# MLOps para escalar

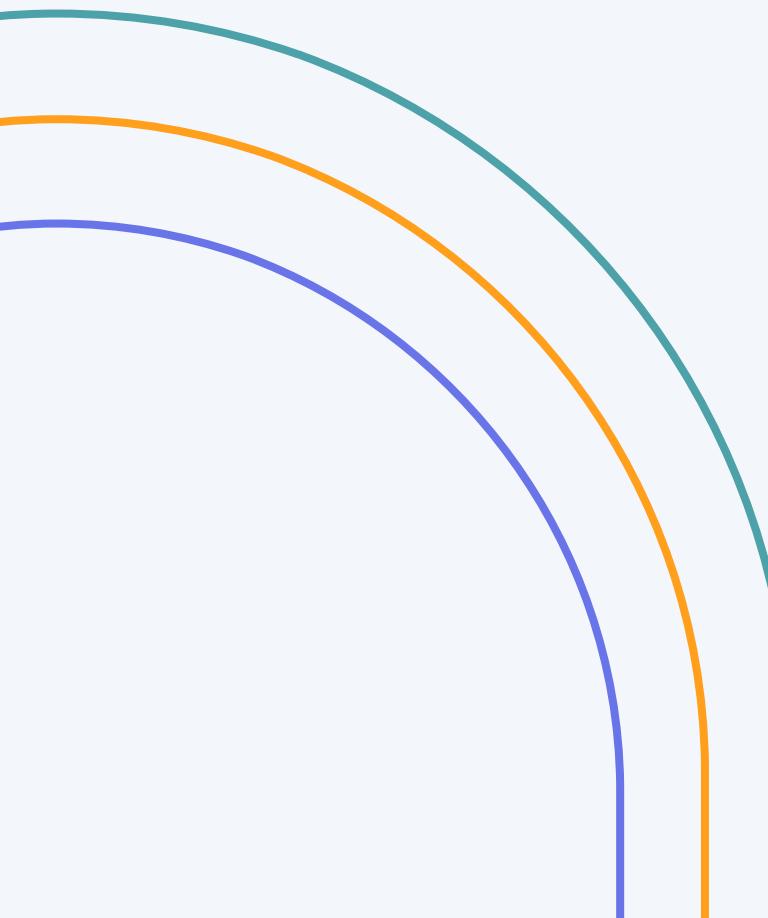


# MLOps para adaptarnos





# Academia vs Industria.



# Academia vs Industria

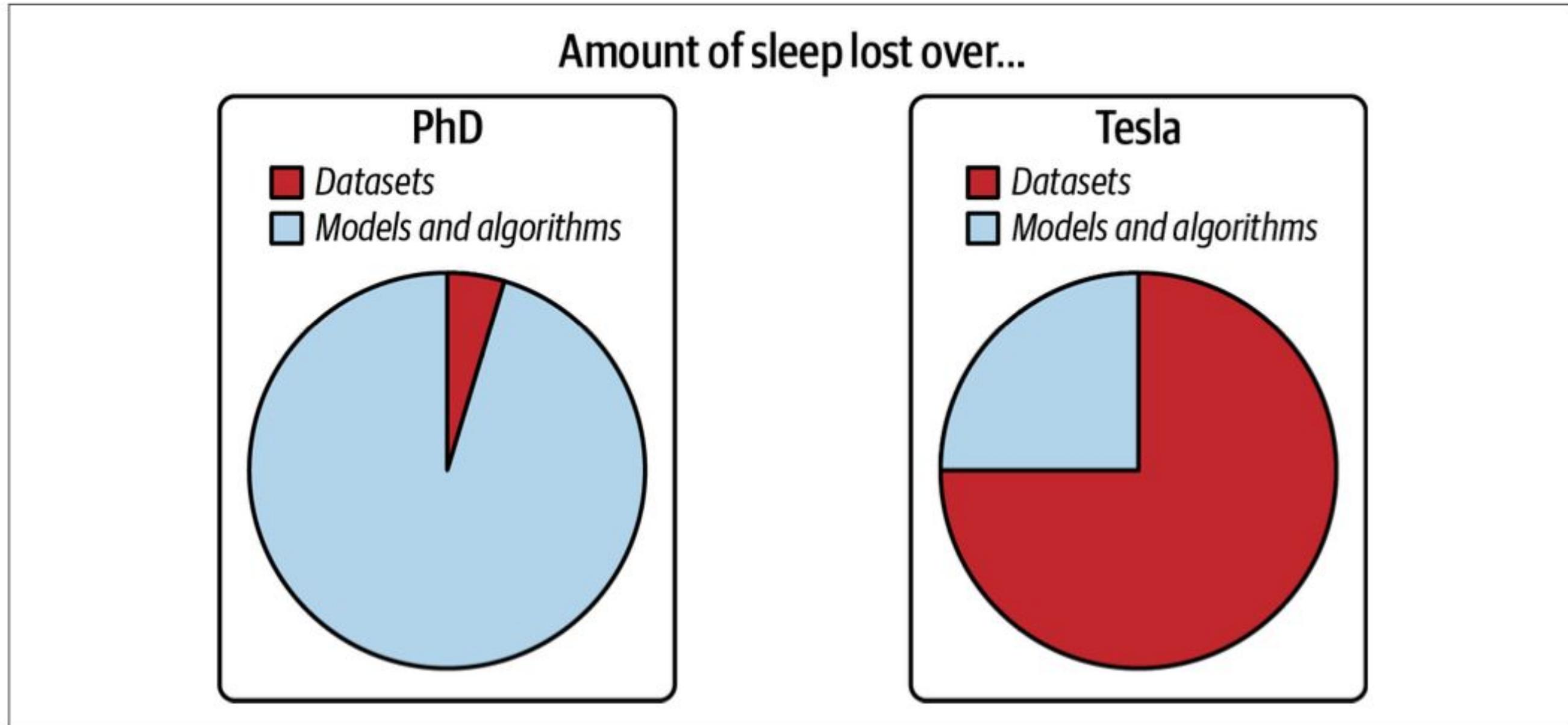
*Table 1-1. Key differences between ML in research and ML in production*

	Research	Production
Requirements	State-of-the-art model performance on benchmark datasets	Different stakeholders have different requirements
Computational priority	Fast training, high throughput	Fast inference, low latency
Data	Static <sup>a</sup>	Constantly shifting
Fairness	Often not a focus	Must be considered
Interpretability	Often not a focus	Must be considered

<sup>a</sup> A subfield of research focuses on continual learning: developing models to work with changing data distributions. We'll cover continual learning in [Chapter 9](#).



# Academia vs Industria



*Figure 1-5. Data in research versus data in production. Source: Adapted from an image by Andrej Karpathy<sup>24</sup>*



# Tiempos de trabajo

- En la academia, los investigadores se dedican a un solo problema hasta obtener resultados y luego publicarlos.
- En la industria no se busca la perfección, sino algo funcional que pueda mejorarse con el tiempo.
- Los profesionales de la industria pueden estar involucrados en varios proyectos al mismo tiempo y estos pueden ser pausados o re-priorizados según el mercado.



# Evaluación

- En la academia, se evalúa el desempeño del modelo en una métrica específica, sin importar otros aspectos.
- En la industria, se deben considerar múltiples factores adicionales al rendimiento del modelo.
- Es importante que el modelo pueda adaptarse a la infraestructura .



# Novedad

- En la academia, se busca innovar y empujar los límites del conocimiento.
- En la industria, se prefieren métodos y técnicas ya probados.
- Implementar soluciones no probadas puede representar un riesgo para una empresa.
- Gran parte del presupuesto se destina a la estabilidad y al funcionamiento de algoritmos conocidos.



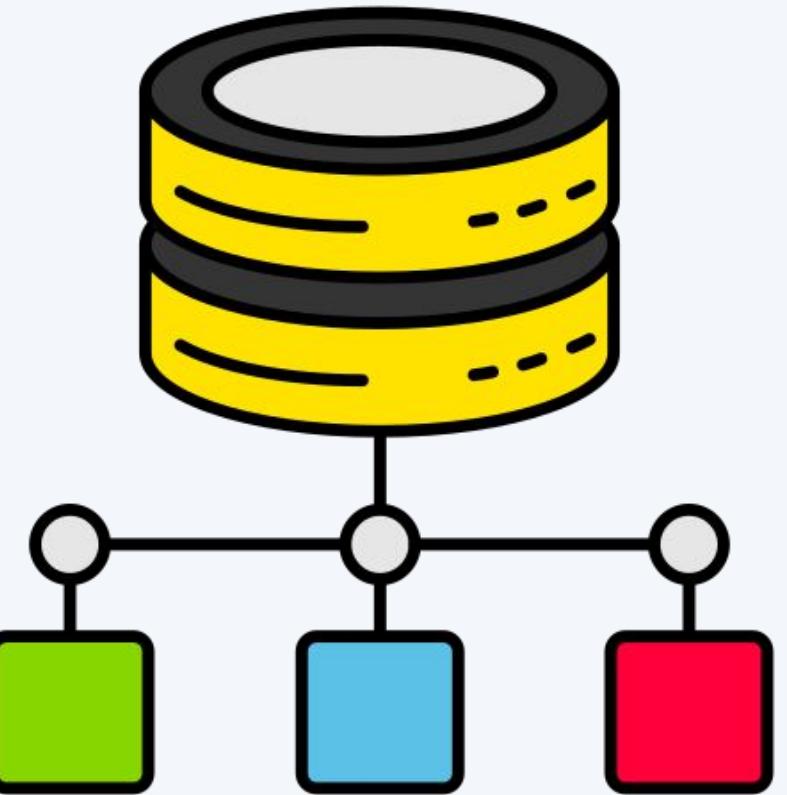
# Interpretabilidad

- En la industria, la interpretabilidad de los modelos es crítica debido a regulaciones gubernamentales y presiones sociales.
- Regulaciones como el GDPR Europeo, la PDPA de Singapur y la CCPA Californiana obligan a tomar precauciones en la interpretación de las decisiones de los modelos.



# Datos

- En la academia se evalúan nuevos modelos y algoritmos en datasets conocidos y estáticos para medir el avance en el estado del arte.
- En la industria, los datasets son personalizados y centrados en el negocio de la empresa, se actualizan con mayor frecuencia y los modelos de machine learning deben adaptarse a esos cambios.



