**1. Antecedentes**

(a) **Investigación previa:**

* La investigación previa sobre la evaluación de modelos de generación de texto se ha centrado en una variedad de metodologías, métricas y desafíos.
* Las metodologías comunes incluyen enfoques cuantitativos como el puntaje BLEU, el puntaje ROUGE y el puntaje METEOR, así como enfoques cualitativos como la evaluación humana.
* Las métricas comunes incluyen precisión, fluidez, creatividad e informatividad.
* Los desafíos comunes incluyen sesgo de datos, sesgo del modelo y la dificultad de definir lo que constituye "buen" texto.

(b) **Pregunta de investigación principal:**

¿Qué tan precisos son los resultados de texto generados por modelos creados en Google AI Studio?

(c) **Preguntas de investigación adicionales:**

* ¿Cómo varía la precisión de los resultados de texto según el tipo de modelo?
* ¿Cómo varía la precisión de los resultados de texto según la tarea?
* ¿Cómo evalúan los desarrolladores la precisión de los resultados de texto generados por modelos creados en Google AI Studio?

**2. Diseño**

(a) **Enfoque de estudio de caso:**

* Se utilizará un diseño de estudio de caso múltiple, con tres casos que representen diferentes tipos de modelos (por ejemplo, modelos de lenguaje, modelos de subtitulado de imágenes, modelos de texto a código).
* Se utilizará un enfoque holístico, ya que el estudio examinará todos los aspectos de la precisión de los resultados de texto generados por modelos creados en Google AI Studio.

(b) **Objeto de estudio:**

* El objeto de estudio serán los resultados de texto generados por los tres modelos en cada caso.
* Los detalles relevantes a recopilar incluirán la precisión de los resultados, las tareas para las que se generaron y los comentarios de los desarrolladores que han utilizado los modelos.

(c) **Proposiciones y medidas:**

**Proposición 1:** La precisión de los resultados de texto generados por modelos creados en Google AI Studio varía según el tipo de modelo.

* Medida: La precisión de los resultados se medirá utilizando una variedad de métricas, incluidas el puntaje BLEU, el puntaje ROUGE y el puntaje METEOR.

**Proposición 2:** La precisión de los resultados de texto generados por modelos creados en Google AI Studio varía según la tarea.

* Medida: La precisión de los resultados se medirá para una variedad de tareas, como traducción, resumen y respuesta a preguntas.

**Proposición 3:** Los desarrolladores evalúan la precisión de los resultados de texto generados por modelos creados en Google AI Studio utilizando una variedad de criterios.

* Medida: Los desarrolladores serán entrevistados para comprender sus criterios para evaluar la precisión de los resultados de texto.

**3. Selección**

(a) **Criterios de selección de casos:**

* Los tres modelos en cada caso se seleccionarán en función de los siguientes criterios:
  + Tipo de modelo (por ejemplo, modelo de lenguaje, modelo de subtitulado de imágenes, modelo de texto a código)
  + Puntos de referencia de precisión
  + Adopción de usuarios

**4. Procedimientos**

(a) **Procedimientos de campo:**

* Se recopilarán los siguientes datos:
  + Los resultados generados por los modelos para una variedad de tareas
  + La precisión de los resultados, medida por una variedad de métricas
  + Los comentarios de los desarrolladores que han utilizado los modelos
* Los datos se recopilarán utilizando una variedad de métodos, incluidos:
  + Evaluación manual de los resultados
  + Evaluación automatizada utilizando métricas
  + Entrevistas con desarrolladores
* Se tomarán en cuenta las consideraciones éticas, como anonimizar los datos y obtener el consentimiento informado de los participantes.

(b) **Roles y responsabilidades:**

* El equipo de investigación estará formado por los siguientes miembros:
  + Investigador principal (IP)
  + Asistente de investigación (RA)
  + Analista de datos
* El IP será responsable de la dirección general del estudio, incluida la elaboración del plan de investigación, la recopilación y el análisis de datos y la redacción del informe.
* El RA será responsable de la recopilación de datos, incluida la evaluación manual de los resultados y las entrevistas con los desarrolladores.
* El analista de datos será responsable del análisis de datos, incluida el cálculo de métricas e identificación de patrones.

**5. Recopilación de datos**

(a) **Tipos de datos:**

* Se recopilarán los siguientes tipos de datos:
  + Resultados de texto generados por los modelos
  + Precisión de los resultados, medida por una variedad de métricas
  + Comentarios de los desarrolladores

b) **Plan de Recopilación de Datos:**

* Los datos se recopilarán durante un periodo de seis meses.
* Los resultados generados por los modelos se recopilarán mediante la evaluación manual de una muestra de los resultados.
* La precisión de los resultados se medirá utilizando una variedad de herramientas automatizadas que calculen métricas como BLEU, ROUGE y METEOR.
* Los comentarios de los desarrolladores se recopilarán a través de entrevistas semi-estructuradas realizadas de forma individual o en grupos pequeños.
* Todas las interacciones con los desarrolladores se grabarán digitalmente con su consentimiento informado y se transcribirán para su posterior análisis.
* Se anonimizarán todos los datos personales de los desarrolladores para garantizar la confidencialidad.

(c) **Almacenamiento de Datos:**

* Los datos se almacenarán en un servidor seguro con acceso restringido.
* Se realizarán copias de seguridad periódicas de los datos para evitar la pérdida.
* Los datos se anonimizarán antes de cualquier uso externo del estudio.

