



# Sistema de Análisis de la Opinión Pública en la Red Social X

Autor:

Federico Emilio Otero

Director:

Título y Nombre del director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos  
entre el 26 de agosto de 2025 y el 14 de octubre de 2025.*

## Índice

Descripción técnica-conceptual del proyecto a realizar . . . . .	5
2. Identificación y análisis de los interesados . . . . .	7
3. Propósito del proyecto . . . . .	7
4. Alcance del proyecto . . . . .	7
5. Supuestos del proyecto. . . . .	8
6. Product Backlog . . . . .	8
7. Criterios de aceptación de historias de usuario . . . . .	9
8. Fases de CRISP-DM . . . . .	10
9. Desglose del trabajo en tareas . . . . .	10
10. Planificación de Sprints . . . . .	11
11. Diagrama de Gantt (sprints) . . . . .	13
12. Normativa y cumplimiento de datos (gobernanza) . . . . .	13
13. Gestión de riesgos . . . . .	14
14. Sprint Review . . . . .	15
15. Sprint Retrospective . . . . .	16

## Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	26 de agosto de 2025
1	Se completa hasta el punto 5 inclusive	9 de septiembre de 2025

## Acta de constitución del proyecto

Buenos Aires, 26 de agosto de 2025

Por medio de la presente se acuerda con el Federico Emilio Otero que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Sistema de Análisis de la Opinión Pública en la Red Social X” y consistirá en la implementación de un sistema basado en procesamiento de lenguaje natural (NLP) para analizar las publicaciones en la red social X y comprender la evolución de la opinión pública en términos emocionales y cognitivos. El trabajo tendrá un presupuesto preliminar estimado de 600 horas y un costo estimado de 15.000 USD, con fecha de inicio el 26 de agosto de 2025 y fecha de presentación pública en junio de 2026.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg  
Director posgrado FIUBA

Nombre del cliente  
Empresa del cliente

Título y Nombre del director  
Director del Trabajo Final

## 1. Descripción técnica-conceptual del proyecto a realizar

En contextos políticos dinámicos, el análisis de la opinión pública es necesario para entender el estado emocional y cognitivo de la ciudadanía, así como también sus principales demandas y su nivel de aprobación de las políticas implementadas. Las redes sociales en general, y X en particular, se han convertido en una plataforma clave para comprender estas cuestiones.

Este proyecto es parte de un emprendimiento personal donde se busca aplicar diferentes herramientas de la inteligencia artificial a las ciencias sociales. Consiste en implementar un sistema de análisis de la opinión pública en la red social X, que podrá complementarse con los estudios de opinión tradicionales para tener un panorama mucho más completo. A diferencia de las encuestas clásicas, que capturan opiniones de forma puntual y estructurada, el análisis de la opinión pública digital brindará una visión continua, no intrusiva y con alto volumen de datos.

Los clientes de este sistema tienen perfiles diversos, y van desde las consultoras de opinión pública, pasando por los espacios políticos hasta los científicos sociales, periodistas y la ciudadanía en general. Se considera que un sistema así será muy útil, ya que facilitará la toma de decisiones y la implementación de políticas públicas, el asesoramiento de diferentes espacios políticos y la investigación científica.

Este sistema constará de una interfaz visual donde podrá consultarse, para un determinado período de tiempo elegido por el usuario, métricas relacionadas al grado de aprobación o desaprobación del gobierno, las emociones involucradas, los sesgos ideológicos, los principales temas en la agenda y la postura frente a ellos. Se podrá visualizar también representaciones gráficas de comunidades ideológicas o cámaras de eco a través del análisis de retweets y menciones.

El sistema se desarrollará en un entorno de Python, utilizando herramientas modernas de NLP como spaCy, HuggingFace Transformers y modelos pre-entrenados para análisis de sentimiento en redes sociales, como BERT o RoBERTa, adaptados a X. Se utilizará una combinación de scraping con snsrape y la API de X para recolectar publicaciones relacionadas con términos clave, hashtags y cuentas oficiales. Los textos serán pre-procesados mediante técnicas de limpieza, normalización y lematización. Luego, se aplicarán distintos modelos, entre ellos:

- Modelos para desambiguar términos o expresiones y poder aclarar o traducir ciertos pasajes donde se usan ciertos recursos literarios como el sarcasmo.
- Modelos para entender los sesgos ideológicos y categorizarlos.
- Modelos de análisis de sentimiento para medir la aprobación/desaprobación del gobierno.
- Modelos de detección de emociones (*stance detection*).