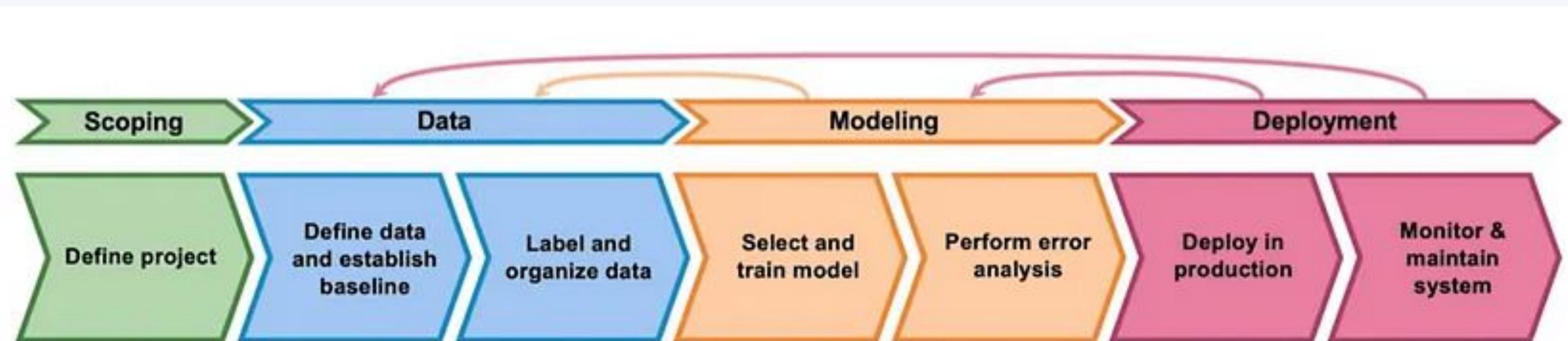


---

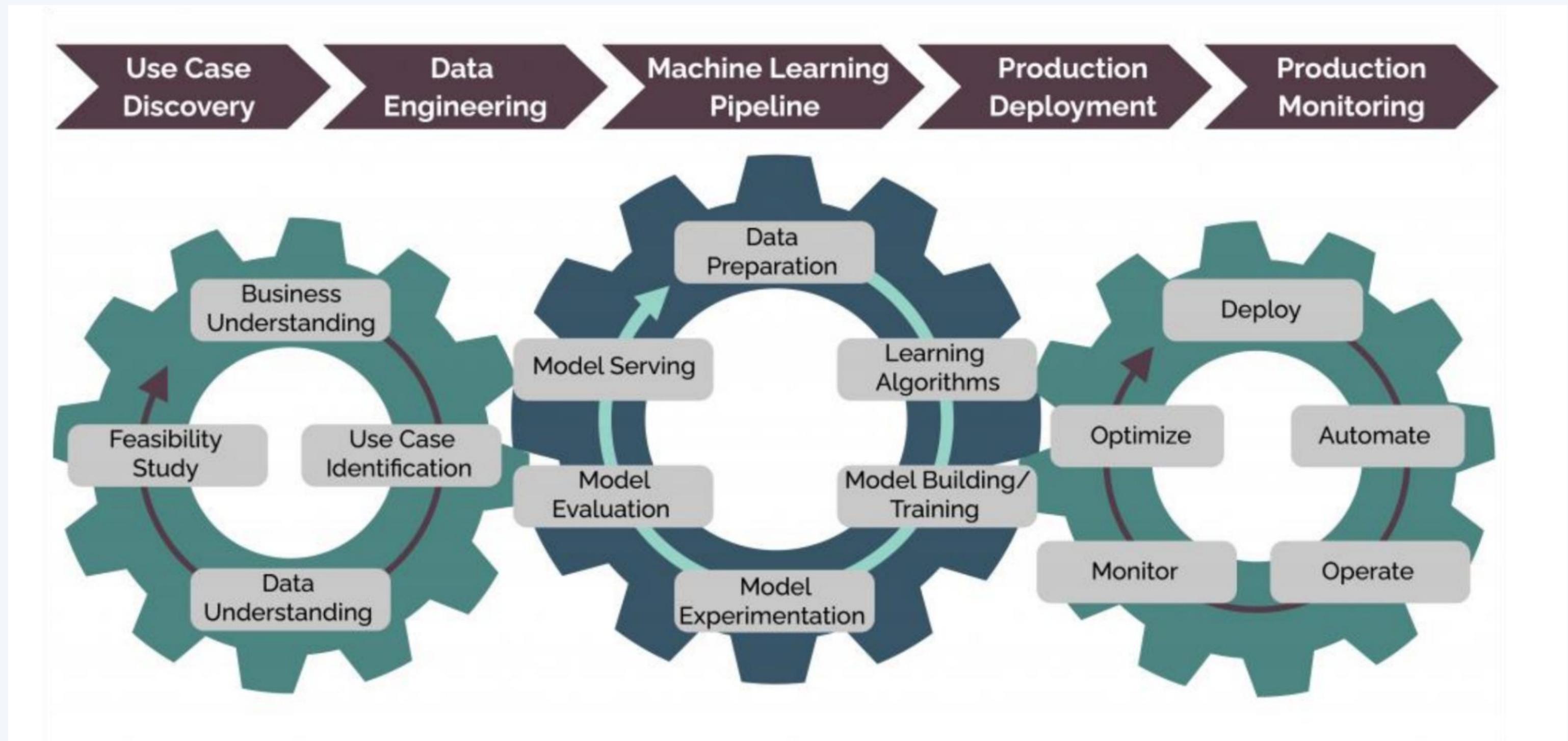
# Ciclo de vida de un proyecto de Machine Learning.



# El ciclo de vida de un proyecto de Machine Learning



# El ciclo de vida de un proyecto ML



# El ciclo de vida de un proyecto ML



Definición del  
proyecto

Organización  
de los datos

Desarrollo  
del modelo

Puesta en  
producción

Monitoreo

# El ciclo de vida de un proyecto ML



Se establece la viabilidad de un proyecto y se establece el alcance de la solución a desarrollar

# El ciclo de vida de un proyecto ML



Se extrae la información necesaria para el modelo y se automatiza la obtención de nuevos datos

# El ciclo de vida de un proyecto ML



Se crea el modelo, ayudado de las mejores prácticas de desarrollo de software

# El ciclo de vida de un proyecto ML



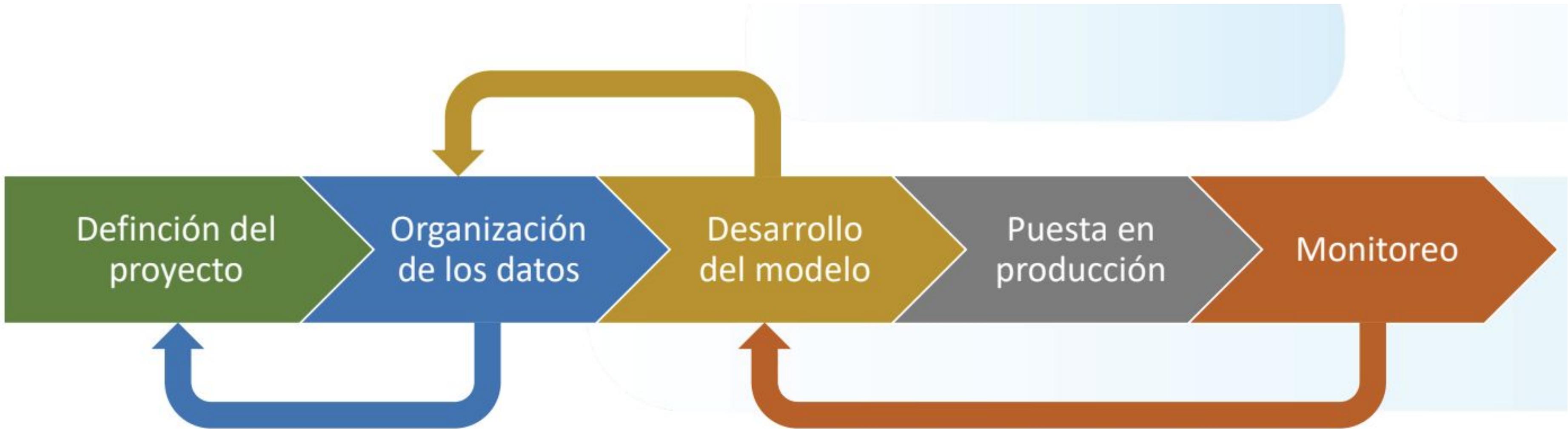
Se operacionaliza el modelo verificando que se integra correctamente con otras partes del sistema

# El ciclo de vida de un proyecto ML



Revisión continua del comportamiento del modelo y de los datos que consume

# El ciclo de vida de un proyecto ML



**NO** es un **Pipeline** de Machine Learning, son conceptos parecidos, pero no son lo mismo.

# Roles en MLOps

---



**Data scientist:**

Why is the model drifting?

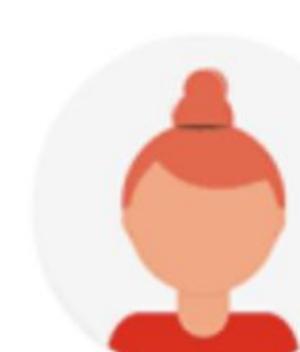


**Model user:**

Can I trust this prediction?

**Data science manager:**

Is it time to retrain?



**Business stakeholder:**

How much value does the model bring?



**Product manager:**

What are the model limitations?



**Support:**

Why did we mark this as spam?

**Data engineer:**

Does it get the right data?

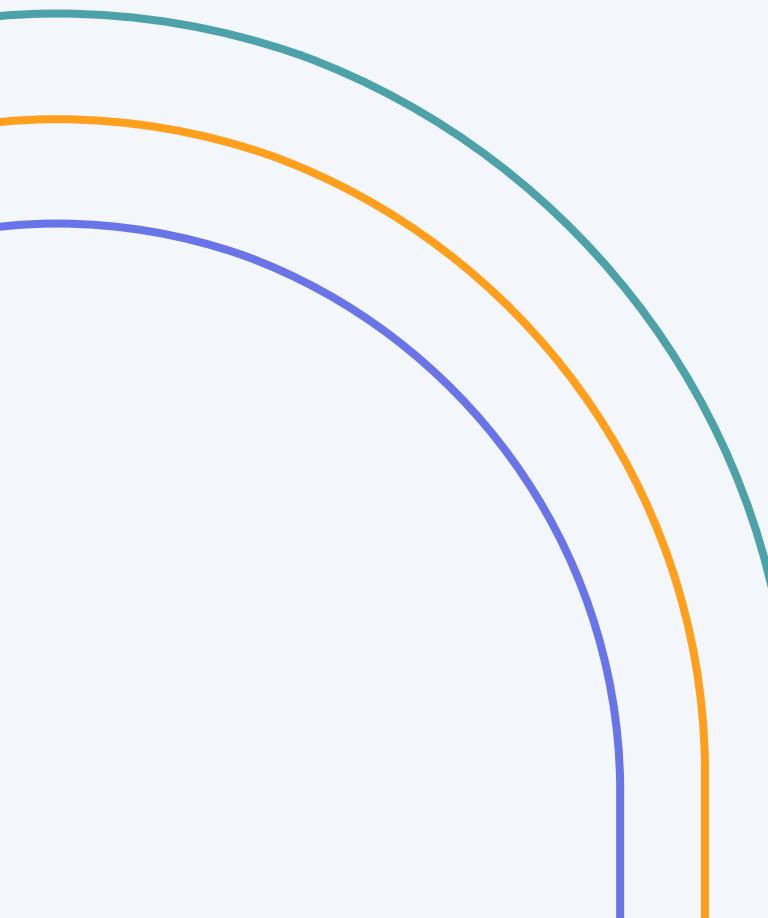


**Compliance:**

Is this model safe to use?



# Herramientas



# Problemas que resuelve



## MLOps

### Versionado

- Datos
- Código
- Modelo



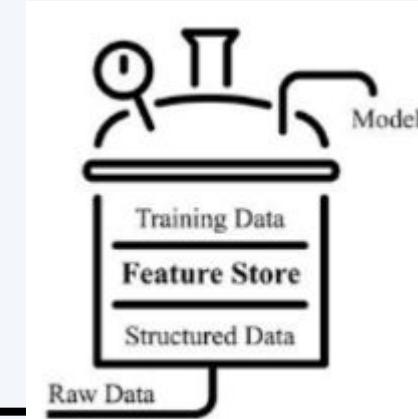
### Seguimiento del modelo

Los modelos en producción pueden degradarse con el tiempo por la deriva de datos. Garantiza la **validez** del modelo a largo plazo.



### Generación de características

MLOps garantiza que las funciones se puedan reutilizar. Permite centrarnos en el diseño/test del modelo. (**Feature store**)



# Problemas que resuelve

## MLOps



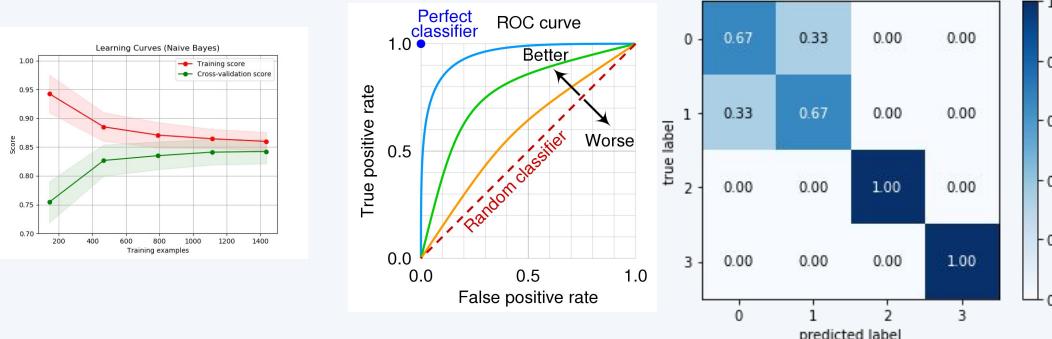
### Registro de modelos

Una vez que se ha entrenado un modelo, se almacena en un **registro de modelo** junto con sus metadatos.



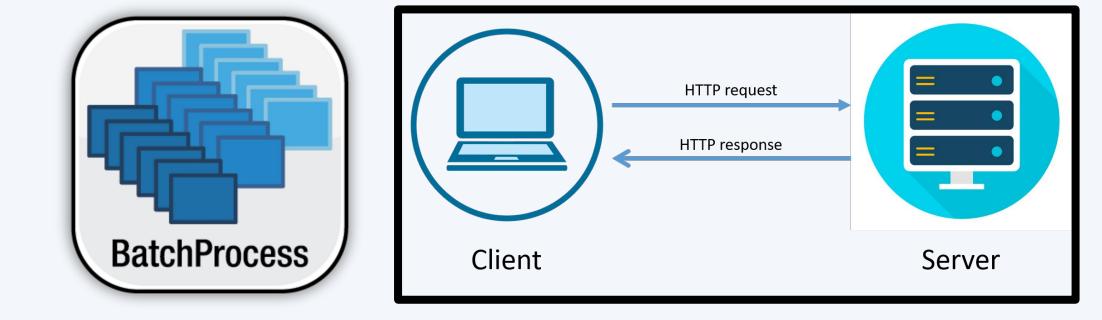
### Almacén de metadatos

El registro es fundamental para la reproducibilidad. Para ello hay que **registrar** desde el seed hasta métricas de evaluación, gráficos.



