# Proyecto Investigación Sistema Multiplataforma con IA

## Aplicaciones de la IA

La IA se ha integrado en herramientas que facilitan tareas como la búsqueda de información, traducción de documentos, generación de contenido multimedia, y corrección de textos, entre otras, demostrando su capacidad para interactuar en lenguaje natural y realizar tareas complejas de forma eficiente.

En el ámbito empresarial, la IA contribuye a la automatización de procesos repetitivos, análisis de datos para predecir tendencias, personalización de la experiencia del cliente, optimización logística, y mejora de la toma de decisiones. Empresas como Amazon, Coca-Cola, IBM, Tesla, y Netflix han adoptado la IA para mejorar su rentabilidad y competitividad.

Algunas Soluciones IA de Operaciones Diarias:

* **Fliki.ai**: Convierte textos en videos usando inteligencia artificial.
* **DALL-E 2**: Crea imágenes realistas o artísticas a partir de descripciones de texto.
* **Speechtext.ai**: Transcribe audio y vídeo a texto.
* **Summarize.tech**: Genera resúmenes de vídeos de YouTube.
* **Grammarly**: Corrige y mejora textos en inglés.

## Inteligencia Artificial en Marketing:

Específicamente en marketing, la IA se emplea en la generación de contenido, optimización de SEO, gestión de relaciones con clientes, y traducción de textos. Herramientas como Frase.io, Salesforce Einstein, y Grammarly, entre otras, utilizan la IA para aumentar la productividad, mejorar la precisión en la comunicación, y ofrecer experiencias personalizadas y predictivas a los clientes. Además, plataformas como Synthesia y Zoho Zia demuestran cómo la IA puede facilitar la creación de contenido multimedia y la mejorar interacción con el cliente.

Algunas Soluciones IA de automatización laboral:

* **Chat GPT (GPT3)**: Facilita la generación de contenido y tiene múltiples usos en marketing digital.
* **Frase.io**: Optimiza el SEO mediante la búsqueda de palabras clave y creación de contenido relevante.
* **Zoho Zia**: Asistente de voz que transforma datos en información útil para la toma de decisiones en empresas.
* **Conversica**: Automatiza la comunicación con posibles clientes hasta convertirlos en oportunidades de venta.
* **Reply**: Diseña flujos de trabajo y realiza búsquedas de correos electrónicos para aumentar bases de datos.
* **Sembly**: Transcribe y analiza reuniones, transformándolas en información procesable para equipos de venta.
* **ClickUp**: Ayuda a la gestión de tareas y proyectos optimizando el tiempo.

## Tecnologías open-source de inteligencia artificial

* TensorFlow y Keras para aprendizaje profundo y redes neuronales.
* Scikit-learn para aprendizaje automático con datos estructurados.
* NLTK o SpaCy para procesamiento de lenguaje natural.
* OpenCV para procesamiento de imágenes y visión por computadora.
* Pandas para análisis de datos estructurados.
* PyTorch desarrollado por Facebook, PyTorch es otra biblioteca de aprendizaje automático que ha ganado popularidad por su flexibilidad y dinamismo.
* FastAI es una biblioteca que se construye sobre PyTorch y simplifica el entrenamiento de modelos de aprendizaje profundo.

## Tecnologías de IA aplicables a CESEL: GPT-Neo y GPT-J

**GPT-Neo** es una implementación de código abierto del modelo GPT diseñada para ser ejecutada en GPUs, con la intención de ofrecer una alternativa abierta a los modelos GPT-3 de OpenAI. GPT-Neo está disponible en distintas versiones que varían en tamaño y capacidad, lo que te permite elegir el modelo que mejor se adapte a tus necesidades y recursos disponibles.

**GPT-J**, por otro lado, es una evolución de GPT-Neo con una arquitectura de 6 billones de parámetros, proporcionando capacidades de generación de texto y comprensión aún más potentes. GPT-J ha demostrado un rendimiento impresionante en una variedad de tareas de procesamiento de lenguaje natural y se considera una alternativa de código abierto muy competente a modelos comerciales como GPT-3.

Requerimiento de hardware

* GPUs: Una sola GPU NVIDIA A100 o H100 puede costar entre $10,000 y $30,000 USD dependiendo de la configuración. Para un cluster de 100 GPUs, es un costo de $1 millón a $3 millones USD solo en GPUs.
* CPUs y Memoria RAM: Un servidor de alta gama con CPUs potentes y gran cantidad de RAM podría costar entre $10,000 y $50,000 USD. Necesitarías varios de estos para soportar las GPUs, así que este costo podría sumar otros cientos de miles de dólares.
* Almacenamiento: Dependiendo de la solución de almacenamiento (por ejemplo, SSDs de alta velocidad), podrías gastar entre $20,000 y $100,000 USD o más.

Alternativas

Dado el alto costo de adquisición y operación de este hardware se podría optar por utilizar servicios en la nube que ofrecen acceso a GPUs de alta gama y otros recursos en un modelo de pago por uso.

