



Síntesis de voz con acento rioplatense basada en inteligencia artificial

Autor:

Ing Ojeda Juan Cruz

Director:

Título y Nombre del director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 11 de marzo de 2025 y el 22 de abril de 2025.*

Índice

| | |
|--|----|
| 1. Descripción técnica-conceptual del proyecto a realizar | 5 |
| 2. Identificación y análisis de los interesados | 6 |
| 3. Propósito del proyecto | 6 |
| 4. Alcance del proyecto | 6 |
| 4.1 El proyecto incluye: | 6 |
| 4.2 El proyecto no incluye: | 7 |
| 5. Supuestos del proyecto. | 7 |
| 6. Requerimientos | 8 |
| 7. Historias de usuarios (<i>Product backlog</i>). | 9 |
| 8. Entregables principales del proyecto | 9 |
| 9. Desglose del trabajo en tareas | 9 |
| 10. Diagrama de Activity On Node. | 10 |
| 11. Diagrama de Gantt | 11 |
| 12. Presupuesto detallado del proyecto | 14 |
| 13. Gestión de riesgos | 14 |
| 14. Gestión de la calidad | 15 |
| 15. Procesos de cierre | 16 |

Registros de cambios

| Revisión | Detalles de los cambios realizados | Fecha |
|----------|--|---------------------|
| 0 | Creación del documento | 11 de marzo de 2025 |
| 1 | Se completa hasta el punto 5 inclusive | 20 de marzo de 2025 |

Acta de constitución del proyecto

Buenos Aires, 11 de marzo de 2025

Por medio de la presente se acuerda con el Ing Ojeda Juan Cruz que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Síntesis de voz con acento rioplatense basada en inteligencia artificial” y consistirá en el desarrollo de un modelo de síntesis de voz con acento rioplatense para un VTuber basado en IA. El trabajo tendrá un presupuesto preliminar estimado de 600 horas y un costo estimado de \$ 0, con fecha de inicio el 11 de marzo de 2025 y fecha de presentación pública el 27 de octubre de 2025.

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

El presente proyecto consiste en el desarrollo de un sistema de síntesis de voz con acento rioplatense, cuyo fin es alimentar la parte vocal de una VTuber basada en Inteligencia Artificial. La motivación surge de la curiosidad y el interés en esta clase de aplicaciones interactivas, donde los usuarios pueden conversar o realizar preguntas a la VTuber y obtener respuestas auditivas en un tono familiar a la región, como el rosarino o el porteño. El poder escuchar bromas, expresiones y matices propios del español rioplatense incrementa la cercanía y la diversión para un público local.

Se trata de un *emprendimiento personal*, sin financiamiento externo ni condiciones especiales en cuanto a confidencialidad o propiedad intelectual. Respecto al *estado del arte*, la mayoría de las soluciones de síntesis de voz se encuentran en inglés, mientras que las alternativas en español —y específicamente en rioplatense— son más limitadas o menos accesibles. Por ello, la propuesta ofrece una contribución novedosa, enfocada en lograr un timbre y una prosodia afines a la región del Río de la Plata.

El proyecto se encuentra en una fase temprana de desarrollo, cercana a la etapa de *prueba de concepto* o MVP (Producto Mínimo Viable). La principal necesidad que se satisface es la generación de voz en tiempo real con un acento rioplatense realista, sin requerimientos de producción profesional o funcionalidades más elaboradas. La innovación radica en brindar una voz sintética atractiva y adaptada culturalmente para el público hispanohablante de Argentina y Uruguay.

En la figura 1 se presenta un diagrama en bloques del sistema propuesto. Como se observa, el texto o las preguntas de los usuarios se envían a un *módulo de IA* que procesa la información y, mediante un *motor de síntesis de voz rioplatense*, devuelve la señal de audio correspondiente al habla sintética. El usuario, finalmente, escucha la respuesta con una entonación acorde al dialecto objetivo.

