



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Departament
d'Economia i
Ciències Socials

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Economía y Ciencias Sociales

Evaluación de la eficiencia de Modelos de Lenguaje
Grande (LLMs) en el análisis de datos cualitativos en el
sector agrario

Trabajo Fin de Máster

Máster Universitario en Economía Agroalimentaria y del Medio
Ambiente

AUTOR/A: Lion, Jason Edward

Tutor/a: Calafat Marzal, Maria Consuelo

Cotutor/a: García Molla, Marta

CURSO ACADÉMICO: 2023/2024

1. TABLA DE CONTENIDO

2.	Tabla de figuras	2
3.	Abstract	3
4.	Introducción	3
4.1	Métodos cualitativos en la investigación agraria	3
4.2	Grandes modelos lingüísticos	4
4.3	Objetivos	4
5.	Marco teórico.....	6
5.1	Desarrollo de LLMs	6
5.2	Limitaciones y problemas	7
5.3	Metodologías LLM	9
6.	Metodología.....	12
6.1	Elección de la metodología de ia	12
6.2	Preparación de datos.....	12
6.3	Análisis LLM	14
6.4	Análisis de datos	14
7.	Resultados	17
8.	Discusión	23
8.1	Interpretación de los resultados.....	23
8.2	Fuentes de error	25
8.3	Limitaciones del estudio	27
8.4	Implicaciones	28
8.5	Orientación futura de la investigación	29
9.	Conclusiones.....	31
10.	Bibliografía	33
11.	Anexos.....	37
	Anexo 1: Metodología de Jones (2014).....	37
	Anexo 2: <i>Prompt</i>	43
	Anexo 3: Ejemplo de respuesta GPT	49

2. TABLA DE FIGURAS

Figura 1 - Estructura de metodología.....	16
Figura 2 - Tasa de respuesta según tipo de participante. Elaboración propia.....	17
Figura 3- Concordancia simple de cada pregunta del prompt. Elaboración propia.	18
Figura 4 - Kappa de Cohen por cada pregunta del prompt. Elaboración propia.	19
Figura 5 – Matriz de confusión de la pregunta “person”. Elaboración propia.....	20
Figura 6 - Matriz de confusión para la pregunta "precio". Elaboración propia.	21
Figura 7 - Matriz de confusión de la pregunta "Why Used". Elaboración propia.	22
Figura 8 - Matriz de confusión de la pregunta "Agro in vil". Elaboración propia.	23
Figura 9 - parte de entrevista de Jones (2014).....	26
Figura 10. Modelo propuesto de sistema de investigación continua basado en metodologías con LLMs. Elaboración propia.	29

3. ABSTRACT

Este estudio explora la integración de Modelos de Lenguaje Grande (LLM), en metodologías cualitativas de investigación agrícola, centrándose en el análisis de transcripciones de entrevistas de un estudio previo que investiga los mercados de semillas y las prácticas agrícolas en África Occidental. Se compararon los resultados del análisis temático asistido por LLM con los de codificadores humanos utilizando métricas como el Kappa de Cohen. Los hallazgos revelan que, aunque los LLM pueden manejar grandes volúmenes de datos cualitativos de manera eficiente, su rendimiento varía según la complejidad de las tareas de codificación. Las preguntas binarias simples mostraron mayores niveles de acuerdo, mientras que las preguntas más complejas y dependientes del contexto mostraron menores puntuaciones de Kappa. El estudio concluye que, aunque los LLMs ofrecen mejoras prometedoras en eficiencia y profundidad del análisis cualitativo, su aplicación requiere ajustes metodológicos y refinamientos continuos para garantizar su fiabilidad y validez. Se sugiere que futuras investigaciones se enfoquen en desarrollar marcos integrados que combinen las capacidades de los LLM con la experiencia humana para mejorar la investigación cualitativa en la agricultura, apoyando así prácticas de desarrollo agrícola más efectivas y sostenibles.

