

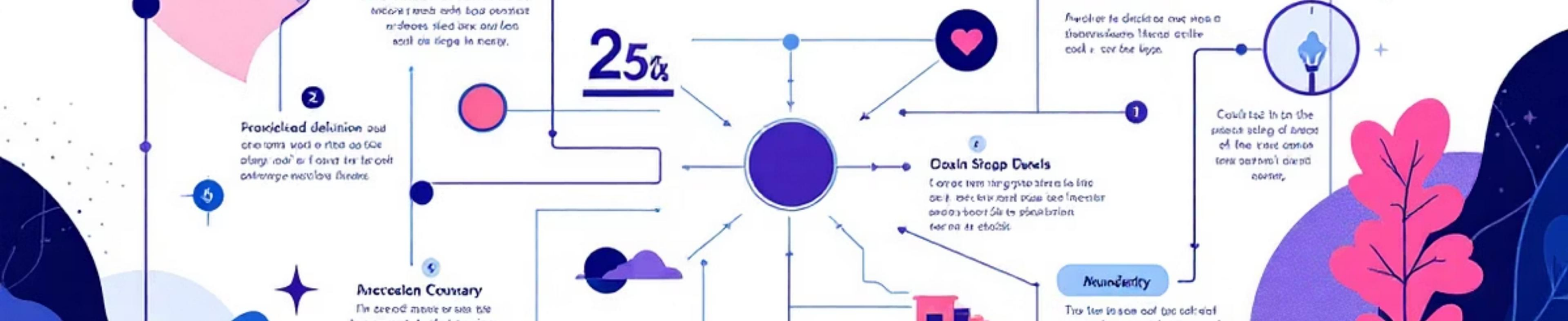
# Definición del Alcance y Planificación de Proyectos de de Machine Learning y NLP



# El ciclo completo



Cada fase es clave para el éxito y la entrega de valor, formando un ciclo integral.



## ¿Por qué definir el alcance final es clave?

### Evita desviaciones

Previene el malgasto de recursos valiosos en proyectos ML/NLP que pueden desviarse fácilmente sin límites claros

### Límites claros

Establece fronteras precisas para entregables y criterios de éxito, facilitando la toma de decisiones

### Comunicación efectiva

Mejora la comunicación entre stakeholders técnicos y de negocio, alineando expectativas desde el inicio

# Metodología Design Thinking aplicada al alcance

01	02	03
<b>Empatizar</b>	<b>Definir</b>	<b>Idear</b>
Comprender las necesidades reales del usuario final y los objetivos estratégicos del negocio a través de entrevistas y observación	Delimitar problemas específicos y establecer objetivos concretos, medibles y alcanzables para el proyecto de ML/NLP	Generar múltiples soluciones creativas y priorizar funcionalidades que aporten mayor valor al usuario
04	05	
<b>Prototipar</b>	<b>Testear</b>	
Crear versiones simplificadas del modelo para validar hipótesis rápidamente antes de invertir en desarrollo completo	Validar prototipos con usuarios reales y ajustar el alcance basándose en feedback concreto y medible	

Los artefactos clave incluyen mapas de empatía, customer journey maps y prototipos de baja fidelidad que guían decisiones informadas



# Backlog del Producto



## Crear lista priorizada

Desarrollar funcionalidades y tareas específicas relacionadas con ML/NLP, desde preparación de datos hasta deployment del modelo.



## Desglosar en historias

Convertir requisitos en historias de usuario claras, medibles y testeable con criterios de aceptación específicos.



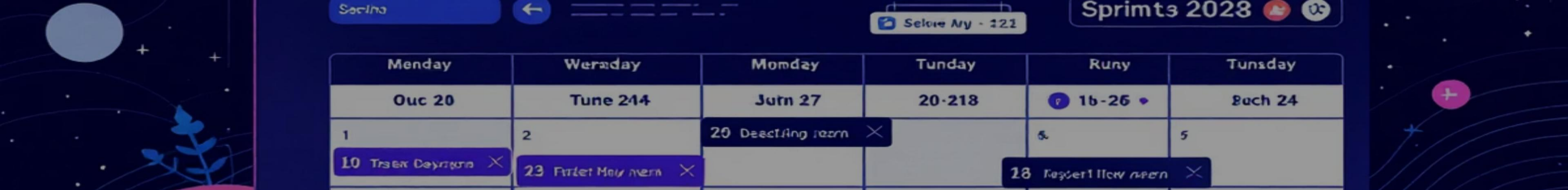
## Refinar continuamente

Mantener el backlog actualizado con feedback de usuarios, métricas del modelo y cambios en requisitos del negocio.