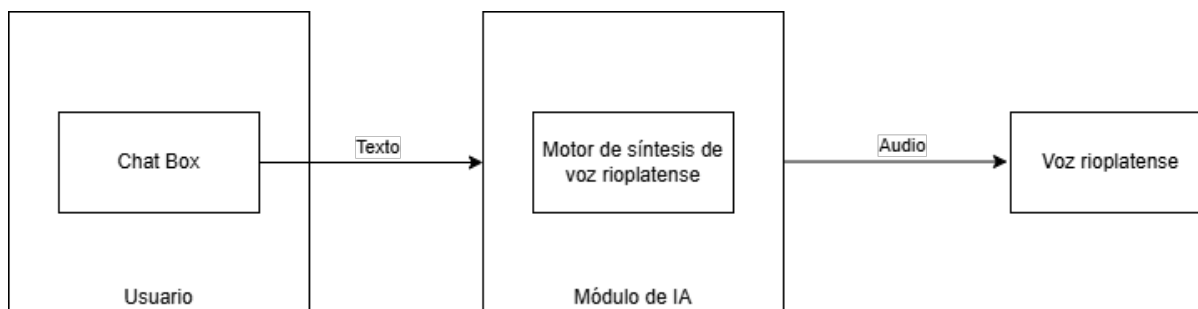


Figura 1. Diagrama en bloques del sistema propuesto para la síntesis de voz rioplatense.

La propuesta se basa en el desarrollo de modelos de aprendizaje automático entrenados con corpus de habla local, lo que permitirá capturar los matices fonéticos y entonativos característicos de la región. De esta forma, se busca brindar a los usuarios una experiencia de interacción más cercana, natural y entretenida, incrementando el valor diferenciador frente a otras soluciones de síntesis de voz en español que no contemplan las particularidades del habla rioplatense.

2. Identificación y análisis de los interesados

En la siguiente tabla se muestran los principales interesados en el proyecto. Cabe destacar que no existe un patrocinador o cliente externo, debido a que se trata de un proyecto de interés personal y académico.

| Rol | Nombre y Apellido | Organización | Puesto |
|---------------|------------------------------|--------------|---|
| Responsable | Ing Ojeda Juan Cruz | FIUBA | Alumno |
| Orientador | Título y Nombre del director | pertenencia | Director del Trabajo Final |
| Usuario final | - | - | Gamers, fans de anime, público interesado en el acento rioplatense. |

A continuación, se describen brevemente las características de cada uno:

- **Responsable (Ing Ojeda Juan Cruz):** Es el alumno a cargo de la planificación y ejecución del proyecto, encargado de recolectar y procesar datos, entrenar el modelo de síntesis de voz y elaborar la documentación final.
- **Usuario final (Gamers, fans de anime, público local):** Corresponde a la audiencia que utiliza o escucha a la VTuber. Valoran la naturalidad, la rapidez de respuesta y la entonación característica del español rioplatense. Aunque no participan activamente en el desarrollo, su satisfacción e interés son cruciales para la validación del proyecto.

3. Propósito del proyecto

Brindar una experiencia de interacción virtual inmersiva mediante la implementación de un sistema de síntesis de voz con acento rioplatense, de forma que los usuarios hispanohablantes puedan disfrutar de respuestas y entonaciones propias de la región, enriqueciendo la autenticidad y la cercanía de la VTuber basada en IA.

4. Alcance del proyecto

El presente proyecto se centrará en desarrollar la síntesis de voz con acento rioplatense para una VTuber basada en IA, de forma coherente con la arquitectura planteada en la descripción técnica (véase Figura 1).

4.1. El proyecto incluye:

- **Módulo de obtención y preparación de corpus rioplatense (Bloque de datos de voz):** Se recolectará y/o empleará un corpus de voz grabada en español rioplatense para capturar la entonación característica de la región (porteno, rosarino, etc.). Este material servirá para entrenar y validar el modelo de síntesis.