En la figura 1 se muestra un diagrama simplificado de la arquitectura del sistema. Se observan cuatro etapas principales del tratamiento de los datos: la extracción o *scrapping*, el pre-procesamiento, la aplicación de distintos modelos y finalmente su almacenamiento y explotación. A su vez, estas etapas están atravesadas por estrategias de integración continua, con el objetivo de automatizar el flujo de los datos.

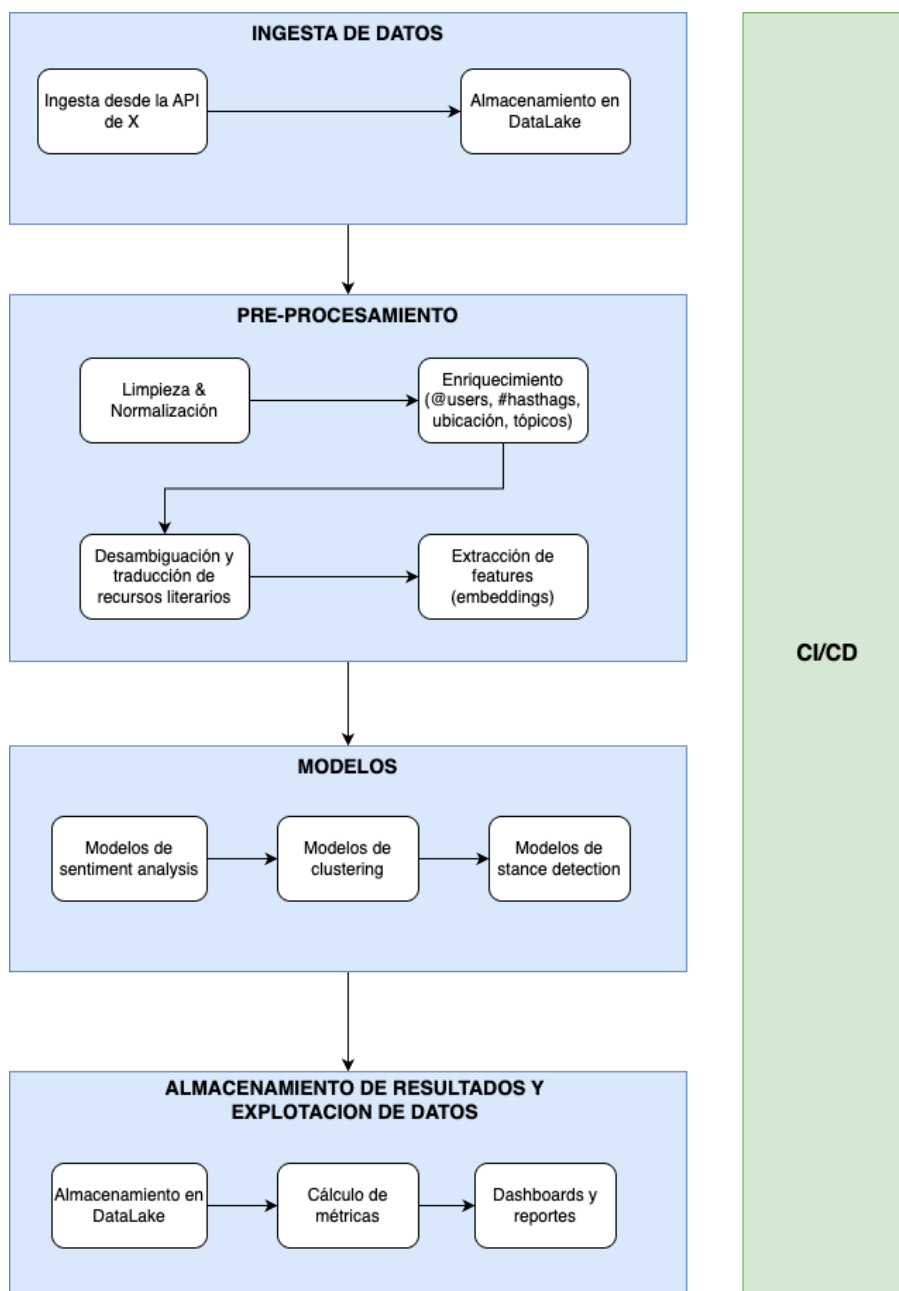


Figura 1. Diagrama de la arquitectura del sistema.