**Resultado:** Backlog vivo que guía el desarrollo iterativo y permite adaptación rápida a cambios en el proyecto de ML/NLP.





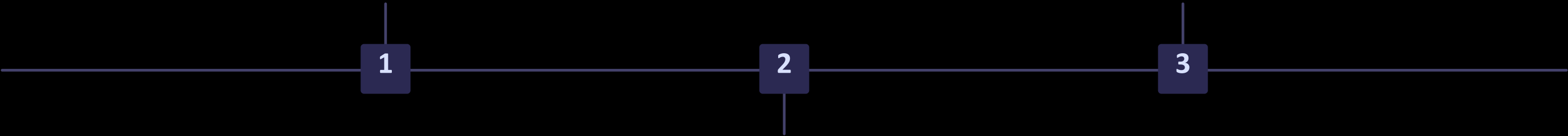
# Cronograma del Sprint

## Estimar capacidad

Evaluar capacidad del equipo considerando habilidades específicas en ML/NLP y duración del sprint (típicamente 1-2 semanas).

## Definir fechas clave

Establecer inicio del sprint, daily standups, sprint review, demo del modelo y retrospectiva del equipo.



## Seleccionar ítems

Elegir elementos del backlog que se puedan completar realísticamente en el sprint, priorizando valor de negocio.

**Resultado:** Cronograma realista y flexible que facilita iteraciones rápidas y entrega de valor continuo en proyectos de inteligencia artificial.

# Presupuesto para el Sprint

## Recursos humanos

- Data scientists senior y junior
- ML engineers y DevOps
- Product owner y Scrum master

## Infraestructura técnica

- Instancias GPU para entrenamiento
- Almacenamiento de datos y modelos
- APIs y servicios cloud especializados

## Herramientas y licencias

- Plataformas ML (AWS SageMaker, Azure ML)
- APIs de NLP (OpenAI, Hugging Face)
- Herramientas de monitoreo y MLOps



**Resultado:** Presupuesto detallado y alineado con el alcance del sprint, incluyendo estimaciones realistas de costos por recurso y tiempo invertido.





# Plan de Validación y Prototipado



## Crear prototipo funcional

Desarrollar MVP del modelo ML o interfaz NLP para validar hipótesis principales con usuarios reales rápidamente.



## Planificar pruebas

Diseñar experimentos controlados con usuarios o stakeholders para validar efectividad y usabilidad del modelo.



## Recoger feedback

Capturar insights cualitativos y cuantitativos para iterar y ajustar el producto según necesidades reales.

**Resultado:** Validación temprana que reduce significativamente riesgos técnicos y de producto, mejorando las probabilidades de éxito del proyecto de ML/NLP.



# Gestión de Riesgos y Contingencias

## Riesgos técnicos

- Datos insuficientes o de baja calidad
- Sesgos en algoritmos de ML
- Problemas de escalabilidad del modelo
- Overfitting o underfitting