- **Diseño y entrenamiento del modelo de síntesis TTS (Bloque de motor de síntesis):** Se desarrollará y ajustará un modelo de aprendizaje automático (Text-to-Speech), capaz de convertir texto a audio con la prosodia y el timbre propios del acento rioplatense. Incluirá el pipeline necesario para el preprocesamiento de datos y la generación de espectrogramas o representaciones intermedias, siguiendo la estructura propuesta en la descripción técnica.
- **Integración básica con la IA conversacional (Bloque de IA conversacional):** Se enlazará el modelo TTS con un asistente o motor conversacional para proporcionar respuestas habladas en tiempo real. El objetivo es demostrar la viabilidad de una VTuber automatizada que hable con un acento rioplatense claro y natural.
- **Pruebas de funcionalidad y validación (Bloque de salida de audio):** Se llevarán a cabo evaluaciones para medir la naturalidad y la precisión del acento. Incluye la revisión subjetiva (feedback de usuarios finales) y la verificación objetiva (métricas de calidad de voz).
- **Documentación técnica esencial:** Se generará un manual o instructivo describiendo la configuración, el uso y la puesta en marcha del sistema, así como los resultados de las pruebas realizadas.

4.2. El proyecto no incluye:

- **Desarrollo o diseño de avatar 2D/3D de la VTuber:** No se implementará la parte gráfica ni animaciones avanzadas para el personaje virtual. El foco radica en la voz y su integración mínima con un asistente de IA.
- **Implementación de lip-sync (sincronización de labios):** No se abordará la correlación detallada entre la locución y los movimientos faciales o labiales en un personaje animado.
- **Plataforma de transmisión en vivo o monetización:** No se creará una infraestructura de streaming ni se definirá la estrategia comercial para explotar el proyecto en redes sociales o eventos.
- **Soporte de otros idiomas o acentos:** La investigación y desarrollo se limita exclusivamente al acento rioplatense. No se explorará la generación de habla en inglés, neutro, u otros dialectos hispanohablantes.
- **Fine-tuning exhaustivo de la IA conversacional:** La labor principal se enfoca en la síntesis de voz. Cualquier trabajo en la parte de conversación será de carácter básico y no incluirá un ajuste o refinamiento profundo del motor conversacional.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Se contará con el tiempo y los recursos técnicos necesarios (por ejemplo, GPU para entrenamiento) para desarrollar y entrenar el modelo de síntesis de voz durante el plazo establecido.

- Se podrá acceder a un corpus de habla rioplatense (o recolectarlo con relativa facilidad) para entrenar el modelo con calidad suficiente.
- Se dispondrá de conexión a internet y la infraestructura necesaria (servicios en la nube, repositorios, etc.) de forma estable durante todo el proceso de desarrollo.
- El equipo o la persona a cargo del proyecto contará con los conocimientos básicos en aprendizaje automático, procesamiento de lenguaje natural y manipulación de datos de audio requeridos.
- No habrá cambios drásticos en las tecnologías base (bibliotecas, *frameworks* de IA, etc.) que impacten la compatibilidad o el desarrollo del proyecto.
- No surgirán restricciones legales o normativas que impidan el uso de datos de voz o la distribución de la aplicación final.
- El público objetivo (hablantes de español rioplatense) mostrará interés en una VTuber con acento local, justificando la focalización en esta región.

6. Requerimientos

Los requerimientos deben enumerarse y de ser posible estar agrupados por afinidad, por ejemplo:

1. Requerimientos funcionales:

1.1. El sistema debe...

1.2. Tal componente debe...

1.3. El usuario debe poder...

2. Requerimientos de documentación:

2.1. Requerimiento 1.

2.2. Requerimiento 2 (prioridad menor)

3. Requerimiento de testing...

4. Requerimientos de la interfaz...

5. Requerimientos interoperabilidad...

6. etc...

Leyendo los requerimientos se debe poder interpretar cómo será el proyecto y su funcionalidad.

Indicar claramente cuál es la prioridad entre los distintos requerimientos y si hay requerimientos opcionales.

!!!No olvidarse de que los requerimientos incluyen a las regulaciones y normas vigentes!!!

Y al escribirlos seguir las siguientes reglas:

- Ser breve y conciso (nadie lee cosas largas).

- Ser específico: no dejar lugar a confusiones.
- Expresar los requerimientos en términos que sean cuantificables y medibles.

7. Historias de usuarios (*Product backlog*)

Descripción: en esta sección se deben incluir las historias de usuarios y su ponderación (*history points*). Recordar que las historias de usuarios son descripciones cortas y simples de una característica contada desde la perspectiva de la persona que desea la nueva capacidad, generalmente un usuario o cliente del sistema. La ponderación es un número entero que representa el tamaño de la historia comparada con otras historias de similar tipo.

