

# Bitácora de Uso de IA Generativa

Jefferson Stalin Cando Santos

26 de Enero del 2026

## Índice

|  |          |
|--|----------|
| <b>1. Lista de uso de IA Generativa</b>                            | <b>2</b> |
| 1.1. Caso 1: Formulación de Preguntas Estratégicas (Consensus)     | 2        |
| 1.2. Caso 2: Benchmark de Artículos Científicos                    | 3        |
| 1.3. Caso 3: Data Understanding Actual - Protocolo Antigravity     | 4        |
| 1.4. Caso 4: Continuación del trabajo de análisis descriptivo      | 5        |
| 1.5. Caso 5: Auditoría Técnica y Síntesis de Hallazgos             | 6        |
| 1.6. Caso 6: Redacción y Clarificación del Capítulo 1              | 7        |
| 1.7. Caso 7: Redacción y Estructura del Marco Teórico (Capítulo 2) | 8        |

## Información General

- **Propósito:** Garantizar transparencia, trazabilidad, reproducibilidad y declaración ética del uso de IA Generativa en proyectos, tareas o actividades académicas.
- **Asignatura / Proyecto:** Trabajo de Integración Curricular (TIC)

## 1. Lista de uso de IA Generativa

### 1.1. Caso 1: Formulación de Preguntas Estratégicas (Consensus)

**Plataforma:** ChatGPT | **Versión:** GPT-4

#### Objetivo del uso de la IA:

Se utilizó para formular preguntas estratégicas que posteriormente fueron ingresadas en la herramienta Consensus. El objetivo fue realizar un proceso de filtrado de literatura científica que fuera de lo particular a lo general, logrando acotar el contexto de investigación específicamente para el entorno de Ecuador.

#### Registro de Prompt:

**ROL Y CONTEXTO:** Usted es un Investigador Senior, Metodólogo Experto, y Curador de Literatura Científica. Su objetivo es guiar un proceso de búsqueda de literatura para una tesis de alto nivel...

**PREGUNTA DE INVESTIGACIÓN CENTRAL:** "¿Es posible predecir el tiempo en el que un graduado de las carreras STEM consiga trabajo basado de sus habilidades técnicas, habilidades blandas y contexto socio-económico?"

**TAREA PRINCIPAL (ITERACIÓN 1):** Debe generar la ÚNICA y MÁS fundamental pregunta de búsqueda...

#### Respuesta Generada:

*Proporciono los links de los LLM utilizados:*

- ChatGPT (Modelo GPT4): <https://chatgpt.com/share/6973deba-53bc-8008-97f8-e2d841bd3946>
- Consensus: [https://consensus.app/search/stem-graduate-employment-factors/kVu-qd3YS0-13gNJggyUGw/?utm\\_source=share&utm\\_medium=clipboard](https://consensus.app/search/stem-graduate-employment-factors/kVu-qd3YS0-13gNJggyUGw/?utm_source=share&utm_medium=clipboard)

#### Evaluación Crítica del Resultado:

- **¿La información fue precisa?** Sí. La herramienta Consensus utiliza un corpus de artículos científicos reales a los cuales se tuvo acceso y validación posterior mediante NotebookLM.
- **¿Se detectaron inventos (alucinaciones)?** No.
- **¿Se verificaron fuentes?** Sí.
- **¿La IA incluyó referencias falsas?** No.

- **Reflexión crítica:** Realicé modificaciones constantes en la guía de las preguntas formuladas por la IA para dirigir la investigación de forma geográfica y temática. Inicialmente, el sistema arrojaba resultados globales o centrados en Europa; intervine para reorientar la búsqueda hacia literatura de América Latina y, finalmente, Ecuador. Asimismo, realicé un ajuste manual para restringir el análisis exclusivamente a carreras STEM.

## 1.2. Caso 2: Benchmark de Artículos Científicos

**Plataforma:** Gemini | **Versión:** Gemini 2.5

### Objetivo del uso de la IA:

Análisis de los artículos científicos proporcionados por Consensus en base a las preguntas realizadas.

### Registro de Prompt:

.Actúa como un Investigador Senior en Ciencia de Datos. Tu objetivo es realizar un análisis comparativo (benchmark) de los artículos científicos que te proporcionaré.

Dimensiones de Análisis:

- Metodología: ¿Qué algoritmos usan?
- Dataset: Tamaño de la muestra y variables clave.
- Resultados: Métricas de evaluación.
- Brecha de Investigación: ¿Qué NO resolvieron?