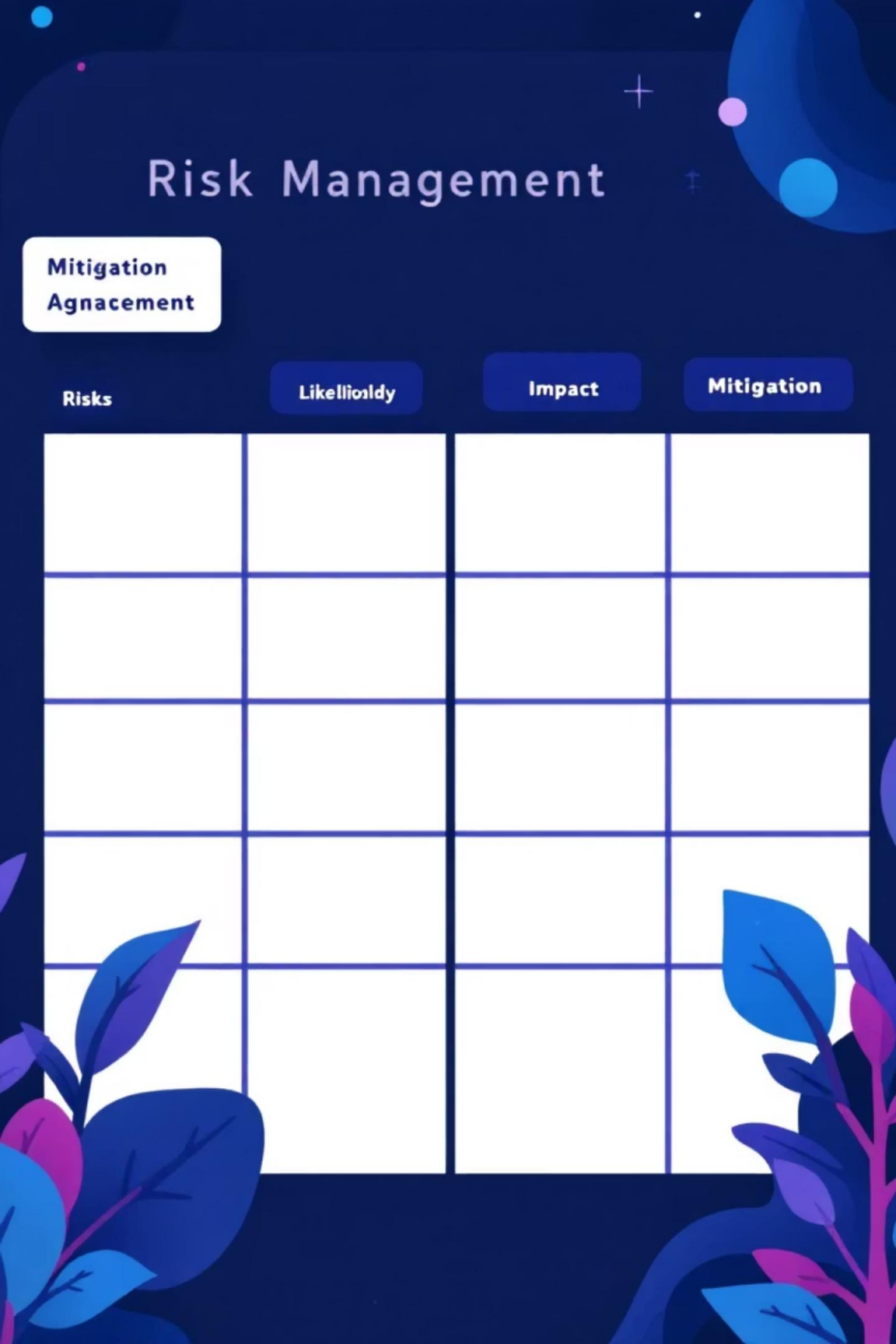
## Plan de mitigación

- Estrategias de data augmentation
- Técnicas de fairness en ML
- Arquitecturas escalables y modularidad
- Validación cruzada rigurosa

## Riesgos de negocio

- Cambios en requisitos del usuario
- Limitaciones presupuestarias
- Competencia en el mercado
- Regulaciones de privacidad de datos

**Resultado:** Mayor resiliencia del proyecto con capacidad de adaptación rápida ante desafíos técnicos y cambios del entorno.



# Etapas

- **Etapa 1: Definición del Alcance y Objetivos**
- **Etapa 2: Recolección y Preparación de Datos**
- **Etapa 3: Exploración y Análisis de Datos (EDA)**
- **Etapa 4: Ingeniería de Características y Selección**
- **Etapa 5: Entrenamiento, Evaluación y Selección del Modelo**
- **Etapa 6: Despliegue y Monitoreo en Producción**

# **Etapas 1: Definición del alcance y objetivos**

## **Identificación del problema problema**

Por ejemplo, la detección automática de anomalías en sistemas de radar de entidades de armada.

## **Objetivos claros**

Definición de objetivos estratégicos y métricas de éxito como precisión y tiempo de respuesta.

## **Alineación con Stakeholders**

Colaboración con personal militar y técnicos para establecer expectativas y asegurar recursos adecuados.