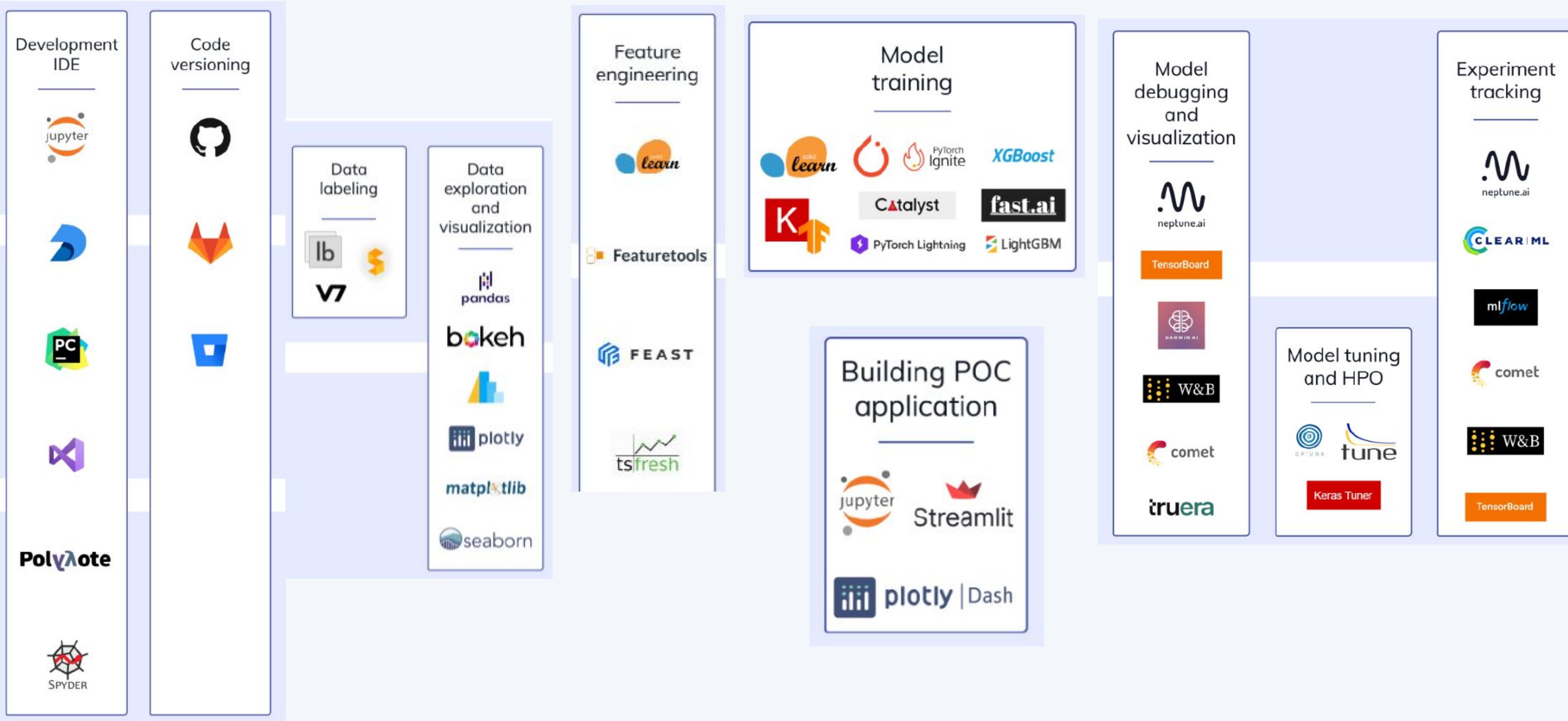
### Model serving

Servir un modelo significa crear las diferentes formas de utilizar ejecutar predicciones con el modelo: **stream, batch.**



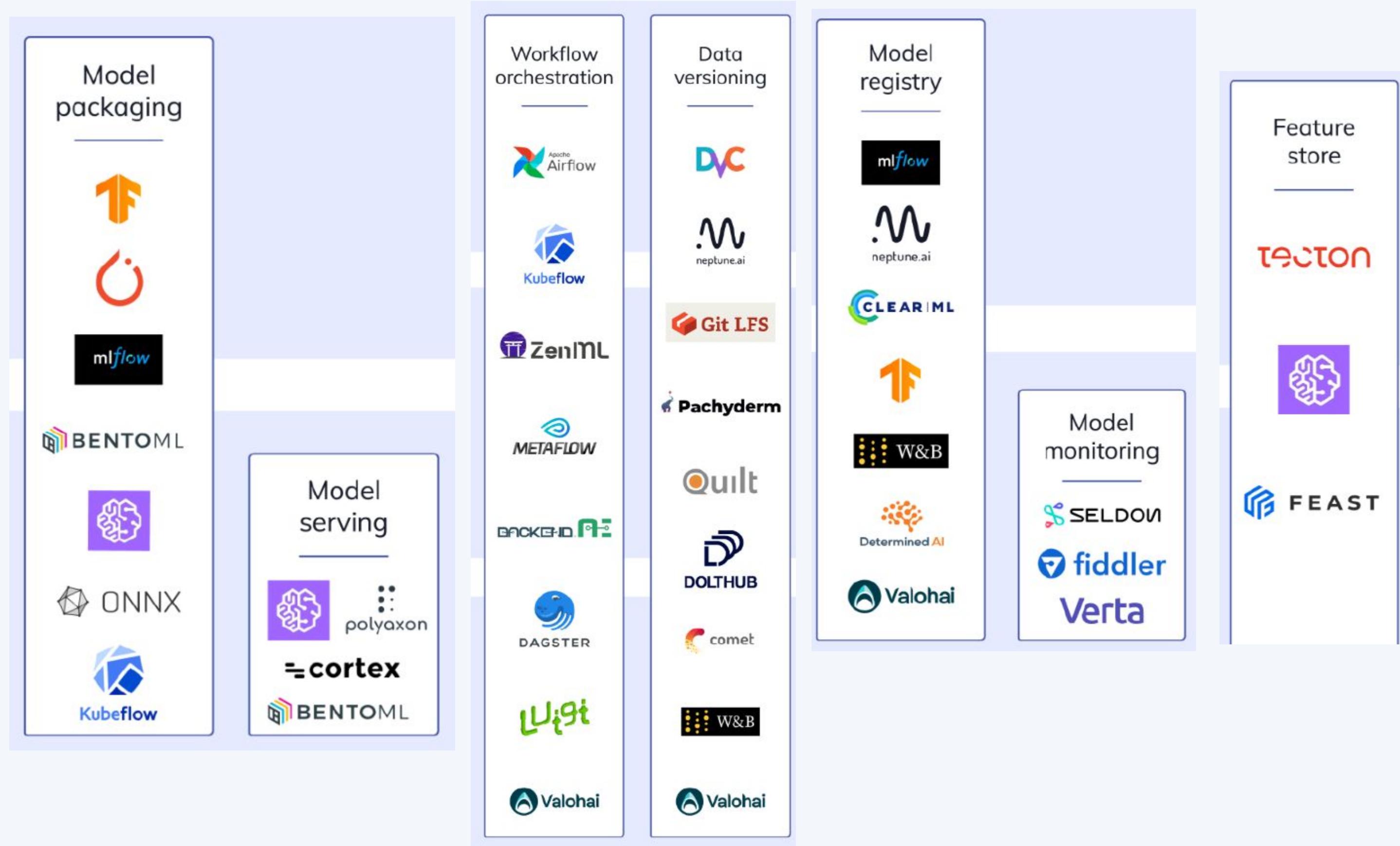


# Herramientas





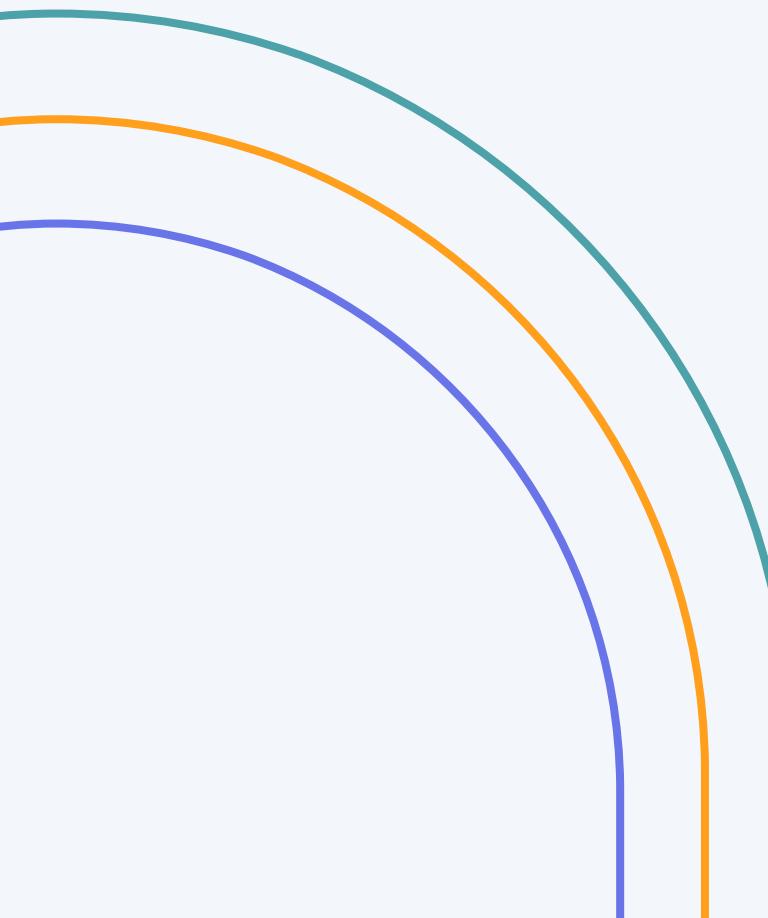
# Herramientas





---

# Etapas de adopción de MLOps.



# Etapas de adopción de MLOps.

**Etapa 1:** Control de versión de modelo y datos

**Etapa 2:** AutoML + Control de versión de código, modelo y datos

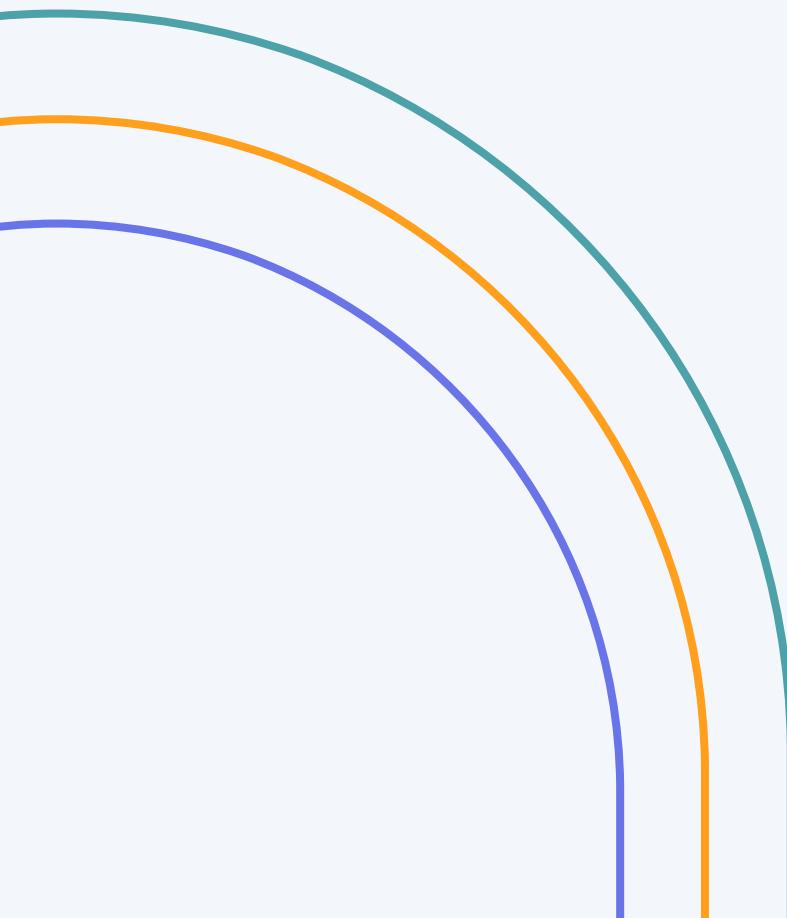
**Etapa 3:** AutoML + Control de versión de código, modelo y datos + Model Serving

