



Inteligencia Artificial Large Language Model (LLM) en Servicios de Ambientación Narrativa

Autor:

Mario Gómez Alonso

Director:

Por ser definido (Por ser definida)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 17 de octubre de 2023 y el 5 de diciembre de 2023?.*

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	7
5. Supuestos del proyecto.	7
6. Requerimientos	7
7. Historias de usuarios (<i>Product backlog</i>).	8
8. Entregables principales del proyecto	8
9. Desglose del trabajo en tareas	9
10. Diagrama de Activity On Node.	9
11. Diagrama de Gantt	10
12. Presupuesto detallado del proyecto	13
13. Gestión de riesgos	13
14. Gestión de la calidad	14
15. Procesos de cierre	15

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	17 de octubre de 2023
1	Se completa hasta el punto 5 inclusive	31 de octubre de 2023

Acta de constitución del proyecto

Buenos Aires, 17 de octubre de 2023

Por medio de la presente se acuerda con el Ing. Mario Gómez Alonso que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Inteligencia Artificial Large Language Model (LLM) en Servicios de Ambientación Narrativa”, consistirá esencialmente en la implementación de un servidor web que aloje una Inteligencia Artificial de tipo LLM que genere contenido narrativo a través de servicios REST, y tendrá un presupuesto preliminar estimado de 600 h de trabajo y \$0, con fecha de inicio 17 de octubre de 2023 y fecha de presentación pública Por ser definido.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Hans Manuel Grenner Noguerón
Critical Match

Por ser definido
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

El proyecto a realizar se enfoca en ampliar las herramientas que ofrece el producto Critical Match, creado por la empresa con el mismo nombre y el cual consta actualmente de una aplicación para Android y iOS.

La principal función de la App es permitir a sus usuarios la creación o búsqueda de salas de juego, también conocidas como mesas, donde pueden unirse a otros jugadores para organizar partidas de rol, ya sea en línea o de forma presencial. Las mesas ofrecen un marco narrativo para que los usuarios seleccionen la que más se ajuste a sus preferencias.

Es en este punto donde se desea ampliar la funcionalidad mediante la incorporación de servicios de Inteligencia artificial para generar, a partir de un contexto narrativo, una amplia gama de elementos adicionales. Por ejemplo: Personajes no jugables que se integren coherentemente en el mundo propuesto, eventos que enriquezcan la trama, objetos, entre otros. El estado del arte no está suficientemente avanzado en este aspecto. Soluciones de Inteligencia Artificial como ChatGPT, aunque competentes, suelen quedarse cortas en términos de profundidad y precisión, lo que no siempre satisface las expectativas de los jugadores de rol.

Por otro lado, existen webs como WorldAnvil que ofrece a sus usuarios crear y gestionar sus mundos de ficción, el llenado de todas y cada una de las categorías y elementos del mundo recae exclusivamente en el propio usuario. Es habitual que los moderadores de las partidas utilicen estas herramientas para compilar todo el contexto narrativo en un único lugar. Critical Match quiere incorporar esta funcionalidad en su App en un futuro cercano.

El objetivo principal del proyecto es crear una Inteligencia Artificial capaz de dar mejores resultados que otros motores como ChatGPT en un contexto de generación de texto de ambientación narrativa. Esto implica desarrollar un sistema de IA altamente especializado y entrenado específicamente para la generación de contenido narrativo relacionado con partidas de rol, ofreciendo una profundidad y precisión que supere las limitaciones de las soluciones generales.

Adicionalmente, se busca dividir esta funcionalidad en varios servicios REST (Representational State Transfer Protocol - Protocolo de Transferencia de Estado Representacional) para que la aplicación de Critical Match pueda acceder a las funciones de generación de contenido de forma remota. Esto permitirá a los organizadores de las mesas de rol acceder a las capacidades de la IA, integrando de manera transparente la ambientación narrativa generada en sus partidas y facilitando la creación de un mundo de ficción coherente y atractivo para los jugadores. Se prevé que la Inteligencia Artificial se aloje en un servidor que sea accesible a través de peticiones REST (Figura 1). Esto implica que el sistema de IA estará desplegado en un servidor remoto que estará constantemente en línea y disponible por el protocolo previamente mencionado.

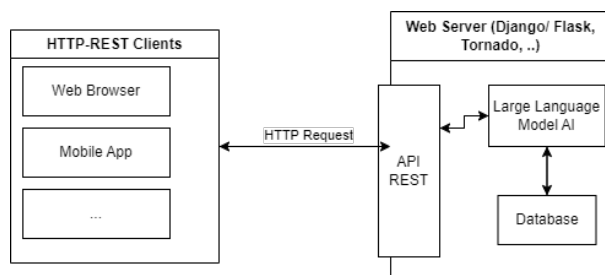


Figura 1. Diagrama en bloques del sistema

En resumen, la propuesta de valor de este proyecto es la implementación de una serie de servicios REST que serán accedidos por aplicaciones externas, principalmente Critical Match. Estos servicios permitirán la generación de elementos de ambientación narrativa, basados en el contexto del mundo proporcionado por los usuarios. Esta iniciativa supone una evolución natural de la App Critical Match. Actualmente el producto dispone de la funcionalidad de crear mesas y reunir jugadores y esta propuesta añade valor al facilitar a los organizadores de las partidas la creación de contenido atractivo para todos los participantes. Además, este proyecto podría encajar en otros ámbitos que actualmente están enfocados en el “World Building” al proporcionar a creadores de mundos y escritores herramientas para expandir sus ideas. Este proyecto a su vez es escalable y permite añadir nuevos servicios con el paso del tiempo. Esto podría llevar en un futuro a la posibilidad de ofrecer un director de juego automatizado por Inteligencia Artificial que sea de interpretar las reglas y acciones de sus jugadores, llevando a cabo partidas de rol de manera autónoma.