Esta etapa es fundamental para sentar las bases del proyecto, garantizando que el esfuerzo de ML se alinee con las necesidades operativas y estratégicas, optimizando la asignación de recursos y mitigando riesgos desde el inicio.



# Etapa 2: Recolección y Preparación de Datos

01

## Obtención de datos

Recopilación de datos históricos de sensores, radares y registros operativos, asegurando la diversidad y volumen necesarios.

02

## Limpieza y normalización

Proceso crítico para eliminar ruido, inconsistencias y datos faltantes, preparando el conjunto para el modelado.

03

## Ejemplo práctico

Filtrado de señales erróneas y etiquetado preciso de eventos relevantes (como firmas de embarcaciones) para el entrenamiento del modelo de detección de anomalías.



# Etapa 3: Exploración y Análisis de Datos (EDA)



- **Análisis estadístico y visual**

Profundización en los datos para descubrir patrones ocultos, correlaciones y anomalías iniciales.

- **Identificación de características clave**

Determinación de las variables más influyentes (como la frecuencia o intensidad de señales de radar) que impactarán directamente el rendimiento del modelo.

- **Detección de tendencias**

Uso de herramientas de visualización avanzadas para identificar tendencias temporales, distribuciones y posibles sesgos en los datos del radar.

El EDA permite a los equipos comprender a fondo el conjunto de datos, informando decisiones críticas sobre la ingeniería de características y la selección del modelo, lo que conduce a resultados más precisos y significativos.



# Etapa 4: Ingeniería de características y selección



## Creación de nuevas variables

Desarrollo de variables derivadas que capturen información relevante, como la tasa de cambio de la señal de un radar a lo largo del tiempo o la interacción entre diferentes sensores.



## Selección de características

Identificación de las características más relevantes y predictivas, eliminando las redundantes o menos informativas para mejorar el rendimiento y la eficiencia del modelo.



## Análisis de importancia de variables

Aplicación de técnicas avanzadas para priorizar señales críticas, como la contribución de la frecuencia de pulso o la potencia de eco en la detección de objetivos.

La ingeniería de características es un arte y una ciencia que transforma datos brutos en información de alto valor, permitiendo que los modelos de ML capturen mejor la complejidad subyacente del problema.



## Etapas 5: Entrenamiento, evaluación y selección del modelo



- **Entrenamiento con algoritmos avanzados:** Implementación de modelos como redes neuronales convolucionales para el procesamiento de señales de radar o árboles de decisión para la clasificación de amenazas.
- **Evaluación con métricas específicas:** Medición del rendimiento utilizando métricas críticas como el **recall** (sensibilidad) para asegurar la detección de todas las posibles amenazas, o la **precisión** para minimizar falsas alarmas.
- **Optimización iterativa:** Ajuste continuo de hiperparámetros y arquitectura del modelo para evitar el sobreajuste y mejorar la generalización en nuevos datos. Se realizan múltiples iteraciones para encontrar el equilibrio perfecto.

# Etapa 6: Despliegue y Monitoreo en Producción



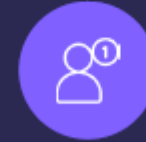
## Integración del Modelo

Implementación fluida del modelo de detección de anomalías directamente en los sistemas operativos y la infraestructura de la armada, garantizando compatibilidad y rendimiento en tiempo real.



## Monitoreo Continuo

Establecimiento de un sistema de vigilancia constante para detectar cualquier degradación en el rendimiento del modelo debido a cambios en los datos o el entorno operativo, como nuevas firmas de radar.



## Sistema de Alerta en Tiempo Real

Configuración de alertas automáticas para notificar a los operadores sobre posibles amenazas detectadas, permitiendo una respuesta inmediata y coordinada.