4. INTRODUCCIÓN

4.1 MÉTODOS CUALITATIVOS EN LA INVESTIGACIÓN AGRARIA

Las metodologías cualitativas son esenciales en la investigación agraria y el desarrollo rural, proporcionando una visión profunda de las dinámicas socioeconómicas y culturales que los métodos cuantitativos no capturan. Estas metodologías ayudan a comprender las dimensiones humanas y sociales de las prácticas agrícolas, fundamentales para diseñar intervenciones eficaces y sostenibles (Prokopy, 2011). Métodos como entrevistas y grupos de discusión permiten entender las experiencias de agricultores y comunidades rurales, identificando necesidades y limitaciones específicas. Jules N. Pretty (1995) subraya la importancia de estos enfoques para un desarrollo agrícola sostenible.

La FAO destaca que los enfoques cualitativos son cruciales para desarrollar políticas y tecnologías adaptadas a pequeños agricultores que enfrentan desafíos como el cambio climático y la degradación de la tierra (Prokopy, 2011). La investigación cualitativa facilita la cocreación de conocimientos entre investigadores y comunidades locales, asegurando que las nuevas tecnologías sean culturalmente aceptables. Prokopy (2011) enfatiza que estos métodos son vitales para explorar motivaciones y barreras en la adopción de nuevas tecnologías.

Además, las metodologías cualitativas abordan eficazmente las dimensiones sociales y culturales del desarrollo rural, dando voz a grupos marginados y asegurando intervenciones equitativas (FIDA, 2023). Prokopy (2010) también señala que estas investigaciones descubren complejidades sociales y culturales que informan estrategias de desarrollo más efectivas. Los resultados de la investigación cualitativa son valiosos para la elaboración de políticas y prácticas, ayudando a los responsables a entender la realidad en el terreno. La FAO utiliza estos datos para desarrollar estrategias que mejoren la productividad agrícola y promuevan el crecimiento sostenible. Pretty (1995) afirma que este enfoque puede orientar políticas efectivas y sostenibles en comunidades rurales.

4.2 GRANDES MODELOS LINGÜÍSTICOS

La aparición de grandes modelos lingüísticos (LLM) como GPT-3 y GPT-4 ha generado atención en la comunidad investigadora, destacando tanto su inmenso potencial como los retos que plantean. Los investigadores han reconocido ampliamente las impresionantes capacidades de los LLM en diversos dominios, como el procesamiento del lenguaje natural, la traducción automática y la generación de contenidos (Vaswani et al., 2017; Brown et al., 2020). Por ejemplo, los sistemas multiagente, donde varios modelos de Inteligencia Artificial (IA) interactúan y se critican mutuamente, han demostrado ser prometedores para mejorar el razonamiento y la precisión de los hechos (Du et al., 2023).

En la investigación jurídica, herramientas como Casetext utilizan LLM para proporcionar análisis más rápidos y precisos, mientras que GitHub Copilot ayuda a los programadores generando y autocompletando código, demostrando las ventajas prácticas de integrar LLM en tareas específicas (Bano, Zowghi, and Whittle, 2023; Zhang et al., 2019). Sin embargo, hay una creciente conciencia sobre los retos asociados a los LLM, como las "alucinaciones", donde generan información plausible pero incorrecta, subrayando la necesidad de datos fiables para garantizar la exactitud de los resultados de la IA.

La comunidad investigadora también enfrenta implicaciones éticas y prácticas, abordando temas como sesgos inherentes, privacidad de datos y la posible propagación de información errónea (Weidinger et al., 2022). El futuro de los LLM implica perfeccionar sus capacidades técnicas y establecer marcos que aseguren resultados coherentes, reproducibles y válidos científicamente, combinando avance tecnológico y rigor metodológico (Patil and Gudivada, 2024).

4.3 OBJETIVOS

El objetivo principal de este estudio es explorar la integración de los grandes modelos de lenguaje (conocido como LLMs por sus siglas en inglés) en metodologías cualitativas para la investigación agrícola y el desarrollo rural. El estudio se estructura en torno a una revisión de la literatura, pruebas empíricas del Análisis de Contenido asistido por LLM (LACA) y una discusión de los resultados, centrándose en las aplicaciones potenciales y las direcciones futuras (Dai et al., 2023; Chew et al., 2023).