Se debe indicar explícitamente el criterio para calcular los *story points* de cada historia.

El formato propuesto es:

1. "Como [rol] quiero [tal cosa] para [tal otra cosa]."

Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)

8. Entregables principales del proyecto

Los entregables del proyecto son (ejemplo):

- Manual de usuario.
- Diagrama de circuitos esquemáticos.
- Código fuente del firmware.
- Diagrama de instalación.
- Memoria del trabajo final.
- etc...

9. Desglose del trabajo en tareas

El WBS debe tener relación directa o indirecta con los requerimientos. Son todas las actividades que se harán en el proyecto para dar cumplimiento a los requerimientos. Se recomienda mostrar el WBS mediante una lista indexada:

1. Grupo de tareas 1 (suma h)
 - 1.1. Tarea 1 (tantas h)
 - 1.2. Tarea 2 (tantas h)

- 1.3. Tarea 3 (tantas h)
2. Grupo de tareas 2 (suma h)
 - 2.1. Tarea 1 (tantas h)
 - 2.2. Tarea 2 (tantas h)
 - 2.3. Tarea 3 (tantas h)
3. Grupo de tareas 3 (suma h)
 - 3.1. Tarea 1 (tantas h)
 - 3.2. Tarea 2 (tantas h)
 - 3.3. Tarea 3 (tantas h)
 - 3.4. Tarea 4 (tantas h)
 - 3.5. Tarea 5 (tantas h)

Cantidad total de horas: tantas.

¡Importante!: la unidad de horas es h y va separada por espacio del número. Es incorrecto escribir “23hs”.

Se recomienda que no haya ninguna tarea que lleve más de 40 h. De ser así se recomienda dividirla en tareas de menor duración.

10. Diagrama de Activity On Node

Armar el AoN a partir del WBS definido en la etapa anterior.

Una herramienta simple para desarrollar los diagramas es el Draw.io (<https://app.diagrams.net/>). Draw.io

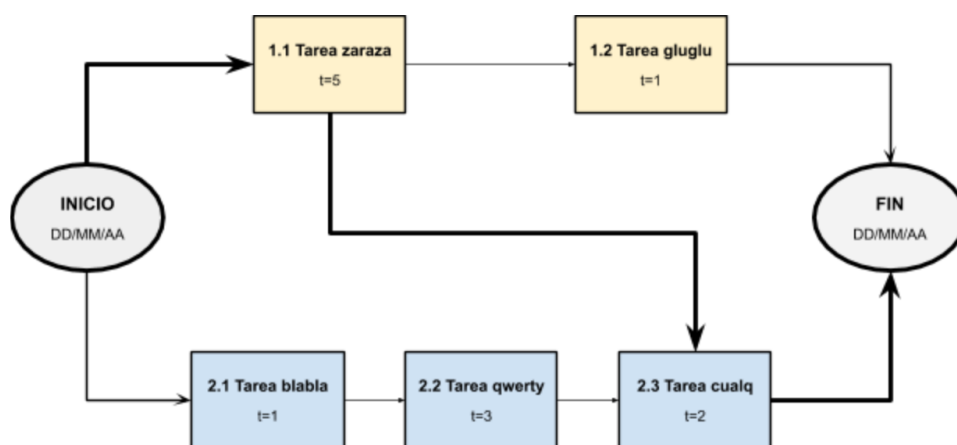


Figura 2. Diagrama de *Activity on Node*.

Indicar claramente en qué unidades están expresados los tiempos. De ser necesario indicar los caminos semi críticos y analizar sus tiempos mediante un cuadro. Es recomendable usar colores y un cuadro indicativo describiendo qué representa cada color.

11. Diagrama de Gantt

Existen muchos programas y recursos *online* para hacer diagramas de Gantt, entre los cuales destacamos:

- Planner
- GanttProject
- Trello + *plugins*. En el siguiente link hay un tutorial oficial:
<https://blog.trello.com/es/diagrama-de-gantt-de-un-proyecto>
- Creately, herramienta online colaborativa.
<https://creately.com/diagram/example/ieb3p3ml/LaTeX>
- Se puede hacer en latex con el paquete *pgfgantt*
<http://ctan.dcc.uchile.cl/graphics/pgf/contrib/pgfgantt/pgfgantt.pdf>

Pegar acá una captura de pantalla del diagrama de Gantt, cuidando que la letra sea suficientemente grande como para ser legible. Si el diagrama queda demasiado ancho, se puede pegar primero la “tabla” del Gantt y luego pegar la parte del diagrama de barras del diagrama de Gantt.