**Etapa 4:** AutoML + Control de versión de código, modelo y datos + Monitorización, Gobernanza y Reentrenamiento.



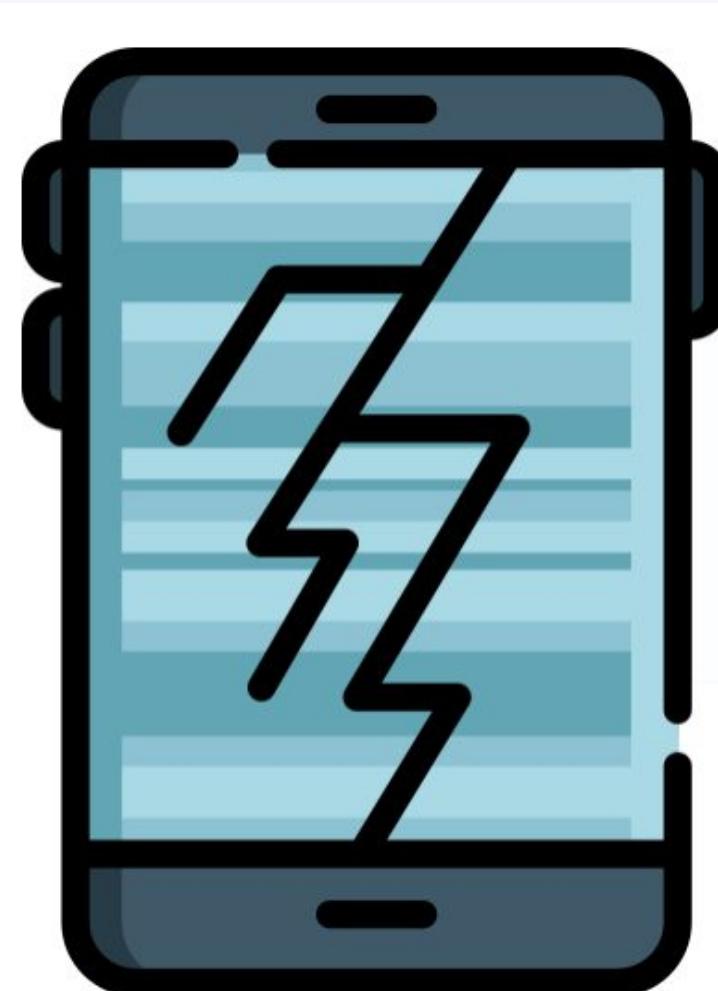


# Grados de automatización



---

# Ejemplo de grados de automatización.



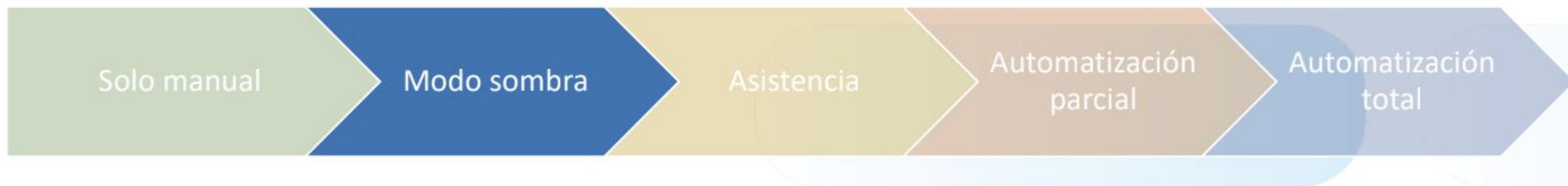
# Solo manual



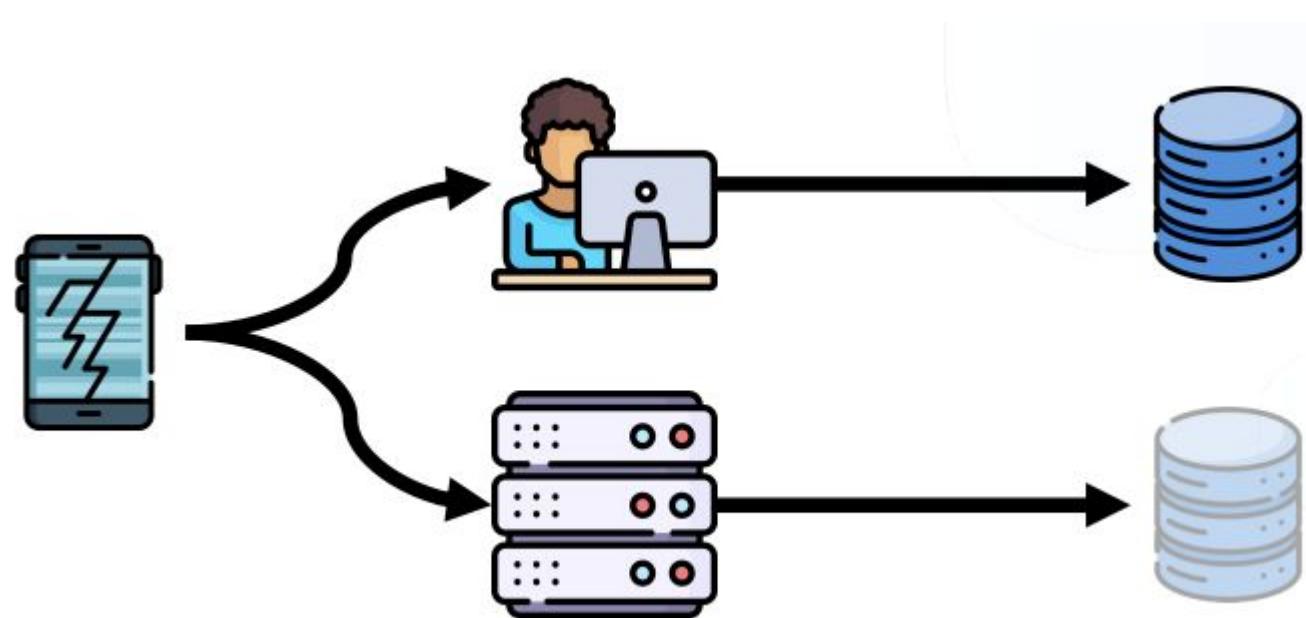
También llamado “**human only**”, empleados revisan manualmente las imágenes de las pantallas.



# Modo sombra (shadow)



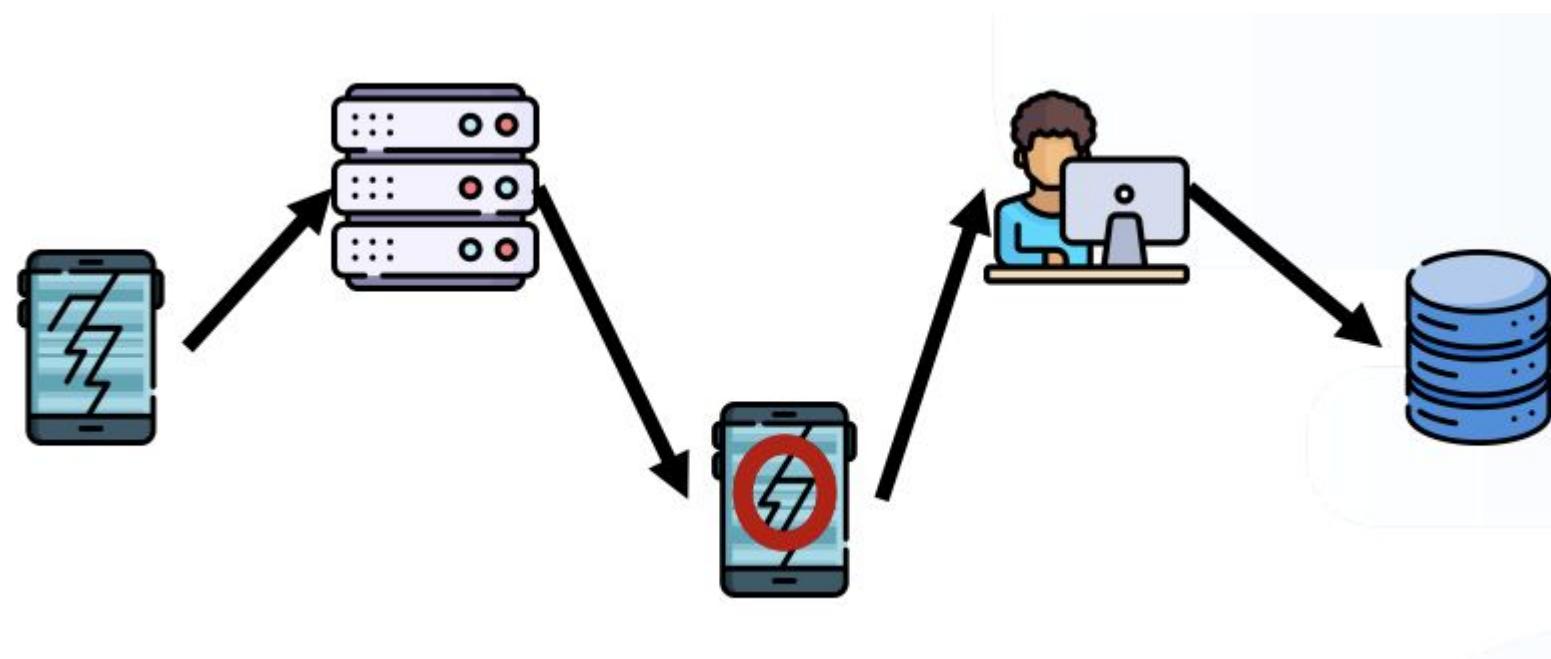
El modelo genera predicciones, pero estas **no son usadas** en producción, son solo recolectadas para análisis.



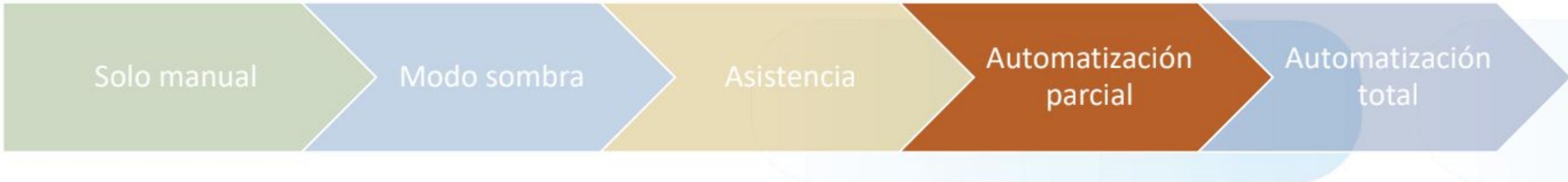
# Asistencia



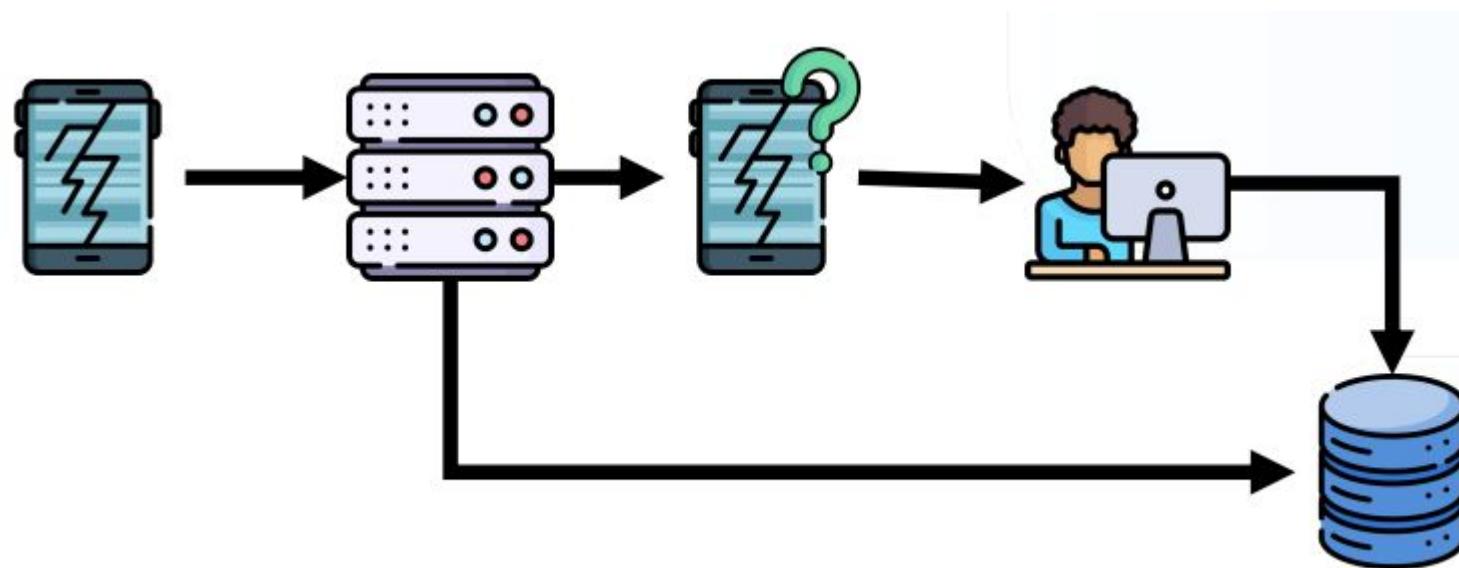
El sistema indica todos los defectos potenciales, por más pequeños que sea y los señala para los usuarios.



# Automatización parcial



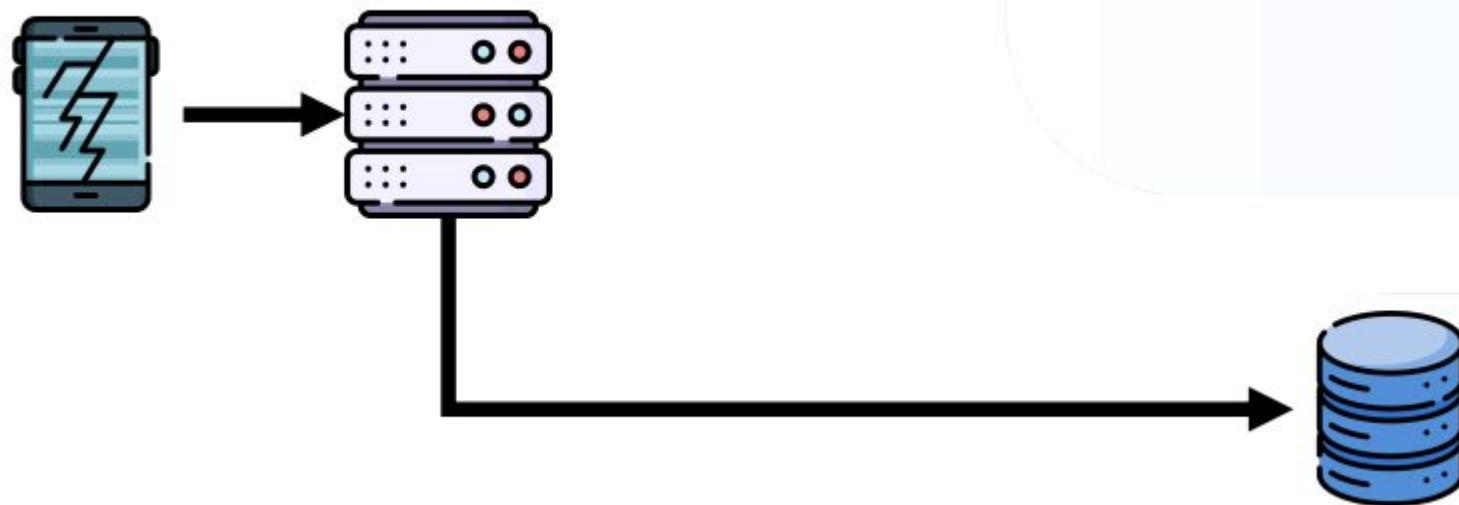
El sistema funciona de forma **correcta** en la **mayoría** de los **casos**, pero hay **ocasiones** en las que se le **difulta**