La primera fase del estudio consiste en realizar una revisión bibliográfica para evaluar las implementaciones actuales de LLMs en la investigación cualitativa agraria. Esta revisión pretende identificar los puntos fuertes y las limitaciones de los métodos existentes, en particular en el contexto del análisis temático y la codificación de contenidos. Las ideas clave de esta revisión informan el desarrollo de metodologías nuevas, sólidas y reproducibles para aplicar los LLM en la investigación agrícola.

A partir de los resultados de la revisión bibliográfica, la segunda fase consiste en probar empíricamente el Análisis de Contenido Asistido por LLM (LACA). Esta fase se centra en la utilización de LLMs como GPT-3.5 y GPT-4 para realizar análisis temáticos de datos cualitativos procedentes de estudios agrícolas. Los resultados de este proceso informarán una comparación sistemática de la precisión y la coherencia de la codificación de los LLMs con analistas humanos (Dai et al., 2023).

La fase final del estudio incluye una discusión de los resultados empíricos y una exploración de las aplicaciones prácticas de los LLM en la investigación agrícola. Pretende analizar las implicaciones prácticas del uso de los LLM para el análisis temático en los proyectos de investigación agrícola existentes, evaluar las ventajas y limitaciones detectadas durante la fase de pruebas empíricas, y explorar las aplicaciones y potencial futuro.

A través de estas fases estructuradas, el estudio pretende aplicar las nuevas metodologías de uso de LLMs a un contexto del sector agrario. Mediante la integración de los LLM en la investigación agrícola, el estudio pretende mejorar la capacidad de llevar a cabo investigación cualitativa en el sector agrícola, en una manera eficiente y equitativa.

5. MARCO TEÓRICO

5.1 DESARROLLO DE LLMs

5.1.1 Historia y evolución

Los modelos lingüísticos han evolucionado mucho en las últimas décadas, empezando por los modelos estadísticos del lenguaje (SLM) a principios de los años noventa. Estos modelos, como el de *n-grams*, se basaban en simples distribuciones de probabilidad para predecir la siguiente palabra de una secuencia a partir de las palabras anteriores (Chu et al., 2024). A principios de la década de 2000, empezaron a utilizar redes neuronales para predecir secuencias de palabras, ofreciendo una comprensión más matizada del contexto y la semántica que sus predecesores estadísticos (Chu et al. 2024, Patil and Gudivada, 2024).

La era actual de los grandes modelos lingüísticos, como la serie GPT de OpenAI, comenzó en serio con GPT-3. Estos modelos se caracterizan por su enorme escala, tanto por el tamaño de los conjuntos de datos en los que se entrenan como por el número de parámetros que contienen. Los LLM suponen un avance significativo con respecto a los modelos anteriores al integrar conjuntos de datos aún mayores y técnicas de entrenamiento más sofisticadas para producir modelos capaces de generar texto similar al humano y realizar una amplia gama de tareas basadas en el lenguaje con una precisión sin precedentes (Chu et al. 2024, Patil y Gudivada, 2024).

Makridakis et al. (2023) destacaron varios factores clave que han impulsado el rápido avance de los grandes modelos lingüísticos (LLM) como GPT-3 y GPT-4. La explosión de datos de texto digital en las dos últimas décadas ha proporcionado un corpus vasto y diverso para el entrenamiento de los LLM. La diversidad del conjunto de datos permite a estos modelos manejar una amplia gama de tareas, desde consultas sencillas hasta la generación de lenguajes complejos, lo que aumenta significativamente su versatilidad y robustez (Makridakis et al., 2023). Además, el crecimiento de la potencia computacional ha sido fundamental en el desarrollo de los LLM. Los avances en hardware especializado han permitido entrenar modelos cada vez más grandes con miles de millones de parámetros como el GPT-4. (Patil y Gudivada, 2024).