Configurar el software para que en la parte de la tabla muestre los códigos del EDT (WBS).
Configurar el software para que al lado de cada barra muestre el nombre de cada tarea.
Revisar que la fecha de finalización coincida con lo indicado en el Acta Constitutiva.

En la figura 3, se muestra un ejemplo de diagrama de gantt realizado con el paquete de *pgfgantt*. En la plantilla pueden ver el código que lo genera y usarlo de base para construir el propio.

Las fechas pueden ser calculadas utilizando alguna de las herramientas antes citadas. Sin embargo, el siguiente ejemplo fue elaborado utilizando [esta hoja de cálculo](#).

Es importante destacar que el ancho del diagrama estará dado por la longitud del texto utilizado para las tareas (Ejemplo: tarea 1, tarea 2, etcétera) y el valor x *unit*. Para mejorar la apariencia del diagrama, es necesario ajustar este valor y, quizás, acortar los nombres de las tareas.

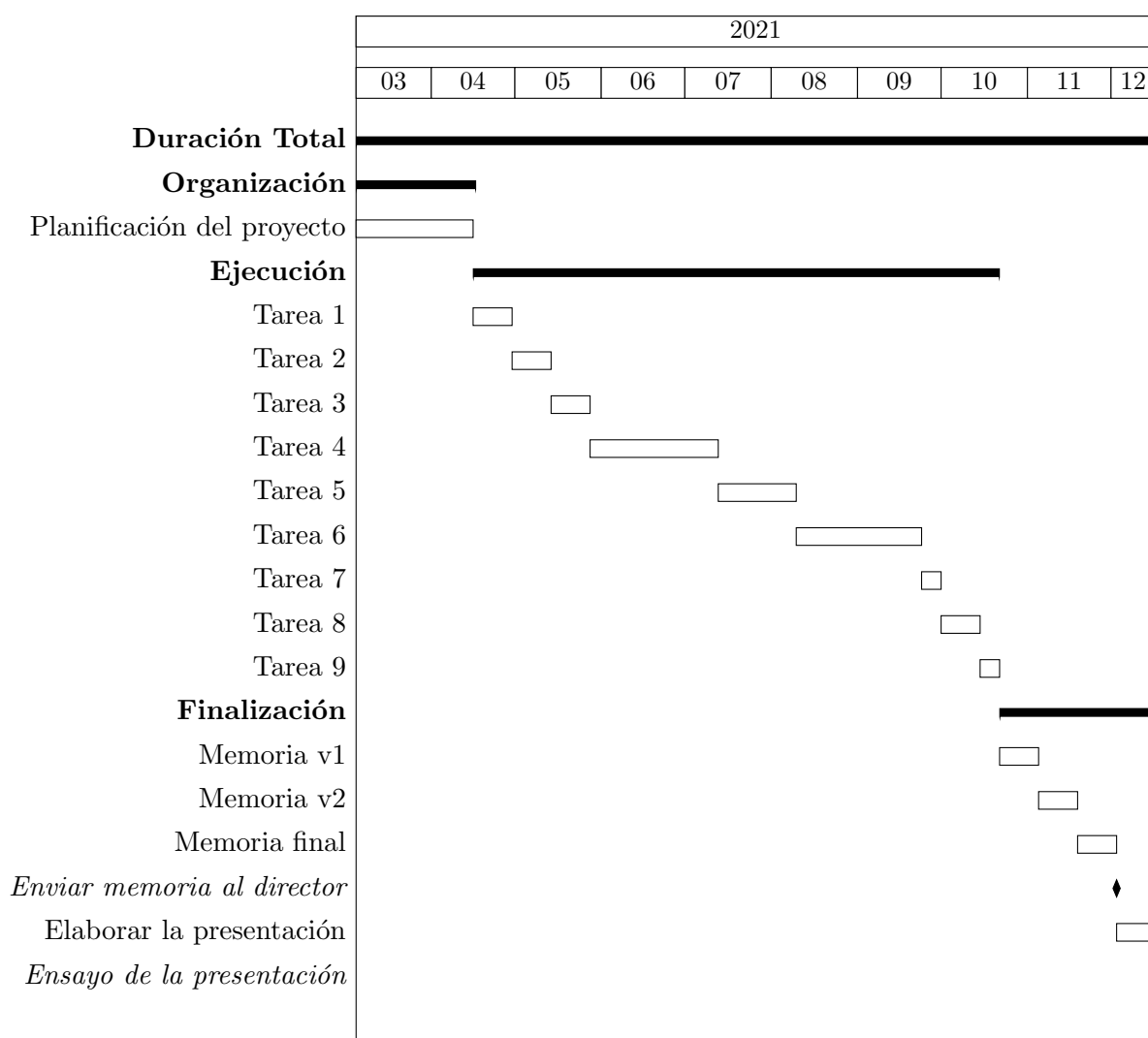


Figura 3. Diagrama de gantt de ejemplo



Figura 4. Ejemplo de diagrama de Gantt (apaisado).

12. Presupuesto detallado del proyecto

Si el proyecto es complejo entonces separarlo en partes:

- Un total global, indicando el subtotal acumulado por cada una de las áreas.
- El desglose detallado del subtotal de cada una de las áreas.

IMPORTANTE: No olvidarse de considerar los **COSTOS INDIRECTOS**.

Incluir la aclaración de si se emplea como moneda el peso argentino (ARS) o si se usa moneda extranjera (USD, EUR, etc). Si es en moneda extranjera se debe indicar la tasa de conversión respecto a la moneda local en una fecha dada.

| COSTOS DIRECTOS | | | |
|-------------------|----------|----------------|-------------|
| Descripción | Cantidad | Valor unitario | Valor total |
| | | | |
| | | | |
| | | | |
| | | | |
| SUBTOTAL | | | |
| COSTOS INDIRECTOS | | | |
| Descripción | Cantidad | Valor unitario | Valor total |
| | | | |
| | | | |
| | | | |
| | | | |
| SUBTOTAL | | | |
| TOTAL | | | |

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 $RPN=S \times O$)