# Automatización total

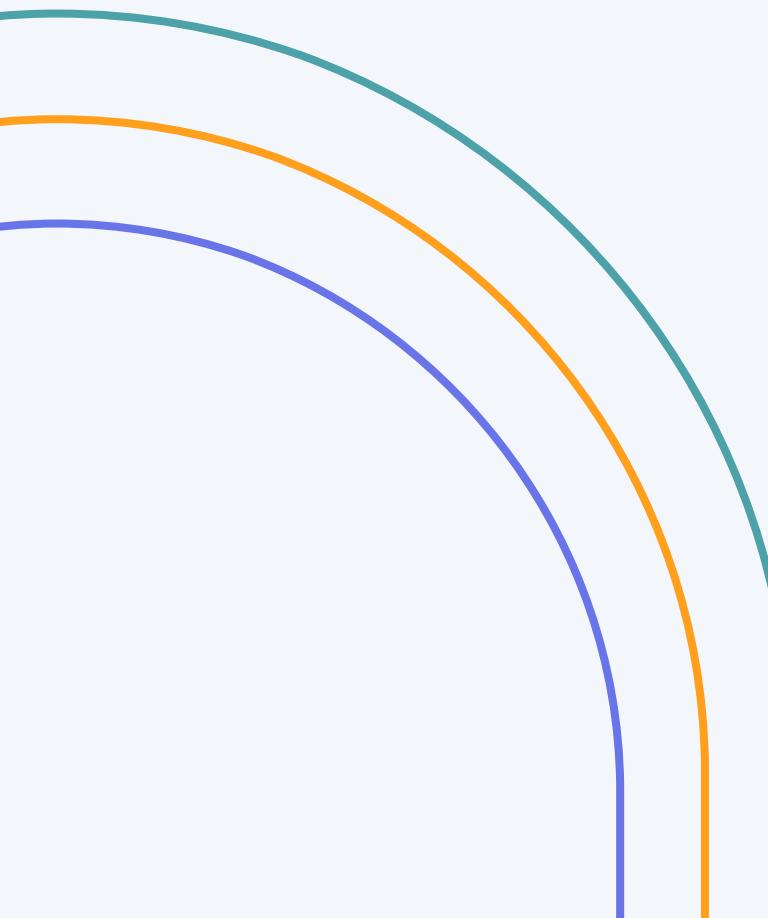


El sistema funciona automáticamente al 100%



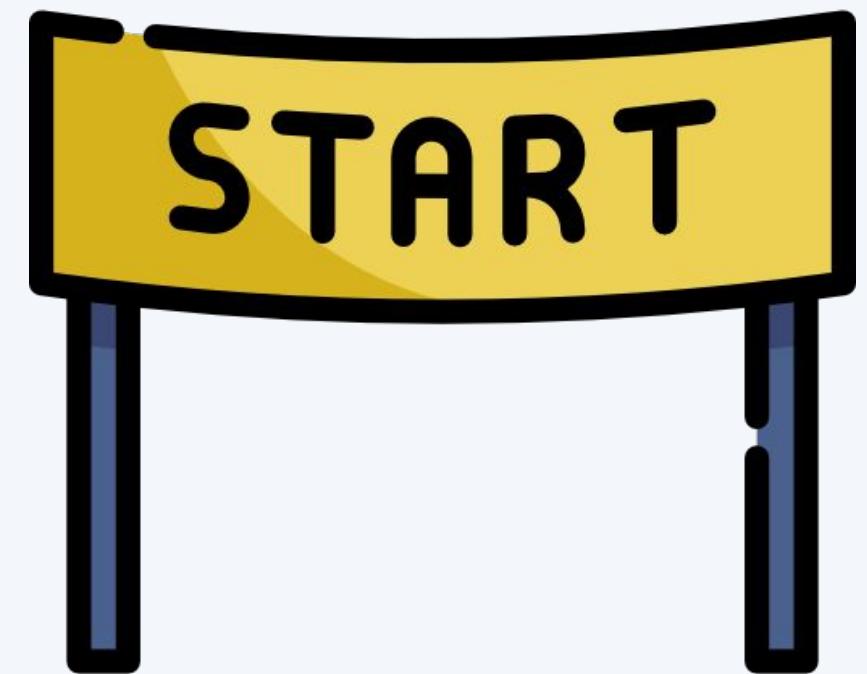


# Describe el problema.



# El primer paso en MLOps

- Debes entender el problema a solucionar
- MLOps no comienza cuando el modelo está terminado
- Habla con stakeholders, los clientes que quieren una solución.



# ¡A veces no se necesita un modelo!

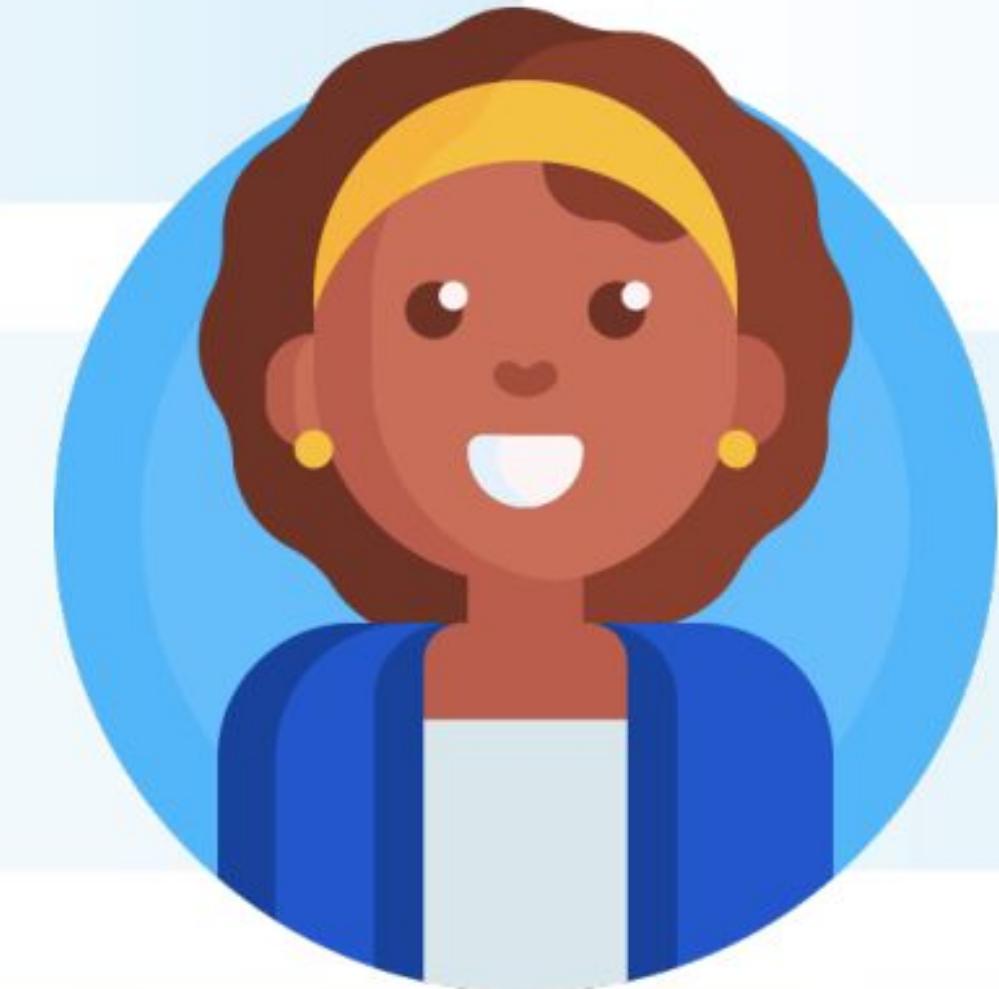
- La gente ve “machine learning” como una solución mágica
- A veces solo necesitan buenas reglas de negocio, análisis bien fundamentados o un dashboard
- Hay que entender las necesidades primero



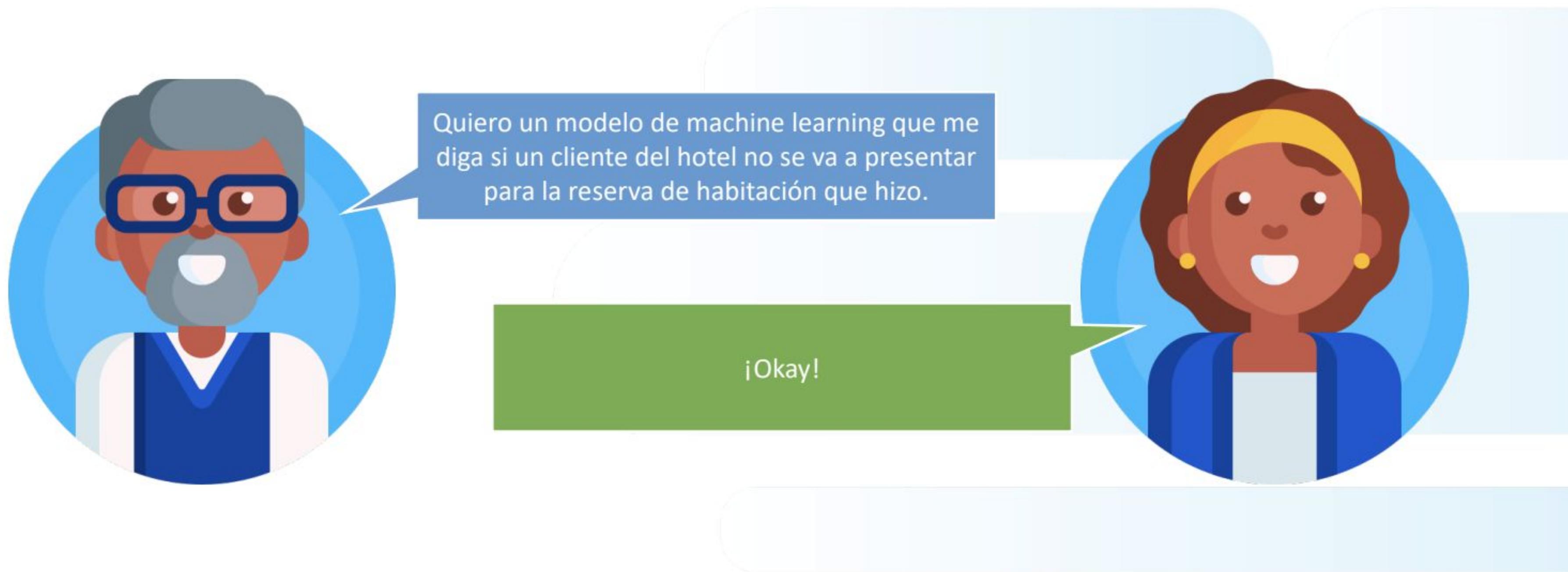
# Interacción 1



Quiero un modelo de machine learning que me diga si un cliente del hotel no se va a presentar para la reserva de habitación que hizo.