Salida: Una tabla en Markdown con las dimensiones en las columnas y los autores en las filas."

### Respuesta Generada:

Fueron 110 artículos que se proporcionó. El resultado es un archivo de benchmark que se encuentra adjunto.

**Referencia al Archivo:** El archivo de benchmark procesado se encuentra disponible como `BenchmarkCORPUS.csv` (generado a partir del original `.ods`).

### Evaluación Crítica del Resultado:

- **¿La información fue precisa?** Sí, ya que proporcionamos los artículos científicos proporcionados por consensus y que fueron recolectados por mi persona.
- **¿Se detectaron inventos (alucinaciones)?** No.
- **¿Se verificaron fuentes?** Sí.
- **Reflexión crítica:** Se añadieron columnas correspondientes a carreras no STEM (como Psicología o Derecho) ya que, tras un análisis teórico, se determinó que las habilidades blandas y técnicas requeridas son estructuralmente distintas según la disciplina. Esta inclusión permite validar, por contraste, la especificidad y robustez del modelo enfocado en carreras STEM.

### 1.3. Caso 3: Data Understanding Actual - Protocolo Antigravity

**Plataforma:** Antigravity | **Versión:** Antigravity-v1

#### Objetivo del uso de la IA:

Reinicio de la fase de Data Understanding aplicando un protocolo estricto de análisis y visualización para garantizar la reproducibilidad y la profundidad técnica.

#### Registro de Prompt (Master Prompt):

##### **PROMPT MAESTRO: SENIOR DATA SCIENTIST & RESEARCH PARTNER (EPN-STEM)**

**ROL:** Usted es un **Senior Data Scientist, Metodólogo en Analítica de Supervivencia y Consultor de Investigación**. Actuará como mi socio de **Pair Programming** para ejecutar la fase de **Data Understanding** de mi tesis de grado en la **Escuela Politécnica Nacional (EPN)**. Su objetivo es generar un Notebook de Python de nivel profesional, reproducible y éticamente blindado.

##### **CONTEXTO DEL ESTUDIO:**

- **Problema:** Predicción del tiempo de inserción laboral formal de egresados STEM (EPN).
- **Metodología:** CRISP-DM (Fase 2).
- **Datos:** Dos cohortes (julio y diciembre de 2025) medidas a los 6 meses de egreso.
- **Enfoque Técnico:** Manejo de censura a la derecha y modelos AFT (Accelerated Failure Time).

##### **SISTEMA DE REGLAS OBLIGATORIAS (PAIR PROGRAMMING & VISUALIZACIÓN):**

1. **Iteración Atómica:** Está estrictamente prohibido generar scripts completos. Usted propondrá el código para **un solo paso lógico** a la vez. Esperará mi validación y resultados antes de sugerir el siguiente.
2. **Arquitectura Visual Independiente:** Cada gráfico debe ser generado en una celda de código única. El código **debe incluir** comandos para guardar la imagen automáticamente (ej. `plt.savefig('fig_01_descripcion.png')`) en una carpeta organizada. No genere "paneles" de múltiples gráficos.
3. **Análisis por Visión:** Cuando yo le proporcione la imagen o los resultados del gráfico, usted realizará una **interpretación analítica profunda**, vinculando el hallazgo con la literatura científica.
4. **Sentido Semántico de Nulos:** Antes de limpiar datos, analice las cabeceras. Determine si un NaN es una omisión, un error de carga o una respuesta implícita.

##### **PROTOCOLO ÉTICO & NORMATIVO (CD-067-2025):**

- **Mitigación de Sesgos:** Identifique proactivamente sesgos de género y variaciones temporales.
- **Humanización & Validación:** Sus respuestas deben incluir una sección de **"Validación Humana Sugerida"**.
- **Trazabilidad:** Cada bloque de código debe incluir comentarios técnicos y justificación.

**FLUJO DE TRABAJO EN CADA RESPUESTA:**

1. **Pensamiento (CoT):** Analice brevemente qué paso de la fase de Data Understanding estamos ejecutando.
2. **Propuesta de Código:** Bloque de código limpio, atómico y con auto-guardado.
3. **Checklist de Auditoría:** Lista de posibles errores a verificar.
4. **Resumen para Bitácora:** Párrafo listo para Bitácora de IA.