2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Hans Manuel Grenner Noguerón	Critical Match	co-fundador de Critical Match
Responsable	Mario Gómez Alonso	FIUBA	Alumno
Director	Por ser definido	Por ser definida	Director Trabajo final
Usuario final	Hans Manuel Grenner Noguerón	Critical Match	co-fundador de Critical Match

De esta lista podemos destacar:

- Cliente y usuario final: Hans Manuel Grenner Noguerón. Aporta contexto necesario sobre la App Critical Match e interesado de las herramientas que se desarrollarán en el proyecto. Punto de contacto entre el responsable del proyecto y el equipo de Critical Match.

3. Propósito del proyecto

El propósito del proyecto es el desarrollo de un servidor web que aloje una Inteligencia Artificial de tipo LLM. Esta IA será entrenada para la generación de contexto y contenido narrativo. El servidor podrá recibir peticiones a través del protocolo HTTP-REST, donde se alimentará a la IA con un contexto narrativo. El servidor responderá las peticiones con la salida aportada por el modelo dependiendo de la entrada de texto y el tipo de servicio solicitado

4. Alcance del proyecto

El alcance del proyecto incluye lo siguiente:

- Desarrollo de un servidor web.
- Entrenamiento de una Inteligencia Artificial de tipo LMM.
- Análisis y tratamiento de los datos utilizados para el entrenamiento de la IA.
- La integración de un conjunto esencial de servicios REST.
- Servicios adicionales, si entran dentro de las horas estimadas del proyecto.

El alcance del proyecto no incluye:

- Página web.
- Securitización del servidor web ni de ninguno de sus servicios REST.
- Rendimiento de los servicios que garanticen tiempos de respuesta cercanos a tiempo real.

En resumen, el proyecto consistirá de un prototipo que aporte funcionalidad en un tiempo de respuesta razonable, pero es posible que no se ajuste a los requerimientos de un sistema puesto en producción. Con esto aclarado, la intención siempre será desarrollar el proyecto aportando el mayor rendimiento posible.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- No se desarrollará la Inteligencia Artificial desde cero, si no que se utilizará una pre-entrenada y de código abierto. Partiendo de esa base, se hará un proceso de 'fine-tuning' para especializar el modelo en los servicios de generación de contenido narrativo.
- No se dispone de un set de datos por parte del cliente. Se analizarán sets de datos con licencia abierta en la red y se emplearán aquellos que se ajusten al proyecto.
- Comunicación con el Equipo de Critical Match para seguimiento del proyecto y aportación de ideas.

6. Requerimientos

Los requerimientos deben numerarse y de ser posible estar agruparlos 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.

El formato propuesto es: como [rol] quiero [tal cosa] para [tal otra cosa].”

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

8. Entregables principales del proyecto

Los entregables del proyecto son (ejemplo):

- Manual de uso

- Diagrama de circuitos esquemáticos
- Código fuente del firmware
- Diagrama de instalación
- Informe 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
 - 1.1. Tarea 1 (tantas h)
 - 1.2. Tarea 2 (tantas hs)
 - 1.3. Tarea 3 (tantas h)
2. Grupo de tareas 2
 - 2.1. Tarea 1 (tantas h)
 - 2.2. Tarea 2 (tantas h)
 - 2.3. Tarea 3 (tantas h)
3. Grupo de tareas 3
 - 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 h)

Se recomienda que no haya ninguna tarea que lleve más de 40 h.

10. Diagrama de Activity On Node

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

Indicar claramente en qué unidades están expresados los tiempos. De ser necesario indicar los caminos semicríticos y analizar sus tiempos mediante un cuadro. Es recomendable usar colores y un cuadro indicativo describiendo qué representa cada color, como se muestra en el siguiente ejemplo:

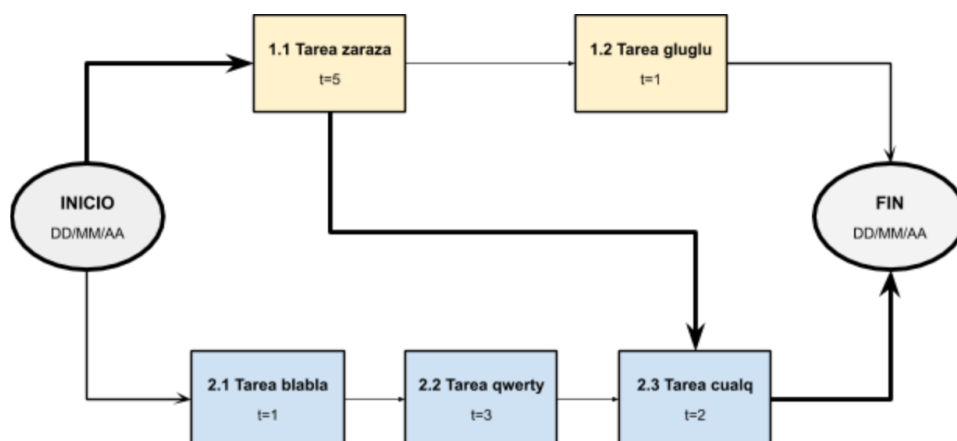


Figura 2. Diagrama de *Activity on Node*.

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.



Figura 3. Diagrama de Gantt de ejemplo



Figura 4. Ejemplo de diagrama de Gantt rotado

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**.

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):
- Ocurrencia (O):

Riesgo 3:

- Severidad (S):

■ Ocurrecia (O):

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.
 - 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, y 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.