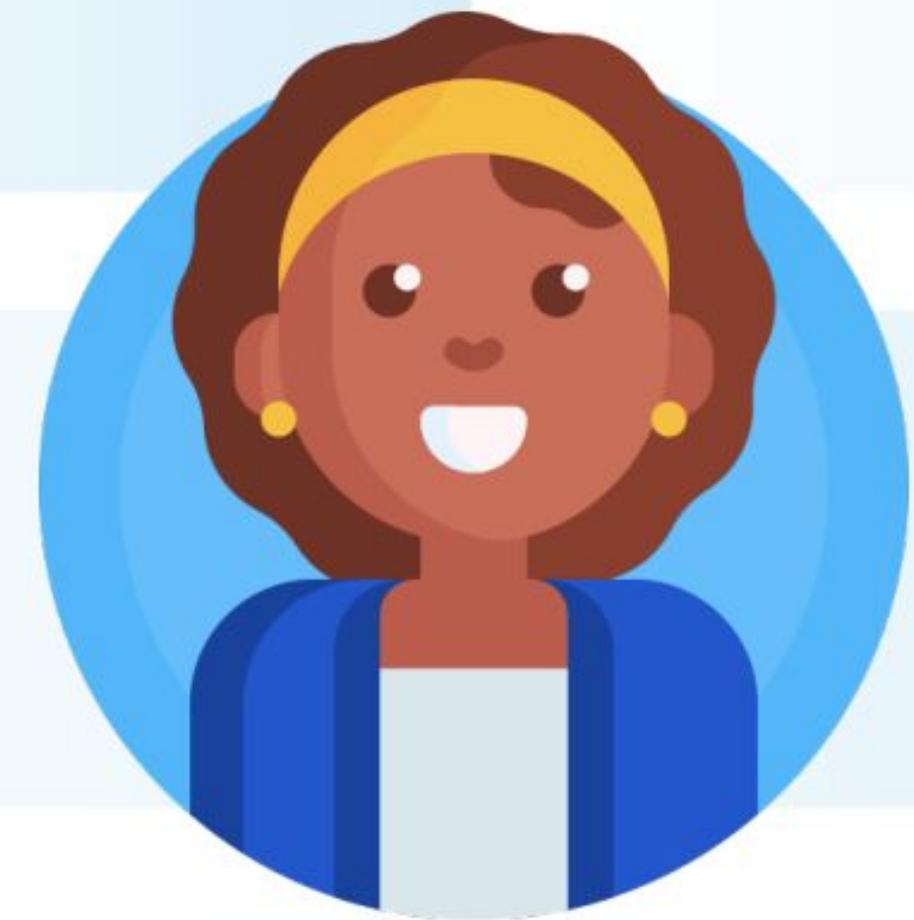
# Interacción 1



# Interacción 2



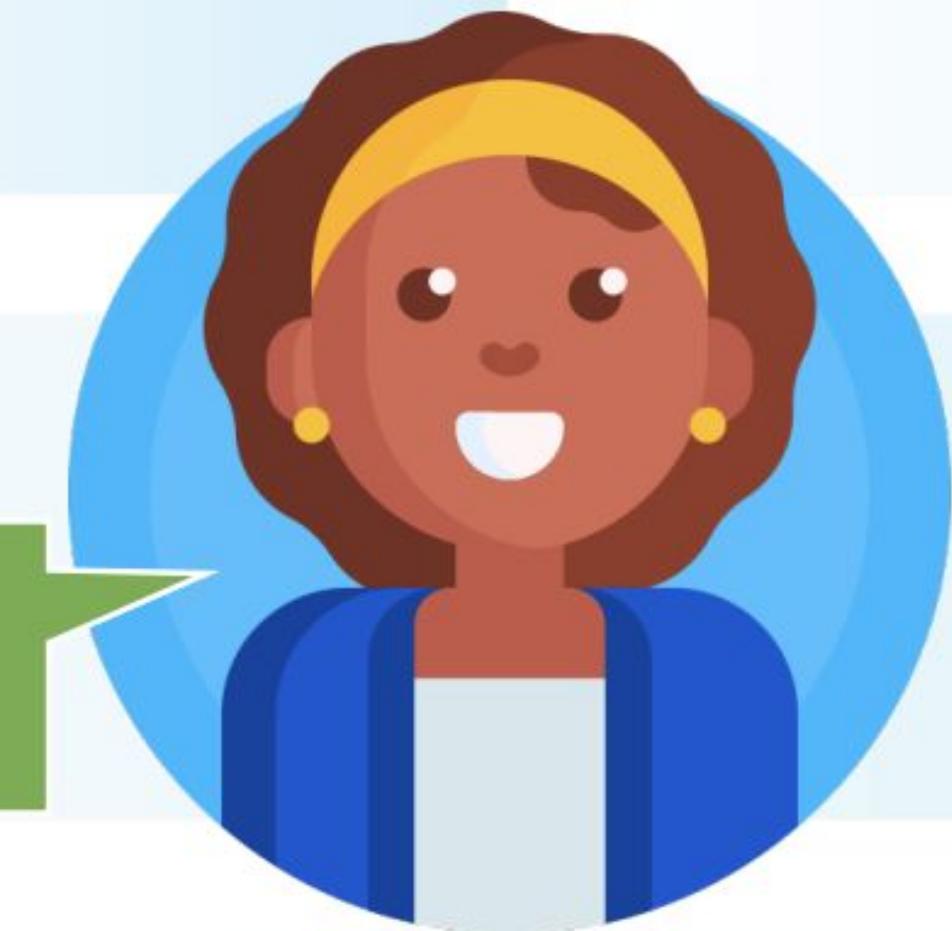
Quiero un modelo de machine learning que me diga si un cliente del hotel no se va a presentar para la reserva de habitación que hizo.



# Interacción 2



Quiero un modelo de machine learning que me diga si un cliente del hotel no se va a presentar para la reserva de habitación que hizo.

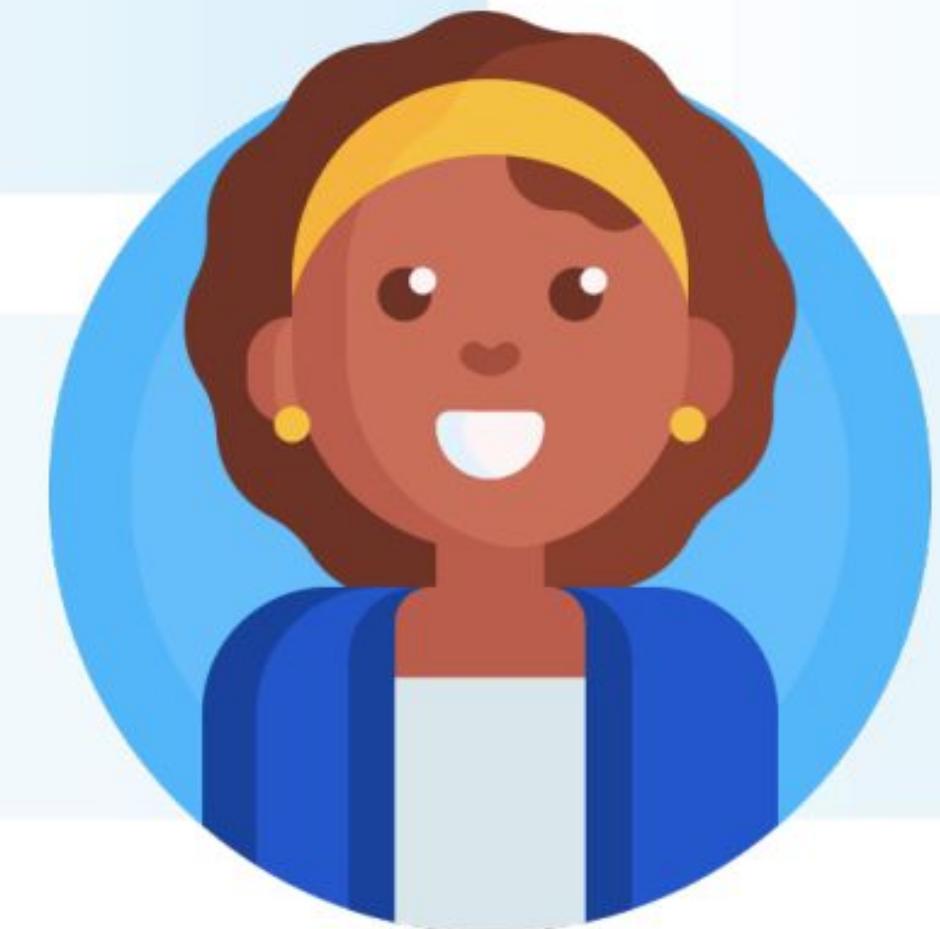


¡Okay! pero, ¿qué problema quieres afrontar con este modelo?

# Interacción 2



Queremos reducir la cantidad de habitaciones vacías; si podemos detectar con antelación a un cliente que va a cancelar podemos darle seguimiento al cliente y ofrecer cancelación gratuita y volver a poner la habitación en el mercado para no perder la tarifa completa.

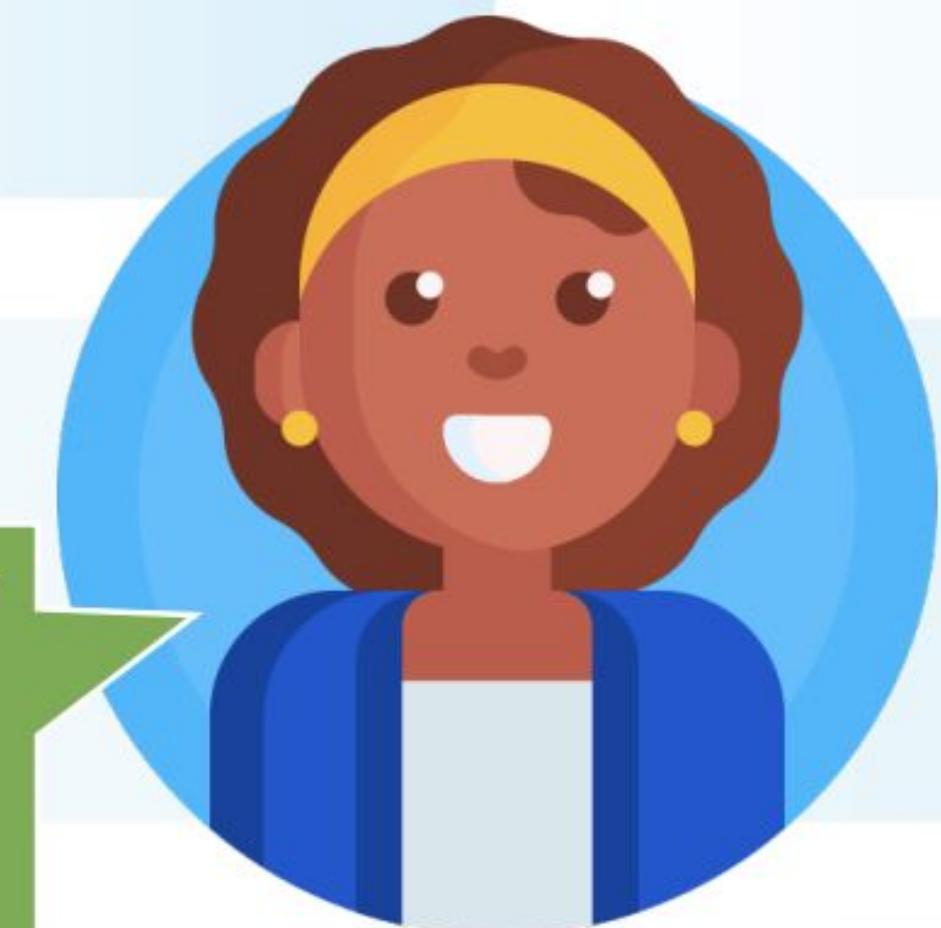


# Interacción 2



Queremos reducir la cantidad de habitaciones vacías; si podemos detectar con antelación a un cliente que va a cancelar podemos darle seguimiento al cliente y ofrecer cancelación gratuita y volver a poner la habitación en el mercado para no perder la tarifa completa.

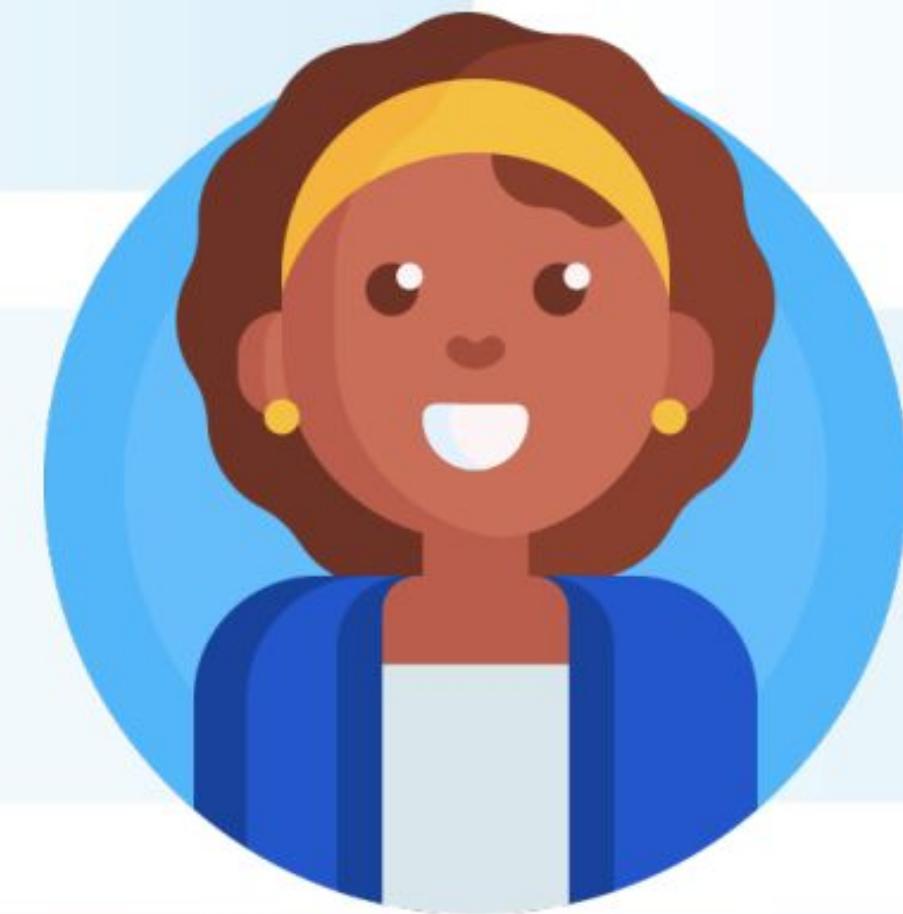
Estaríamos utilizando el historial del cliente con nosotros, además de la información de la reserva a analizar y además de la información de tendencias en redes sociales que nos llega de Twitter. ¿Hay alguna otra información que podríamos consultar?