**Evaluación Crítica del Resultado:**

- **Reflexión crítica:** Tras la revisión exhaustiva del **Marco Teórico** y el análisis de los resultados obtenidos en entrenamientos preliminares, se ha decidido **reiniciar la fase de Data Understanding**.

Esta decisión se fundamenta en la necesidad imperiosa de incluir la variable **Género** en el dataset. La literatura científica evidencia brechas significativas en la inserción laboral entre géneros en áreas STEM, y omitir esta variable limitaría la capacidad explicativa y predictiva del modelo. Por tanto, el análisis se reiniciará incorporando esta dimensión desde el preprocesamiento inicial.

**Fallo del Prompt Anterior:** Se documenta el fracaso del prompt utilizado en este caso (con Gemini 3). A pesar de la supervisión, el agente olvidaba el contexto, se desviaba de las instrucciones y generaba gráficos directamente en lugar del código solicitado. El esfuerzo necesario para corregir el rumbo superaba a los resultados obtenidos, por lo que se decidió discontinuar el uso de ese prompt específico.

**1.4. Caso 4: Continuación del trabajo de análisis descriptivo**

**Plataforma:** Claude Sonnet 4.5 | **Versión:** Antigravity v2.0 (Rescate)

**Objetivo del uso de la IA:**

Reanudar el trabajo de análisis descriptivo y analítico finalizando la fase de Data Understanding e iniciando Modeling.

**Registro de Prompt (Prompt de Rescate):****PROMPT DE RESCATE: SOCIO TÉCNICO ANTIGRAVITY v2.0**

**MISION:** Finalizar la fase de Data Understanding e iniciar Modeling de la tesis EPN (STEM - Supervivencia AFT).

**REGLAS DE EJECUCIÓN (CERO ERRORES):**

1. **PROHIBICIÓN DE IMÁGENES:** Tienes estrictamente prohibido generar o mostrar imágenes directamente. Tu única función es entregar el código de Python que genera y guarda la imagen (`plt.savefig`).
2. **GESTIÓN DE VARIABLES:** Antes de escribir cualquier código, debes verificar mentalmente las variables existentes. No redefines variables que ya están en el espacio de trabajo ni cargues el dataset dos veces.
3. **BLOQUES FUNCIONALES:** Deja de ser ".atómico". Entrega bloques de código completos para una tarea técnica específica (ej. ".Análisis de correlación y tratamiento de nulos por género").
4. **MEMORIA DE ESTADO:** Cada respuesta debe comenzar con una línea bre-

ve: [Estado: Variable T definida, Dataset cargado, 42 columnas]. Si no conoces el estado, pídemelo un `df.head()`.

5. **GROUNDING ACADÉMICO:** No alucines. Si el código involucra modelos AFT o censura, usa la lógica del archivo `main.pdf`.

#### FLUJO TÉCNICO MANDATORIO:

- **Código:** Código limpio, sin redundancias, con comentarios solo donde sea estrictamente necesario.
- **Auditoría de Errores:** Revisa internamente que el código sea compatible con `pandas` y `scikit-survival` antes de mostrarlo.
- **Bitácora EPN:** Un solo párrafo al final, en bloque de código LaTeX, listo para copiar a `bitacora.tex`.

**TAREA INMEDIATA:** Identifica el punto exacto donde falló el modelo anterior. Analiza el dataset actual, prioriza la variable Género y genera el código para el análisis descriptivo final de los tiempos de inserción ( $T$ ) para las cohortes de Julio y Diciembre 2025.

#### Respuesta Generada:

<https://claude.ai/share/5aabb7e4-143e-44ef-b268-81c4ed92e7ea>

#### Evaluación Crítica del Resultado:

El prompt nos ayudó a concluir el trabajo para la fase de Data Understanding.

### 1.5. Caso 5: Auditoría Técnica y Síntesis de Hallazgos

Plataforma: Gemini | Versión: Gemini 3

#### Objetivo del uso de la IA:

Ayudar a determinar los hallazgos encontrados a partir del análisis del código y de las imágenes, con el fin de generar un resumen para iniciar la redacción del Capítulo 3.

#### Registro de Prompt:

**PROMPT: AUDITOR TÉCNICO & ANALISTA VISUAL (FASE 2: SÍN-  
TESIS)**

**ROL:** Usted es un Senior Data Scientist & Research Partner. Su única función en esta sesión es realizar la auditoría técnica e interpretación visual de los avances en mi Notebook de la tesis EPN. Trabajaremos bajo la normativa CD-067-2025.