Actualmente, el sistema se encuentra en su fase inicial, estando en desarrollo el proceso de ingesta de los mensajes (*tuits*) desde la API de X.

## 2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Nombre del cliente	Empresa del cliente	-
Responsable	Federico Emilio Otero	FIUBA	Alumno
Orientador	Título y Nombre del director	pertenencia	Director del Trabajo Final
Usuario final	-	-	-

Las partes interesadas son las siguientes:

- Cliente: En este ítem escribiremos una breve descripción del perfil del cliente del proyecto.
- Orientador: En este ítem escribiremos una breve descripción del perfil del director del proyecto.
- Usuario final: En este ítem escribiremos una breve descripción del perfil del usuario final del proyecto.

## 3. Propósito del proyecto

El propósito del proyecto es lograr un análisis más profundo de la evolución de la opinión pública en redes sociales, en particular X, para poder comprender mejor el estado emocional y cognitivo de la ciudadanía. Elegimos esta red social ya que es ampliamente la más utilizada para el intercambio de opiniones políticas y sociales. Se considera que un sistema así será muy útil, ya que facilitará la toma de decisiones y la implementación de políticas públicas, el asesoramiento de diferentes espacios políticos y la investigación científica.

## 4. Alcance del proyecto

El proyecto incluye:

- Extracción automática de publicaciones.
- Pre-procesamiento y normalización.
- Análisis de sentimiento.
- Detección de sesgos ideológicos.
- Detección de emociones o posturas.
- Visualización de métricas o resultados.

El presente proyecto no incluye:

- Cobertura de otras redes sociales, como Facebook, Instagram o TikTok.

- Análisis multilingüe: Como limitamos el proyecto al análisis de la opinión pública local, el sistema se enfocará exclusivamente en publicaciones en español.
- Análisis de contenido multimedia: El sistema no analizará contenido multimedia como audios, videos o imágenes.
- Veracidad de los contenidos: Por el momento, el sistema no contará con modelos de *fact-checking* para verificar las afirmaciones, ya que el objetivo principal es analizar opiniones y emociones.

## 5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Las librerías de Python y los modelos de machine learning están disponibles y son herramientas estables y robustas.
- Se cuenta con un acceso a la API de X necesario para la recolección de los datos, tanto en cuanto a su contenido como en cuanto a su volumen, conforme a las políticas de uso de la plataforma.
- Se cuenta con los recursos computacionales necesarios para un adecuado entrenamiento de los modelos.
- Se cuenta con la existencia de personas con las competencias técnicas necesarias para desarrollar el sistema.
- Los temas políticos abordados en la red X son representativos, en alguna medida, de la opinión pública digital, aunque no de toda la población general.

## 6. Product Backlog

El Product Backlog debe organizarse en cuatro **épicas** fundamentales del proyecto. Cada épica debe contener al menos dos historias de usuario que describan funcionalidades clave.

El Product Backlog debe permitir interpretar cómo será el proyecto y su funcionalidad. Se deben indicar claramente las prioridades entre las historias de usuario y si hay alguna opcional.

Las historias de usuario deben ser breves, claras y medibles, expresando el rol, la necesidad y el propósito de cada funcionalidad. También deben tener una prioridad definida para facilitar la planificación de los sprints.

Cada historia de usuario debe incluir una ponderación en *Story Points*, un número entero que representa el tamaño relativo de la historia. El criterio para calcular los Story Points debe indicarse explícitamente.

Las historias deben seguir el formato: “*Como [rol], quiero [tal cosa] para [tal otra cosa]*”.

Las épicas deben estructurarse de la siguiente forma:

- **Épica 1**



- HU1
- HU2
- **Épica 2**
  - HU3
  - HU4
- **Épica 3**
  - HU5
  - HU6
- **Épica 4**
  - HU7
  - HU8

#### **Reglas para definir historias de usuario:**

- Ser concisas y claras.
- Expresarlas en términos cuantificables y medibles.
- No dejar margen para interpretaciones ambiguas.
- Indicar claramente su prioridad y si son opcionales.
- Considerar regulaciones y normas vigentes.

### **7. Criterios de aceptación de historias de usuario**

Los criterios de aceptación deben establecerse para cada historia de usuario, asegurando que se cumplan las condiciones necesarias para que la funcionalidad sea validada correctamente.

Cada historia debe tener criterios medibles, específicos y verificables. Deben permitir validar que se cumple con las necesidades del usuario.