Comparación con GPT-3.5

Para modelos específicos como GPT-3, sobre el cual se basa la versión inicial de ChatGPT, el conjunto de datos utilizado para su entrenamiento fue el Common Crawl, un archivo que recoge el contenido de la web, complementado con fuentes adicionales para mejorar la cobertura en áreas específicas y reducir sesgos. El peso aproximado de la data de entrenamiento fue de 70 Tbytes.

Estimación del Tiempo del Proyecto

La duración del desarrollo de un sistema basado en un modelo de inteligencia artificial como GPT-J, alimentado con 40 TB de datos propios de una empresa, puede variar ampliamente según varios factores. Aquí hay algunos aspectos que influirán en el tiempo de desarrollo:

* Preparación de Datos: La fase de preparación de datos, que incluye la recopilación, limpieza, y organización de los datos, puede ser muy prolongada, especialmente con una gran cantidad de datos. Este proceso puede llevar desde semanas hasta meses dependiendo de la calidad y formato de los datos existentes.
* Diseño Experimental: La fase de diseño incluye la selección del modelo, la definición de la arquitectura, y la planificación de los experimentos de entrenamiento. Esta fase puede llevar desde unas pocas semanas hasta varios meses, dependiendo de la complejidad del modelo y los requisitos específicos del proyecto.
* Entrenamiento del Modelo: El tiempo de entrenamiento depende del tamaño del modelo, la capacidad de computación disponible, y la eficiencia de la implementación. Entrenar un modelo del tamaño de GPT-J con 40 TB de datos podría llevar desde semanas hasta varios meses, incluso utilizando cientos a miles de GPUs de alta capacidad.
* Evaluación y Ajustes: Después del entrenamiento, el modelo debe ser evaluado para asegurar que cumple con los objetivos de rendimiento. Basado en los resultados, pueden ser necesarios ajustes y reentrenamientos adicionales, lo que podría añadir semanas o meses adicionales al cronograma.
* Integración y Despliegue: Integrar el modelo entrenado en un sistema operativo y desplegarlo en producción también lleva tiempo. Dependiendo de la complejidad del entorno de TI de la empresa y los requisitos del sistema, esta fase puede llevar desde unas pocas semanas hasta varios meses.

Presupuesto tentativo al utilizar modelos pre-entrenados como GPT-J

* Costo de Hardware: Con infraestructura propia el costo inicial incluiría la adquisición de GPUs de alta capacidad (por ejemplo, NVIDIA A100 o similares), servidores, almacenamiento, y la infraestructura de red.
* Duración del Entrenamiento: El costo también depende de cuánto tiempo necesitas entrenar el modelo. El ajuste fino puede llevar desde algunas horas hasta semanas, dependiendo de la complejidad del modelo y la cantidad de datos
* Costos de Desarrollo y Operación: Incluyen el tiempo del equipo de desarrolladores e investigadores que trabajan en el ajuste fino, la preparación de datos, la evaluación, y la implementación del modelo.

**Presupuesto Tentativo**

|  |  |
| --- | --- |
| **COMPONENTES** | **COSTO** |
| 2 GPU DE ALTA GAMA - MARCA NVIDIA A100 DE 40 GB O 80 GB VRAM | US$ 20,000.00 |
| 1 CPU DE ALTA GAMA SERIES AMD Ryzen Threadripper o los Intel Xeon Scalable | US$ 4,999.00 |
| ALMACENAMIENTO SSD DE 40 TBytes Y/O HDD EXTRA | US$ 2,408.97 |
| MEMORY SMART HPE 128GB | US$ 5,634.00 |
| **TOTAL** | **US$ 33,041.97** |

**Requerimientos:**

* 2 GPU DE ALTA GAMA - MARCA NVIDIA A100 DE 40 GB O 80 GB VRAM
* 1 CPU DE ALTA GAMA SERIES AMD Ryzen Threadripper o los Intel Xeon Scalable
* ALMACENAMIENTO SSD DE 40 TBytes Y/O HDD EXTRA
  + SSD PARA EL ALMACENAMIENTO DEL DATASET
  + HDD PARA EL ALMACENAMIENTO DE DATOS A LARGO PLAZO
* MEMORIA RAM DE 256 GB O 512 GB

**Precio por unidad:**

* NVIDIA A100 /unidad = 10.000 dólares
* AMD Ryzen Threadripper /unidad = 4,999 dólares
* Intel Xeon Scalable /unidad = 2,750 dólares
* SSD 5 TBytes /unidad = 249.99 dólares
* HDD 5TBytes /unidad = 1500 soles
* Memory Smart HPE 128GB /unidad = 2817 dólares

**Costo Total:**

* 33,042 dólares - 35,000 dólares

**Otras alternativas:**

* **HPE Proliant DL320 Gen11** -> $2563 - $4012
* **Servicio Nube Azure con 40 TB** -> $3890,83 al mes