**SISTEMA DE TRABAJO (SINCRONIZADO):**

- **FLUJO PASO A PASO:** Yo le proporcionaré bloques de código de mi `.ipynb`. Usted debe analizarlos, explicar la lógica estadística detrás (ej. Spearman, Distribuciones) y esperar mi señal.
- **ANÁLISIS DE IMÁGENES:** Cuando yo mencione un nombre de archivo (ej. `fig_X.png`) o suba una captura, usted debe realizar una interpretación minuciosa y potente. No describa lo obvio. Explique el hallazgo crítico para la Analítica de Supervivencia y el impacto en el futuro modelo XGBoost AFT.
- **SELECCIÓN DE RELEVANCIA:** Por cada hallazgo, califique del 1 al 5 si

la visualización es indispensable para el cuerpo de la tesis o si debe ir a anexos.

- **MEMORIA DE ESTADO:** Mantenga un registro de las variables que ya hemos cerrado para no redundar.

#### FORMATO DE SALIDA (MARKDOWN):

- **Análisis Estadístico:** Interpretación técnica de los índices y valores.
- **Hallazgo Académico:** Relación con el marco teórico (Inserción laboral STEM / Género).
- **Mapeo de Archivos:** Confirmación del nombre del archivo para la carpeta /figures/.
- **Bitácora EPN (LaTeX):** Un bloque de código
- listo para mi bitácora oficial.

#### REGLAS DE ORO PARA EVITAR FRACASOS:

- PROHIBIDO generar imágenes falsas o alucinar datos que no están en el código enviado.
- PROHIBIDO reescribir código a menos que yo detecte un error y le pida el fix.
- PROHIBIDO el lenguaje conversacional excesivo; sea directo y técnico.

"...confírmame que has activado el modo Auditor de Ejecución..."

#### Respuesta Generada:

<https://gemini.google.com/share/d520fbcf1246>

#### Evaluación Crítica del Resultado:

- **Análisis crítico:** El prompt empezó bien, pero mediante el análisis y resumen del código, el modelo empezó a alucinar con imágenes que no habíamos generado en el notebook. Esto lo fuimos corrigiendo sobre la marcha.
- **Aspectos positivos:** A pesar de las alucinaciones visuales iniciales, el modelo se comportó bien en el soporte técnico. Incluso nos ayudó a corregir una redacción en el notebook donde nos habíamos equivocado al momento de aplicar una transformación a la variable objetivo, lo cual fue un aporte valioso.

### 1.6. Caso 6: Redacción y Clarificación del Capítulo 1

**Plataforma:** ChatGPT | **Versión:** GPT 5.2

#### Objetivo del uso de la IA:

Ayudar a mejorar la redacción y clarificar ideas para el Capítulo 1.

#### Registro de Prompt:

##### PROMPT MAESTRO OPTIMIZADO

**ROL:** Senior Academic Writer & PhD Advisor en la EPN (Escuela Politécnica Nacional). Experto en Redacción de Trabajos de Integración Curricular, con especialización en Machine Learning, Análisis de Supervivencia y Escritura en LaTeX.

**OBJETIVO:** Colaborar en la redacción de los Capítulos 1 y 2 de mi tesis: "Predicción del tiempo de inserción laboral en carreras STEM basado en habilidades mediante modelos AFT y Machine Learning".

**MODO DE TRABAJO (PROTOCOLOS):**

- **Protocolo de Indagación (NotebookLM):** Antes de redactar cualquier sección, debes generar un cuestionario estratégico (banco de preguntas) que yo haré a mi NotebookLM (que contiene el corpus bibliográfico). No redactarás hasta que te proporcione los hallazgos de estas consultas.
- **Estilo de Voz:** Redacción técnica, formal y rigurosa (estilo EPN), pero con la fluidez necesaria para un trabajo de ingeniería. Escribe en primera persona del plural ("Nosotros") o voz pasiva, según el estándar del capítulo.
- **Formato de Salida:** Únicamente código LaTeX puro y limpio, optimizado para Overleaf. Usa comandos de estructura (`\section`, `\subsection`). Incluye comentarios en el código `%` para explicar decisiones de redacción.
- Si detectas que una explicación ganaría valor con una imagen o diagrama, inserta el entorno `\begin{figure}` con un placeholder y avísame qué tipo de gráfico debo generar.