Se estructuran de forma análoga a las épicas del backlog:

- **Épica 1**
  - Criterios de aceptación HU1
  - Criterios de aceptación HU2
- **Épica 2**
  - Criterios de aceptación HU3
  - Criterios de aceptación HU4
- **Épica 3**

- Criterios de aceptación HU5
- Criterios de aceptación HU6
- **Épica 4**
  - Criterios de aceptación HU7
  - Criterios de aceptación HU8

#### **Reglas para definir criterios de aceptación:**

- Medibles y verificables.
- Especificar cuándo una historia se considera completada.
- Incluir condiciones específicas.
- No ambiguos.
- Probables de testear funcional o técnicamente.
- Mínimo 3 criterios por HU.

### **8. Fases de CRISP-DM**

1. **Comprensión del negocio:** objetivo, valor agregado de IA, métricas de éxito.
2. **Comprensión de los datos:** tipo, origen, cantidad, calidad.
3. **Preparación de los datos:** características clave, transformaciones necesarias.
4. **Modelado:** tipo de problema, algoritmos posibles.
5. **Evaluación del modelo:** métricas de rendimiento.
6. **Despliegue del modelo (opcional):** tipo de despliegue y herramientas.

### **9. Desglose del trabajo en tareas**

A partir de cada Historia de Usuario (HU) definida en la sección 6, descomponer el trabajo en tareas técnicas concretas, medibles y acotadas en el tiempo.

- Seleccionar entre 2 y 3 tareas por cada historia de usuario.
- Cada tarea debe estar claramente formulada, ser técnica, accionable y con una estimación horaria entre 2 y 8 horas.
- Evitar tareas genéricas (como "desarrollar funcionalidad") o demasiado amplias.
- Si una tarea supera las 8 horas, debe dividirse en subtareas.
- Indicar la prioridad relativa de cada tarea (Alta, Media o Baja).

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU1	Tarea 1 HU1	6 h	Alta
HU1	Tarea 2 HU1	8 h	Alta
HU2	Tarea 1 HU2	5 h	Media
HU2	Tarea 2 HU2	6 h	Alta
...	...	...	...

#### Criterios para estimar tiempos:

- Considerar la complejidad técnica, el nivel de incertidumbre y la experiencia previa.
- Evitar subestimar el esfuerzo: estimar el tiempo realista que llevaría implementar, testear y documentar cada tarea.
- Basar la estimación en la experiencia propia o en referencias de tareas similares.

#### Sobre la prioridad:

- Asignar una prioridad relativa (Alta, Media o Baja) según la relevancia funcional de la tarea y su impacto en los entregables.
- Priorizar tareas que estén vinculadas a criterios de aceptación de las HU o que sean necesarias para desbloquear otras.
- Incluir tareas opcionales solo si están bien justificadas.

#### Recomendaciones generales:

- -Enfocarse en tareas que surgen directamente de las HU planteadas.
- No es necesario cubrir las 600 horas del proyecto en esta sección: el foco está en el desglose de funcionalidades clave.
- Este trabajo será la base para organizar algunos de los sprints y elaborar el cronograma del proyecto, por lo que debe ser claro y realista.

## 10. Planificación de Sprints

Organizar las tareas técnicas del proyecto en sprints de trabajo que permitan distribuir de forma equilibrada la carga horaria total, estimada en 600 horas.

#### Consigna:

- Completar una tabla que relacione sprints con HU y tareas técnicas correspondientes.
- Incluir estimación en horas para cada tarea.
- Indicar responsable y porcentaje de avance estimado o completado.
- Contemplar también tareas de planificación, documentación, redacción de memoria y preparación de defensa.

### Conceptos clave:

- Una épica es una unidad funcional amplia; una historia de usuario es una funcionalidad concreta; un sprint es una unidad de tiempo donde se ejecutan tareas.
- Las tareas son el nivel más desagregado: permiten estimar tiempos, asignar responsables y monitorear progreso.

### Duración sugerida:

- Para un proyecto de 600 h, se recomienda planificar entre 10 y 12 sprints de aproximadamente 2 semanas cada uno.
- Asignar entre 45 y 50 horas efectivas por sprint a tareas técnicas.
- Reservar 100 a 120 h para actividades no técnicas (planificación, escritura, reuniones, defensa).

### Importante:

- En proyectos individuales, el responsable suele ser el propio autor.
- Aun así, desagregar tareas facilita el seguimiento y mejora continua.