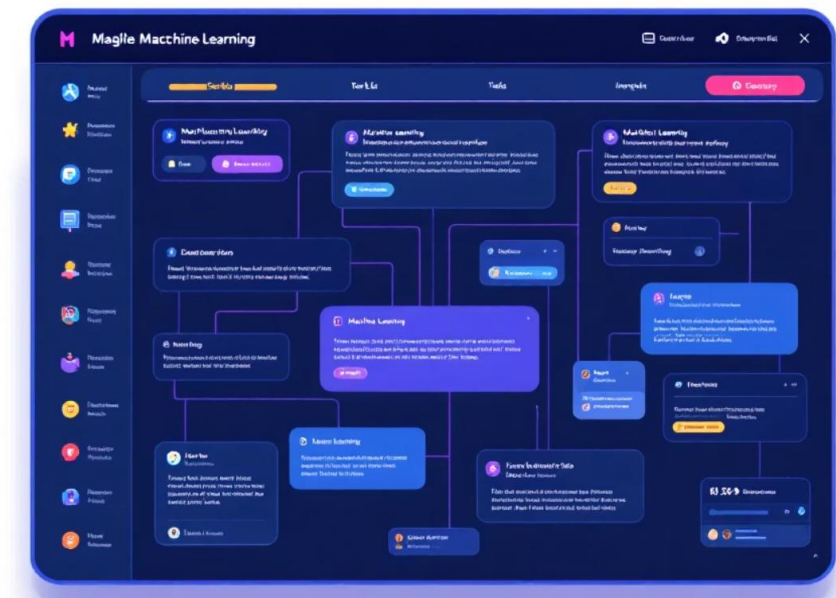
El despliegue no es el final, sino el inicio de una fase crítica donde el modelo demuestra su valor. El monitoreo proactivo asegura su vigencia y eficacia a largo plazo en un entorno dinámico.



# Estructuración del Plan de Trabajo: Sprint Plan para ML y NLP

## Características del Sprint en ML/NLP

- **Duración:** Ciclos cortos de 1 semana con objetivos claros y entregables específicos
- **Fases iterativas:** Recolección de datos, limpieza, ingeniería de features, modelado, evaluación y despliegue
- **Herramientas:** Tableros Kanban o metodología Scrum para seguimiento continuo y adaptación ágil
- **Flexibilidad:** Capacidad de pivotar basada en resultados de cada sprint



📄 **Sprint Ejemplo:** Sprint 1 - Dataset curado y análisis exploratorio completo; Sprint 2 - Desarrollo de modelos base y evaluación inicial de rendimiento





## Criterios de éxito y entregables en proyectos ML/NLP

1

### Métricas cuantificables

Definir indicadores precisos como accuracy, precision, recall, F1-score, AUC-ROC para clasificación, y RMSE, MAE para regresión

2

### Entregables tangibles

Datasets limpios y documentados, modelos entrenados y versionados, reportes de resultados detallados, código reproducible con documentación

3

### Validación con usuarios

Pruebas con stakeholders y usuarios finales para asegurar que el modelo genera valor real y cumple expectativas de negocio

4

### Documentación técnica

Crear documentación clara y completa que facilite el mantenimiento, actualizaciones futuras y escalabilidad del proyecto

# Artefactos Design Thinking en el sprint plan



## Mapa de empatía

Herramienta visual para comprender profundamente a los usuarios finales y sus necesidades específicas de datos y funcionalidades



## Storyboards

Visualización del flujo completo de trabajo que identifica puntos críticos, cuellos de botella y oportunidades de optimización



## Prototipos simples

Modelos de baja complejidad que permiten validar hipótesis clave rápidamente antes de invertir en desarrollo completo



## Loops de feedback

Ciclos integrados de retroalimentación que permiten iterar, aprender y mejorar continuamente el plan de trabajo



# Recomendaciones para un plan de trabajo exitoso

## 1 Definición clara del problema

Establecer desde el inicio objetivos específicos, medibles y alineados con la estrategia de negocio para evitar desviaciones costosas

## 2 Calidad sobre cantidad

Priorizar la limpieza, consistencia y representatividad de los datos por encima del volumen, ya que modelos requieren datos de alta calidad

## 3 Iteración rápida

Implementar ciclos cortos de prototipado y validación frecuente que permitan aprender rápido y ajustar la dirección según sea necesario

## 4 Documentación continua

Mantener documentación detallada de cada etapa para facilitar colaboración en equipo, transferencia de conocimiento y escalabilidad futura





# Plan estructurado

## Reduce riesgos

La claridad en alcance y planificación estratégica minimiza riesgos técnicos y mejora significativamente los resultados finales

## Potencia alineación

Integrar Design Thinking fortalece la alineación entre necesidades reales de usuarios y objetivos estratégicos de negocio

## Acelera el éxito

La selección cuidadosa de datos y metodología sprint estructurada acelera el camino hacia resultados de valor