| Riesgo | S | O | RPN | S* | O* | RPN* |
|--------|---|---|-----|----|----|------|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Criterio adoptado:

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

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. Gestión de la calidad

Elija al menos diez requerimientos que a su criterio sean los más importantes/críticos/que aportan más valor y para cada uno de ellos indique las acciones de verificación y validación que permitan asegurar su cumplimiento.

- Req #1: copiar acá el requerimiento con su correspondiente número.

- Verificación para confirmar si se cumplió con lo requerido antes de mostrar el sistema al cliente. Detallar.
- Validación con el cliente para confirmar que está de acuerdo en que se cumplió con lo requerido. Detallar.

Tener en cuenta que en este contexto se pueden mencionar simulaciones, cálculos, revisión de hojas de datos, consulta con expertos, mediciones, etc.

Las acciones de verificación suelen considerar al entregable como “caja blanca”, es decir se conoce en profundidad su funcionamiento interno.

En cambio, las acciones de validación suelen considerar al entregable como “caja negra”, es decir, que no se conocen los detalles de su funcionamiento interno.

15. Procesos de cierre

Establecer las pautas de trabajo para realizar una reunión final de evaluación del proyecto, tal que contemple las siguientes actividades:

- Pautas de trabajo que se seguirán para analizar si se respetó el Plan de Proyecto original:
 - Indicar quién se ocupará de hacer esto y cuál será el procedimiento a aplicar.
- Identificación de las técnicas y procedimientos útiles e inútiles que se emplearon, los problemas que surgieron y cómo se solucionaron:
 - Indicar quién se ocupará de hacer esto y cuál será el procedimiento para dejar registro.
- Indicar quién organizará el acto de agradecimiento a todos los interesados, y en especial al equipo de trabajo y colaboradores:
 - Indicar esto y quién financiará los gastos correspondientes.