### Conversión opcional de Story Points a horas:

- 1 SP  $\approx$  2 h como referencia flexible.
- Tener en cuenta aproximaciones tipo Fibonacci.

Cuadro 1. Formato sugerido

Sprint	HU o fase	Tarea	Horas / SP	Responsable	% Completado
Sprint 0	Planificación	Definir alcance y cronograma	10 h	Alumno	100 %
Sprint 0	Planificación	Reunión con el tutor/cliente	5 h	Alumno	50 %
Sprint 0	Planificación	Ajuste de los entregables	6 h	Alumno	25 %
Sprint 1	HU1	Tarea 1 HU1	6 h / 3 SP	Alumno	0 %
Sprint 1	HU1	Tarea 2 HU1	10 h / 5 SP	Alumno	0 %
Sprint 2	HU2	Tarea 1 HU2	7 h / 5 SP	Alumno	0 %
...	...	...	...	...	...
Sprint 5	Escritura	Redacción memoria	50 h / 34 SP	Alumno	0 %
Sprint 6	Defensa	Preparación de la exposición	20 h / 13 SP	Alumno	0 %

### Recomendaciones:

- Verificar que la carga horaria por sprint sea equilibrada.
- Usar sprints de 1 a 3 semanas, acordes al cronograma general.
- Actualizar el % completado durante el seguimiento del proyecto.
- Considerar un sprint final exclusivo para pruebas, revisión y ajustes antes de la defensa.

## 11. Diagrama de Gantt (sprints)

Visualizar en un diagrama de Gantt la planificación temporal del proyecto, tomando como base los sprints definidos en la sección anterior. Debe contemplar todas las horas del proyecto.

Consigna:

- Elaborar un diagrama de Gantt que muestre la secuencia temporal de los sprints.
- Cada fila debe representar un sprint (con su número o nombre), y el eje horizontal debe indicar el tiempo (en semanas o fechas concretas).
- Las tareas técnicas derivadas de HU deben diferenciarse visualmente (por ejemplo, con un color distinto) de las tareas no técnicas (planificación, redacción, defensa).
- Incluir todas las tareas estimadas en cada sprint.

Recomendaciones para el Gantt:

- Podés usar herramientas gratuitas como TeamGantt, ClickUp, GanttProject, [Google Sheets], [Trello + Planyway], entre otras.
- Ordená los sprints de forma cronológica, comenzando con Sprint 0 (planificación) y finalizando con el sprint de defensa.
- Asegurate de reflejar la duración realista de cada sprint según tu disponibilidad y el cronograma general del posgrado.
- Incluí hitos importantes: reuniones, entregas parciales, defensa.

Incluir una imagen legible del diagrama de Gantt. Si es muy ancho, presentar primero la tabla y luego el gráfico de barras.

## 12. Normativa y cumplimiento de datos (gobernanza)

En esta sección se debe analizar si los datos utilizados en el proyecto están sujetos a normativas de protección de datos y privacidad, y en qué condiciones se pueden emplear.

**Aspectos a considerar:**

- Evaluar si los datos están regulados por normativas como GDPR, Ley 25.326 de Protección de Datos Personales en Argentina, HIPAA u otras según jurisdicción y temática.

- Determinar si el uso de los datos requiere consentimiento explícito de los usuarios involucrados.
- Indicar si existen restricciones legales, técnicas o contractuales sobre el uso, compartición o publicación de los datos.
- Aclarar si los datos provienen de fuentes licenciadas, de acceso público o bajo algún tipo de autorización especial.
- Analizar la viabilidad del proyecto desde el punto de vista legal y ético, considerando la gobernanza de los datos.

Este análisis es clave para garantizar el cumplimiento normativo y evitar conflictos legales durante el desarrollo y publicación del proyecto.

### 13. Gestión de riesgos

a) Identificación de los riesgos (al menos cinco) y estimación de sus consecuencias:

Riesgo 1: detallar el riesgo (riesgo es algo que si ocurre altera los planes previstos de forma negativa)

- Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).
- Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2:

- Severidad (S): X.  
Justificación...
- Ocurrencia (O): Y.  
Justificación...

Riesgo 3:

- Severidad (S): X.  
Justificación...
- Ocurrencia (O): Y.  
Justificación...

b) Tabla de gestión de riesgos: (El RPN se calcula como  $RPN=S \times O$ )

Criterio adoptado:

Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a...