# Interacción 2



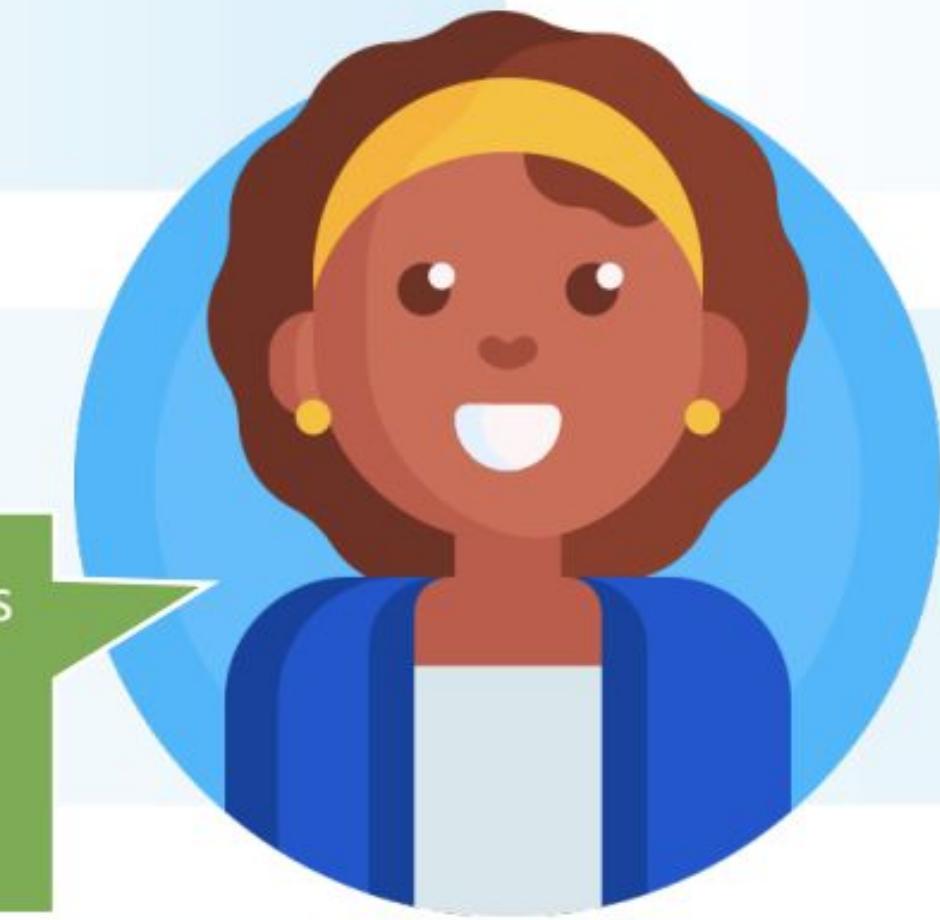
Ehm, tenemos el acuerdo de datos con TripAdvisor, nos da información sobre qué atracciones son más populares



# Interacción 2



Ehm, tenemos el acuerdo de datos con TripAdvisor, nos da información sobre qué atracciones son más populares

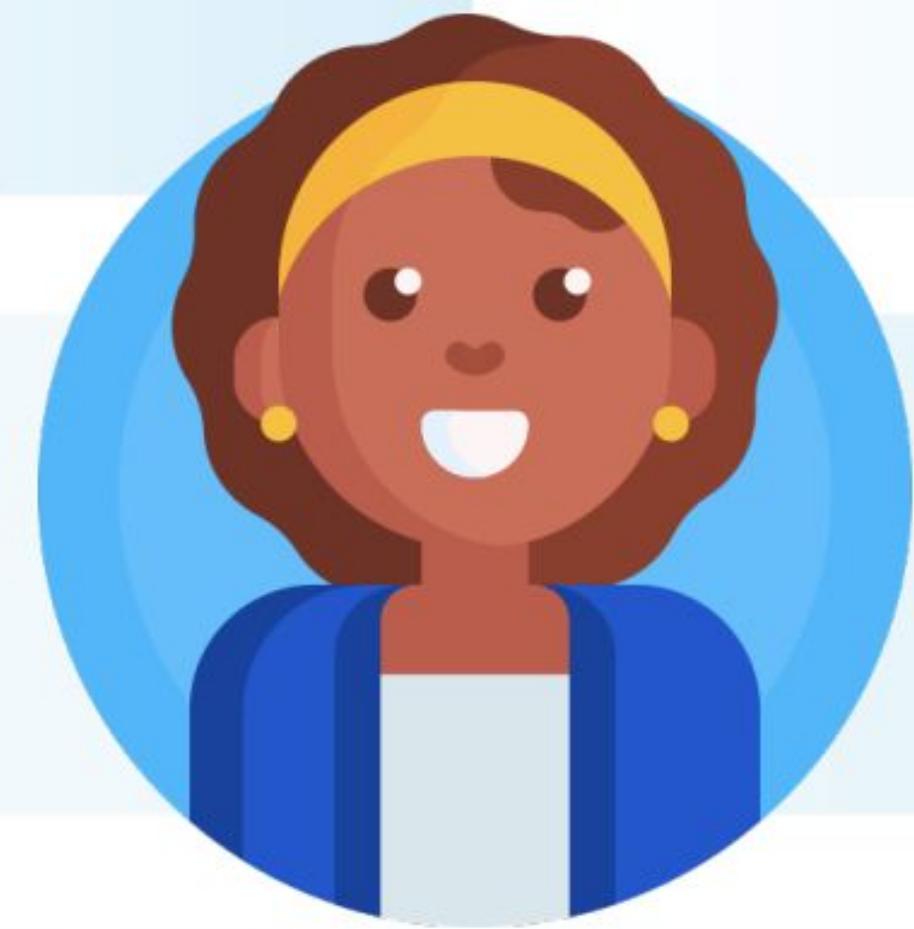


Perfecto, lo consideramos. Ahora, supongamos que se pueda detectar clientes con riesgo de cancelación, ¿cómo se efectuarían las llamadas?

# Interacción 2



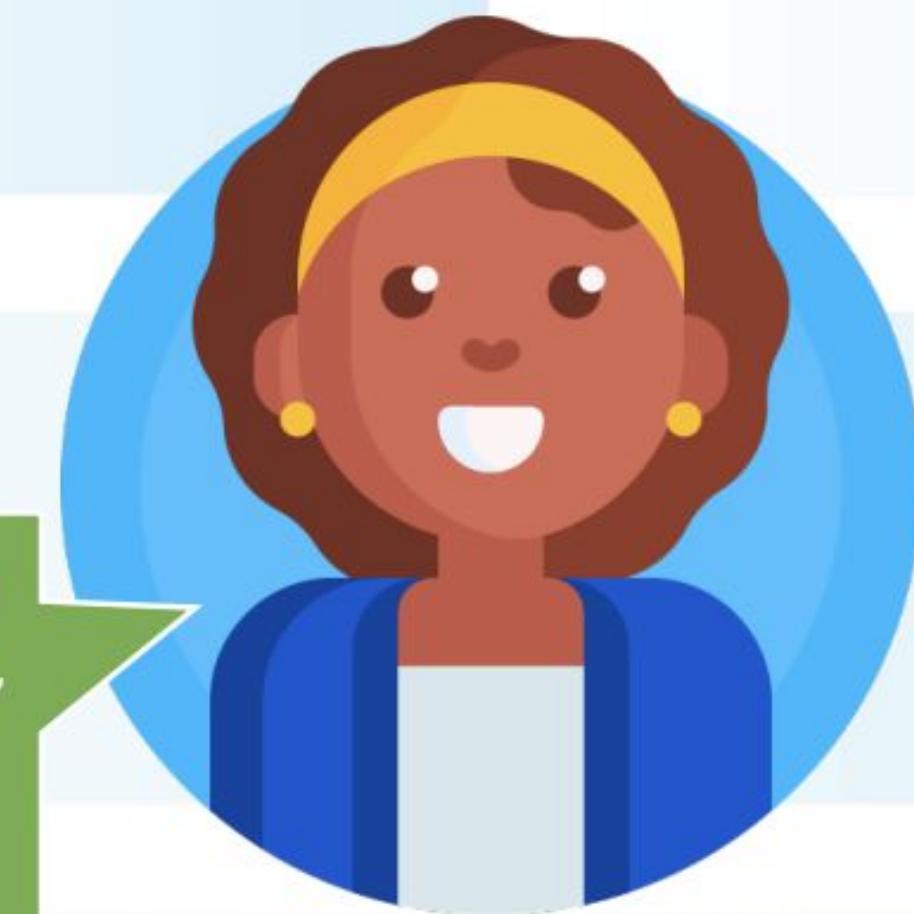
Tenemos al personal de servicio a clientes, queremos que una vez al día puedan ver qué clientes están en riesgo de cancelación y les llamen para confirmar la reserva.



# Interacción 2



Tenemos al personal de servicio a clientes, queremos que una vez al día puedan ver qué clientes están en riesgo de cancelación y les llamen para confirmar la reserva.

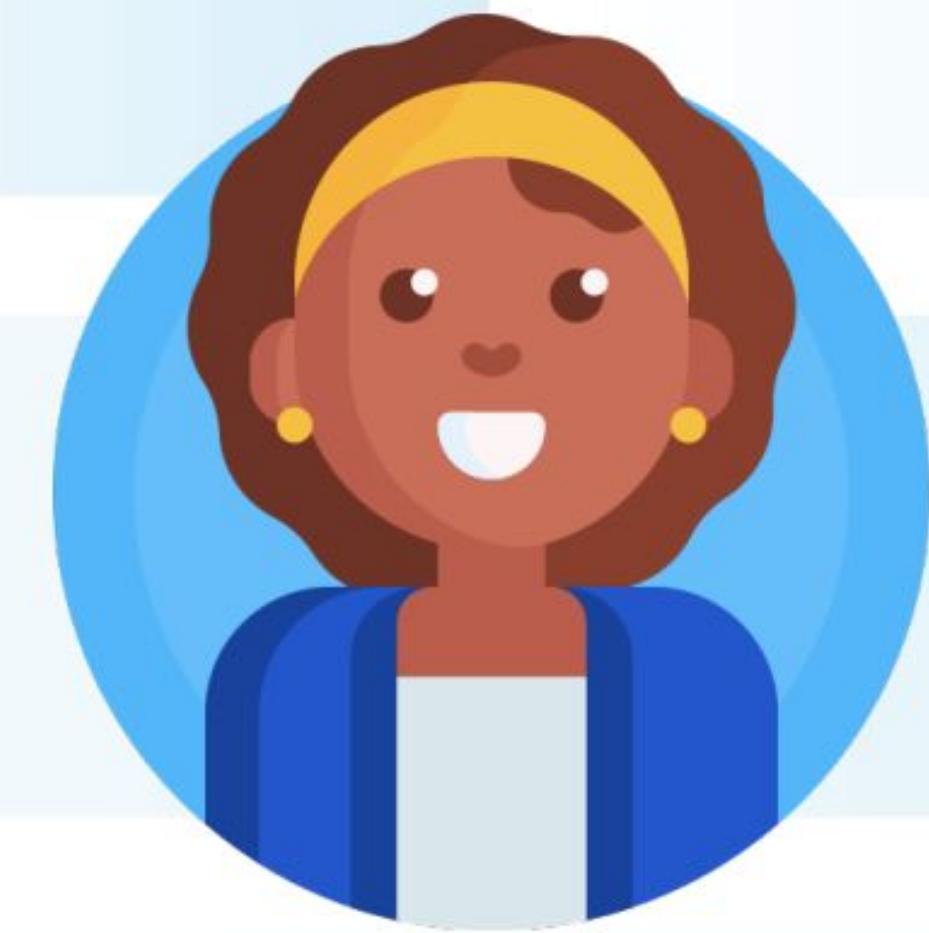


Estuvimos trabajando hace poco en un modelo de priorización de llamadas para el *call center*, ¿deberíamos implementar algo similar para la predicción de cancelaciones? ¿en lugar de decirte quién va a cancelar o no, darte una especie de porcentaje de riesgo?

# Interacción 2



Eso estaría perfecto, así podemos llamarle  
primero a aquellos con mayor riesgo.

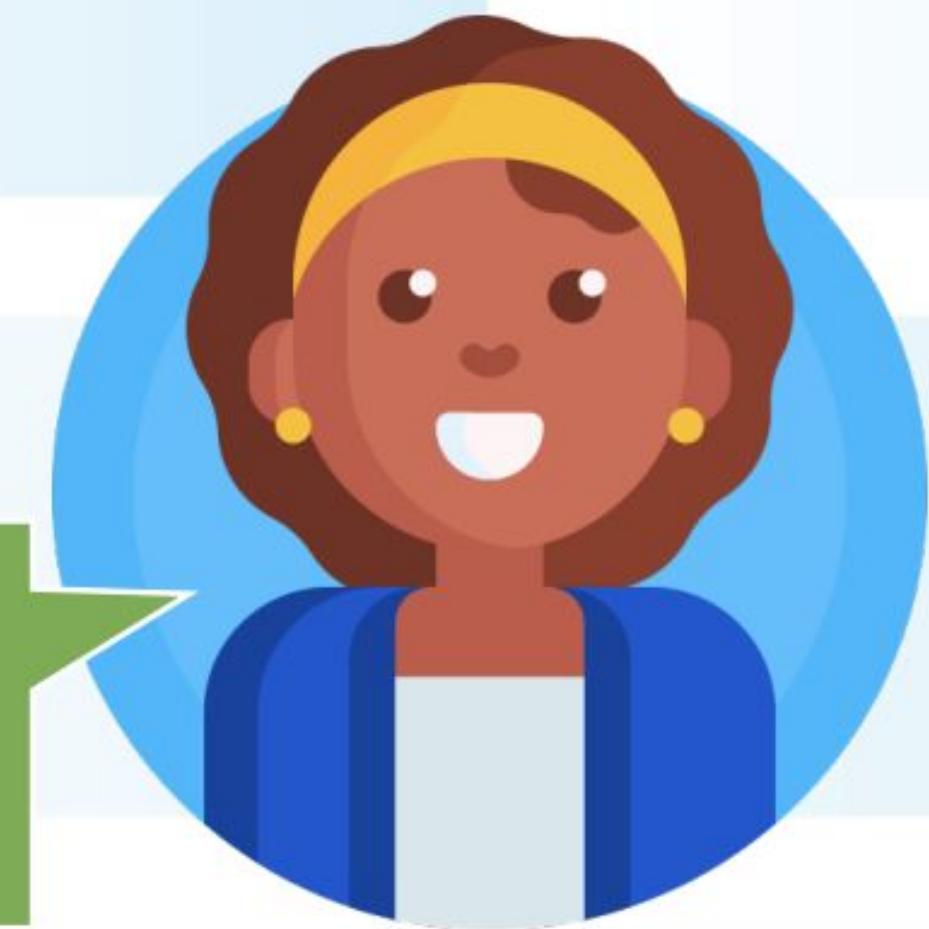


# Interacción 2



Eso estaría perfecto, así podemos llamarle primero a aquellos con mayor riesgo.