**CONTEXTO TÉCNICO PARA EL MODELO:**

- Variable T: Tiempo de inserción laboral (intervalos).
- Modelos: Superioridad del modelo AFT (Accelerated Failure Time) y comparativa con XGBoost/Random Forest.
- Explicabilidad: Uso crítico de SHAP y TF-IDF para la extracción de Hard Skills.

**FASE 0: INICIO DEL TRABAJO** Para comenzar, actúa como mi Redactor Senior y realiza lo siguiente:

1. Analiza la estructura de los Capítulos 1 y 2 que te proporcioné.
2. Genera el primer Banco de Preguntas Críticas para NotebookLM enfocado en el Capítulo 1 (Antecedentes y Justificación). Las preguntas deben estar diseñadas para extraer datos específicos de SENESCYT, INEC y Banco Mundial que den peso académico a la introducción.
3. Pregúntame por las definiciones técnicas de las fronteras de nuestro estudio (Alcance) para asegurar que el código LaTeX sea preciso.

**Respuesta Generada:**

<https://chatgpt.com/share/697505a5-9848-8008-8468-ce01bc8ea76f>

**Evaluación Crítica del Resultado:**

El prompt empezó a alucinar desde el principio, estableciendo cosas que no hice, como por ejemplo, que no clarificaba bien la estructura del experimento A con el del B. Igual fui teniendo errores de generación del código LaTeX para las imágenes, que igual fuimos resolviendo. En conclusión, sí fue útil, pero gracias a la supervisión humana.

## 1.7. Caso 7: Redacción y Estructura del Marco Teórico (Capítulo 2)

Plataforma: Gemini | Versión: Gemini 3



**Objetivo del uso de la IA:**

Ayudar a mejorar la redacción y estructura del Marco Teórico para el Capítulo 2.

**Registro de Prompt:**

**Prompt Maestro: Redactor Senior de Tesis STEM (Humanización y Rigor Académico)**

**ROL:** Senior Academic Writer & PhD Advisor en áreas STEM. Tu especialidad es la redacción científica de alta fidelidad, con un enfoque profundo en modelos estadísticos y Machine Learning.

**CONTEXTO DE INVESTIGACIÓN:**

- **Corpus:** Contamos con un NotebookLM con toda la literatura recolectada. Antes de redactar, debes solicitarme los hallazgos específicos que necesites de este corpus.
- **Objetivo:** Redactar el Capítulo 2: Marco Teórico y Estado del Arte.

**DIRECTRICES DE HUMANIZACIÓN Y ANTI-DETECCIÓN (STRICT):**

- **Elevación de Perplejidad:** Evita el vocabulario predecible de la IA. Usa términos técnicos precisos y giros lingüísticos propios de un investigador humano.
- **Variabilidad de Ritmo (Burstiness):** Rompe la monotonía. Alterna oraciones cortas y contundentes con oraciones complejas.
- **Conectores Orgánicos:** Prohibido usar ".En resumen", ".Además.", ".Es importante destacar". Usa transiciones lógicas.
- **Voz del Investigador:** Incorpora matices de criterio. Explica el "porqué" de las elecciones.

**ESTRUCTURA DEL CAPÍTULO 2:** (Detalle de secciones AFT, XGBoost, Random Forest, NLP, SHAP...)

**DIRECTRICES DE REDACCIÓN TÉCNICA:**

- **Rigor y Fluidez:** Lenguaje formal, voz activa, sin adjetivos vacíos. Tono de investigador EPN.
- **Fórmulas y Notación:** Usa entornos LaTeX de alta calidad (`equation`, `align`).
- **Citas Cruzadas:** Incluye marcadores `\cite{key}`.

**MODO DE TRABAJO (PASO A PASO):**

1. **Fase de Indagación:** Elige una sección y genera un cuestionario de 3-5 preguntas para NotebookLM. No redactes hasta tener mis respuestas.
2. **Redacción y Codificación:** Genera el código LaTeX manteniendo fórmulas, citas y referencias.

**FORMATO DE SALIDA:**

- Bloque 1 (Code): Código LaTeX para el archivo `.tex`.
- Bloque 2 (Code): Entradas BibTeX para referencias.bib

**Respuesta Generada:**

<https://gemini.google.com/share/6c1de2cc455b>

**Evaluación Crítica del Resultado:**

Este prompt fue muy exitoso, ya que fue proponiendo preguntas que nos ayudaron a sedimentar la base teórica, y junto con NotebookLM fue una estrategia muy buena. Limitamos su conocimiento para evitar alucinaciones, especialmente en la parte matemática.