Riesgo	S	O	RPN	S*	O*	RPN*

Nota: los valores marcados con (\*) en la tabla corresponden luego de haber aplicado la mitigación.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 1: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).  
Nueva asignación de S y O, con su respectiva justificación:

- Severidad (S\*): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).
- Probabilidad de ocurrencia (O\*): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

Riesgo 3: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

## 14. Sprint Review

La revisión de sprint (*Sprint Review*) es una práctica fundamental en metodologías ágiles. Consiste en revisar y evaluar lo que se ha completado al finalizar un sprint. En esta instancia, se presentan los avances y se verifica si las funcionalidades cumplen con los criterios de aceptación establecidos. También se identifican entregables parciales y se consideran ajustes si es necesario.

Aunque el proyecto aún se encuentre en etapa de planificación, esta sección permite proyectar cómo se evaluarán las funcionalidades más importantes del backlog. Esta mirada anticipada favorece la planificación enfocada en valor y permite reflexionar sobre posibles obstáculos.

**Objetivo:** anticipar cómo se evaluará el avance del proyecto a medida que se desarrollen las funcionalidades, utilizando como base al menos cuatro historias de usuario del *Product Backlog*.

Seleccionar al menos 4 HU del Product Backlog. Para cada una, completar la siguiente tabla de revisión proyectada:

**Formato sugerido:**

HU seleccionada	Tareas asociadas	Entregable esperado	¿Cómo sabrás que está cumplida?	Observaciones o riesgos
HU1	Tarea 1	Módulo funcional	Cumple criterios de aceptación definidos	Falta validar con el tutor
	Tarea 2			
HU3	Tarea 1	Reporte generado	Exportación disponible y clara	Requiere datos reales
	Tarea 2			
HU5	Tarea 1	Panel de gestión	Roles diferenciados operativos	Riesgo en integración
	Tarea 2			
HU7	Tarea 1	Informe trimestral	PDF con gráficos y evolución	Puede faltar tiempo para ajustes
	Tarea 2			

## 15. Sprint Retrospective

La retrospectiva de sprint es una práctica orientada a la mejora continua. Al finalizar un sprint, el equipo (o el alumno, si trabaja de forma individual) reflexiona sobre lo que funcionó bien, lo que puede mejorarse y qué acciones concretas pueden implementarse para trabajar mejor en el futuro.

Durante la cursada se propuso el uso de la **Estrella de la Retrospectiva**, que organiza la reflexión en torno a cinco ejes:

- ¿Qué hacer más?
- ¿Qué hacer menos?
- ¿Qué mantener?
- ¿Qué empezar a hacer?
- ¿Qué dejar de hacer?

Aun en una etapa temprana, esta herramienta permite que el alumno planifique su forma de trabajar, identifique anticipadamente posibles dificultades y diseñe estrategias de organización personal.

**Objetivo:** reflexionar sobre las condiciones iniciales del proyecto, identificando fortalezas, posibles dificultades y estrategias de mejora, incluso antes del inicio del desarrollo.

Completar la siguiente tabla tomando como referencia los cinco ejes de la Estrella de la Retrospectiva (*Starfish* o estrella de mar). Esta instancia te ayudará a definir buenas prácticas desde el inicio y prepararte para enfrentar el trabajo de forma organizada y flexible. Se deberá completar la tabla al menos para 3 sprints técnicos y 1 no técnico.

**Formato sugerido:**



<b>Sprint tipo y N°</b>	<b>¿Qué hacer más?</b>	<b>¿Qué hacer menos?</b>	<b>¿Qué mantener?</b>	<b>¿Qué empezar a hacer?</b>	<b>¿Qué dejar de hacer?</b>
Sprint técnico - 1	Validaciones continuas con el alumno	Cambios sin versión registrada	Pruebas con datos simulados	Documentar cambios propuestos	Ajustes sin análisis de impacto
Sprint técnico - 2	Verificar configuraciones en múltiples escenarios	Modificar parámetros sin guardar historial	Perfiles reutilizables	Usar logs para configuración	Repetir pruebas manuales innecesarias
Sprint técnico - 8	Comparar correlaciones con casos previos	Cambiar parámetros sin justificar	Revisión cruzada de métricas	Anotar configuraciones usadas	Trabajar sin respaldo de datos
Sprint no técnico - 12 (por ej.: “Defensa”)	Ensayos orales con feedback	Cambiar contenidos en la memoria	Material visual claro	Dividir la presentación por bloques	Agregar gráficos difíciles de explicar