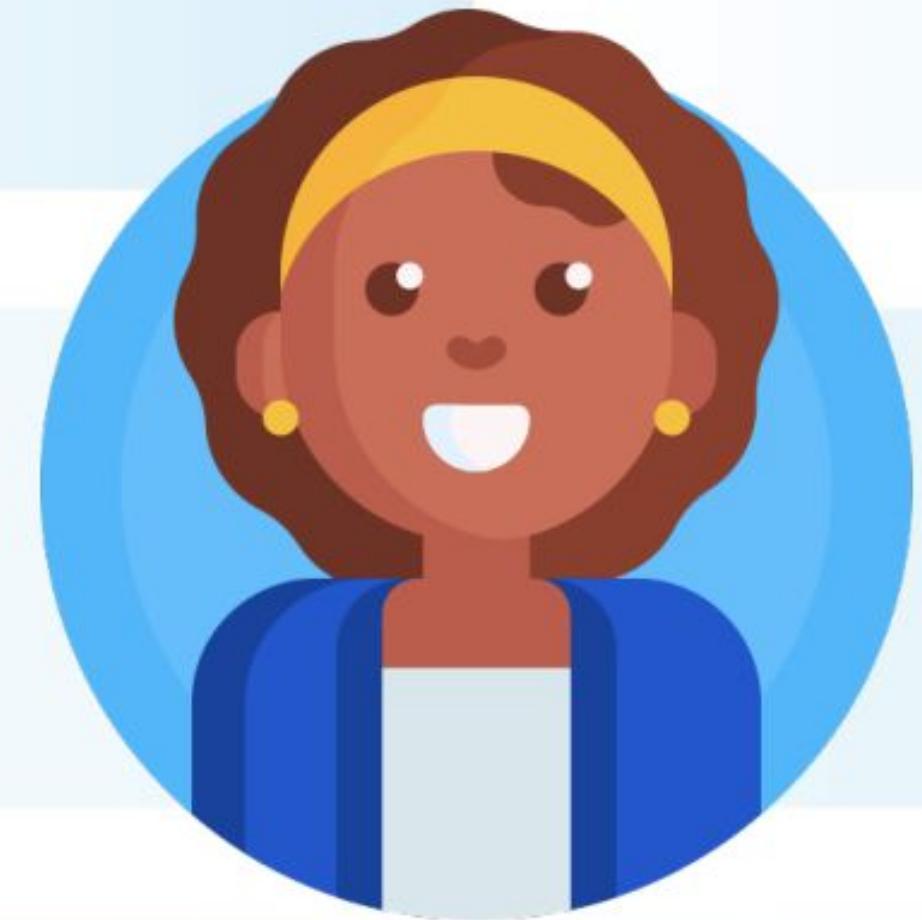
Como las predicciones son una vez al día podríamos efectuar el proceso por la noche para que cuando los agentes lleguen por la mañana tengan ya las calificaciones.



# Interacción 2



¿Puede tu sistema conectarse con el *call center* directamente?



# Interacción 2



¿Puede tu sistema conectarse con el *call center* directamente?

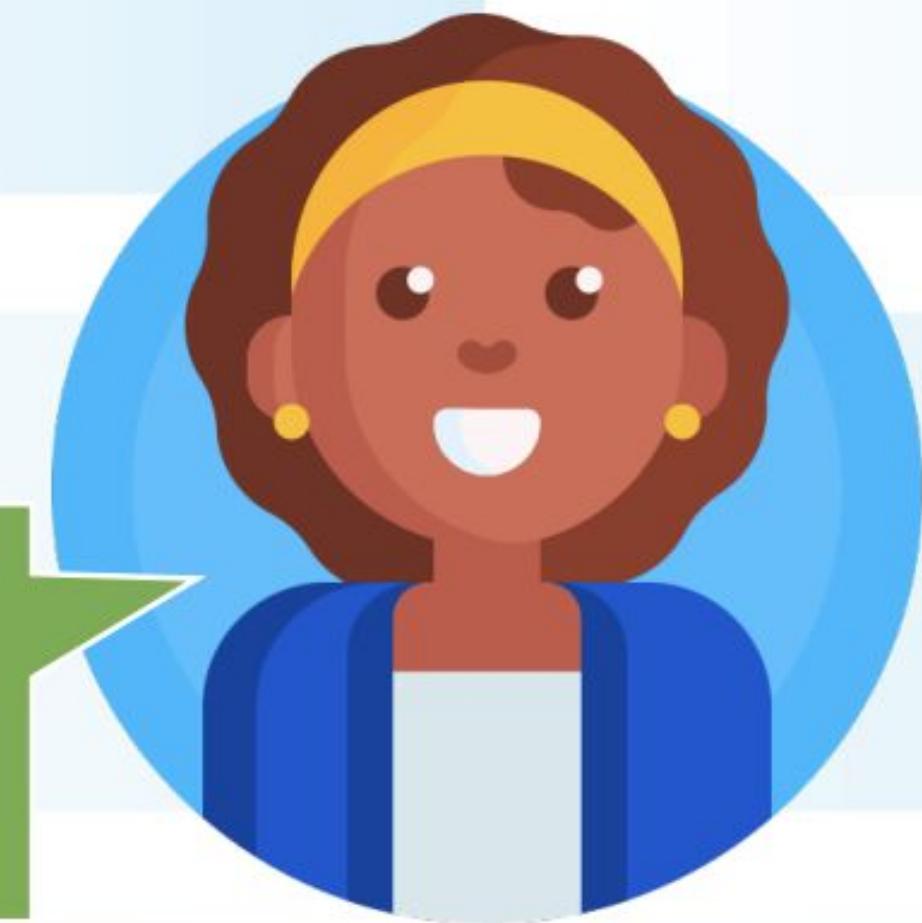
No directamente, pero hemos trabajado ya con el *call center*, sabemos que su sistema está hecho en Java y no tendríamos problema en hacernos cargo de dejar las calificaciones listas para que ellos las consuman. Si confirmemos la viabilidad del modelo nos pondremos en contacto con el equipo para coordinar la entrega, desde luego, estarías presente para las pláticas iniciales.



# Interacción 2



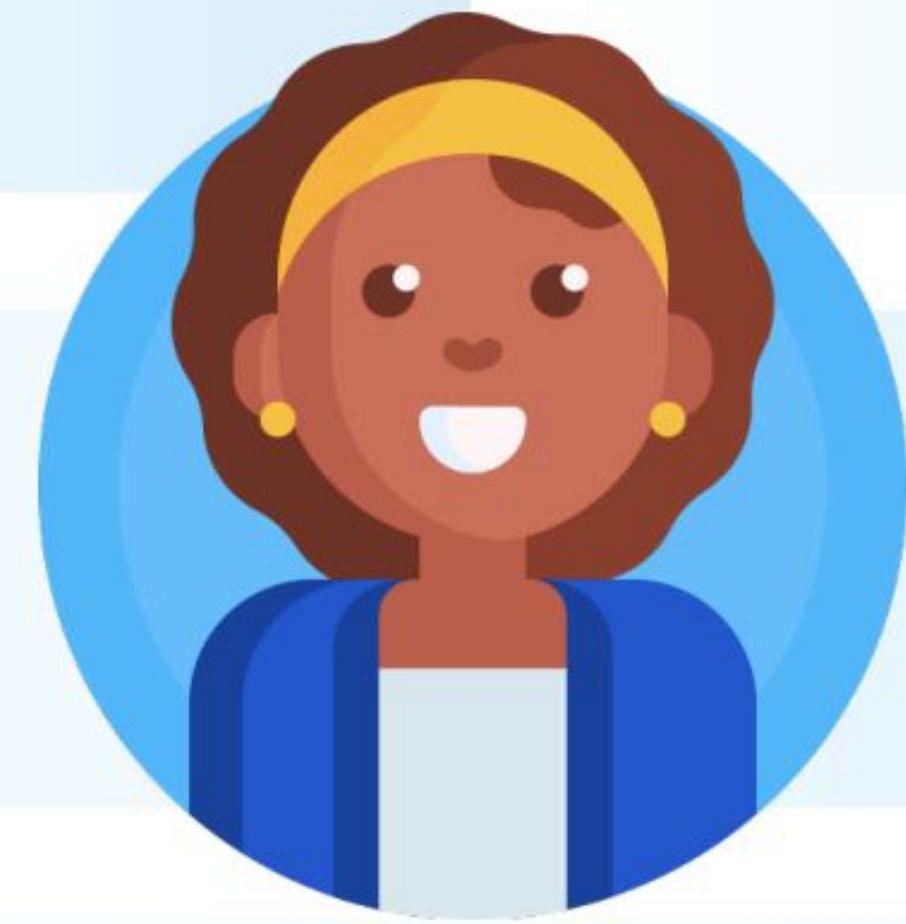
Por el momento lo que podemos hacer es comenzar el análisis y presentarte un avance en un par de semanas para estés al tanto de la viabilidad del modelo.



# Interacción 2



Muchas gracias.

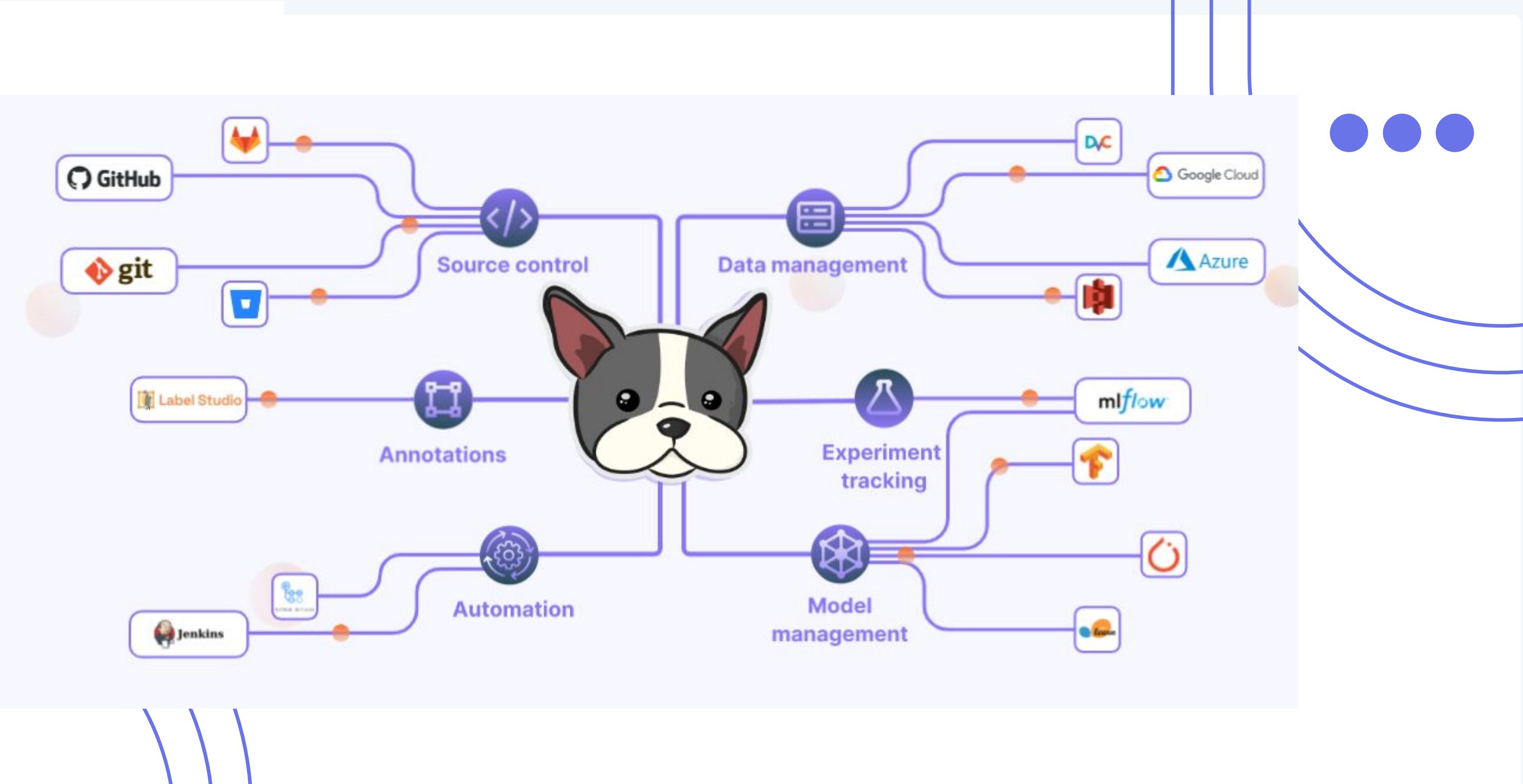




# DagsHub

...

**Focus on what's important while we do the MLOps heavy lifting for you. With DagsHub you can host your code, data, experiments and models, manage projects and collaborate with your teammates – all in one place.**





Muchas gracias!  
¿Alguna duda?

The science is for everyone !!!



# Recursos de interés

- Effective MLOps: Model Development
- CI/CD for Machine Learning (GitOps).
- A curated list of MLOps projects

<https://www.wandb.courses/pages/w-b-courses>

<https://mlops.toys/#projects>