**6. Análisis:**

(a) **Criterios de Interpretación:**

* Los hallazgos del estudio de caso se interpretarán considerando la exhaustividad, consistencia y posibles sesgos de los datos.
* Se utilizará el análisis triangular para comparar y verificar la coherencia entre las diferentes fuentes de datos (resultados de texto, métricas y entrevistas de desarrolladores).
* Se aplicará el razonamiento crítico para evaluar la plausibilidad y relevancia de los hallazgos en relación con las preguntas de investigación.

(b) **Mapeo de Datos-Pregunta:**

* Se identificarán los elementos de datos específicos que se relacionan con cada pregunta de investigación, subpregunta y proposición.
* Se utilizarán técnicas de análisis temático para codificar y categorizar los datos textuales (entrevistas y transcripciones).
* Se utilizarán análisis estadísticos para comparar y contrastar los resultados de las diferentes métricas de precisión.
* Se combinarán los datos de diversas fuentes para obtener una comprensión más completa de la precisión de los resultados de texto generados por los modelos.

(c) **Explicación Alternativa:**

* Se considerarán y se identificarán posibles explicaciones alternativas para los hallazgos del estudio, más allá de las hipótesis planteadas.
* Se recopilarán datos adicionales, si es necesario, para diferenciar entre las posibles explicaciones y fortalecer la validez de las conclusiones.
* Se discutirán las limitaciones del estudio y las posibles direcciones para futuras investigaciones.

(d) **Análisis Iterativo:**

* El análisis de datos se realizará de forma continua a lo largo del estudio de caso, lo que permitirá realizar ajustes y obtener nuevas perspectivas a medida que surjan nuevos datos.
* Se realizarán informes preliminares periódicos para compartir los hallazgos emergentes con el equipo de investigación y obtener retroalimentación.
* El informe final del estudio reflejará la evolución del análisis y presentará las conclusiones finales de manera clara y convincente.

**7. Validez del Plan:**

(a) **Validación mediante checklist:**

* Se comparará el plan de estudio con checklists de diseño de estudios de caso y recolección de datos, como los propuestos por Ho6st y Runeson. Esto asegurará que el plan tenga rigor científico y siga buenas prácticas de investigación.
* Se documentarán los puntos en los que el plan cumple los criterios del checklist y se explicarán las razones para cualquier desviación.

(b) **Validez de constructo:**

* Se garantizará que las métricas elegidas (BLEU, ROUGE, METEOR) reflejen adecuadamente la precisión de los resultados de texto generados (fluidez, gramática, coherencia).
* Se utilizarán múltiples fuentes de datos para evaluar la precisión, incluyendo la evaluación automática con métricas, la evaluación manual por expertos y las opiniones de los desarrolladores.
* Se consultará con expertos en generación de texto e inteligencia artificial para validar la pertinencia de las métricas y los métodos de evaluación.

(c) **Validez interna (no aplicable al estudio de caso actual):**

* Debido a que este es un estudio de caso exploratorio y descriptivo, no se intenta establecer una relación causal entre variables.
* Sin embargo, se controlarán las variables externas que puedan afectar la precisión de los modelos, como la tarea específica, el tamaño de la muestra y la configuración del modelo.

(d) **Validez externa:**

* Se definirá el alcance de la generalización de los hallazgos, reconociendo que los resultados obtenidos son específicos a los modelos y tareas analizados.
* Se considerará el uso de la teoría existente sobre la generación de texto para ampliar la aplicabilidad de los hallazgos a otros modelos y contextos.
* Se explorará la posibilidad de realizar estudios similares con otros tipos de modelos y tareas para aumentar la validez externa.

**8. Limitaciones del estudio:**

* Se reconocerán las limitaciones inherentes al estudio, como:
  + El número limitado de modelos analizados.
  + La subjetividad de la evaluación manual por parte de expertos y desarrolladores.
  + La posibilidad de sesgos en los datos de entrenamiento de los modelos.
* Se explicará cómo estas limitaciones podrían afectar la interpretación de los resultados y se sugerirán formas de mitigarlas en futuras investigaciones.

**9. Reportes:**

* El público objetivo del estudio serán investigadores en inteligencia artificial, desarrolladores de software y profesionales interesados en la generación de texto.
* Los hallazgos del estudio se utilizarán para:
  + Informar el desarrollo de mejores métricas y métodos de evaluación para modelos de generación de texto.
  + Proporcionar a los desarrolladores información sobre las fortalezas y debilidades de los diferentes modelos en cuanto a la precisión.
  + Contribuir a la investigación general sobre la inteligencia artificial y la generación de lenguaje natural.

**10. Calendario:**

* Se estima un tiempo de seis meses para la realización del estudio, distribuido de la siguiente manera:
  + Planificación y selección de modelos: 1 meses.
  + Recopilación de datos: 2 meses.
  + Análisis de datos: 2 meses.
  + Redacción del informe final: 1 mes.
* Se reconoce la naturaleza iterativa del proceso de investigación y se deja espacio para ajustes en el calendario si es necesario.

**11. Apéndices:**

(a) **Informe de validación:**

* Se incluirá como apéndice una tabla que compare el plan de estudio con los criterios del checklist de Host y Runeson, marcando los puntos cumplidos y explicando las desviaciones.

(b) **Registro de divergencias:**

* Se mantendrá un registro a lo largo del estudio documentando cualquier cambio realizado al protocolo original, junto con la justificación lógica para cada cambio.