Además, las innovaciones en las técnicas de preentrenamiento y ajuste fino han impulsado aún más las capacidades de los LLM. Métodos como el aprendizaje en contexto y el aprendizaje por transferencia permiten a los modelos generalizar mejor y obtener buenos resultados en diversas tareas sin necesidad de ajustes específicos. Estas técnicas permiten a los LLM aprovechar eficientemente el conocimiento previo, mejorando su rendimiento en aplicaciones posteriores (Patil y Gudivada, 2024).

5.1.2 Modelos actuales

Hoy en día existen varios modelos de LLM de uso comercial como LLaMA de Meta, Gemini de Google, y Claude de Anthropic, pero ninguno ha llegado a tener tanta utilización como el GPT de OpenAI. Estos LLMs se han empleado en diversos contextos para abordar retos específicos de la investigación cualitativa, las recomendaciones agrícolas y los marcos educativos, entre otros (Vaswani et al., 2017; Brown et al., 2020; Touvron et al., 2023).

GPT-3 ha sido una opción popular debido a sus sólidas capacidades de comprensión y generación de lenguaje. Por ejemplo, en el estudio de Ibrahim et al. (2024), GPT-3 se utilizó para responder a las preguntas de los agricultores sobre el cultivo de arroz de regadío en Nigeria, demostrando una mejora significativa en la calidad de las respuestas en comparación con los agentes de extensión tradicionales. La versión actualizada, GPT-3.5, se ha destacado por sus capacidades mejoradas de comprensión y respuesta contextual. Chew et al. (2023) utilizaron GPT-3.5 en su Análisis de Contenido Asistido por LLM (LACA), donde facilitó una codificación deductiva más precisa en el análisis temático al interpretar datos cualitativos complejos. El proceso iterativo entre los analistas humanos y GPT-3.5 garantizó que los resultados temáticos no sólo fueran rápidos, sino también matizados y fiables.

Con el lanzamiento de GPT-4, los investigadores han observado nuevas mejoras en la precisión y una comprensión contextual más profunda, lo cual es crucial en campos que requieren una gran precisión, como la asesoría agrícola y la entrega de contenidos educativos (Brown et al. 2020). Silva et al. (2023) exploraron la aplicación de GPT-4 como asistente de agrónomo, donde se puso a prueba frente a exámenes de certificación agrícola, demostrando que el LLM no solo podía igualar el rendimiento de un experto humano, sino que en algunos casos lo superaba.

Estos usos subrayan la evolución y diversificación de las capacidades del LLM. Las primeras conclusiones sugieren que, si bien los modelos más antiguos, como GPT-3, son suficientes para generar respuestas genéricas, los modelos más recientes, como GPT-4, ofrecen avances significativos a la hora de tratar contenidos complejos, matizados o altamente especializados. A medida que la tecnología LLM siga evolucionando, será necesario seguir investigando para explorar todo el potencial de estos modelos en aplicaciones académicas y prácticas, garantizando que se utilicen de la forma más eficaz para complementar la experiencia humana y mejorar los procesos de toma de decisiones en diversos ámbitos.

5.2 LIMITACIONES Y PROBLEMAS

Aunque los LLM ofrecen un potencial transformador en diversos ámbitos, estas limitaciones y restricciones ponen de manifiesto la necesidad de una investigación continua, una aplicación reflexiva y una evaluación rigurosa para garantizar que estas herramientas se utilicen de forma ética, eficaz e inclusiva. Abordar estos retos será crucial para aprovechar todo el potencial de los LLM en la investigación y las aplicaciones prácticas.

5.2.1 Alucinación

Las alucinaciones, o la generación de respuestas objetivamente incorrectas o sin sentido por parte de los LLM, representan un reto importante. Estos errores pueden socavar la fiabilidad de los modelos, especialmente en aplicaciones críticas como el asesoramiento médico o la orientación agrícola, donde la precisión es primordial (Bano, Zowghi, and Whittle, 2023). Por ejemplo, Ibrahim et al. (2024) observaron que, aunque ChatGPT proporcionaba a menudo respuestas de alta calidad, en ocasiones generaba información incorrecta sobre prácticas agrícolas, lo que podía inducir a error a los usuarios y reducir la confianza en los sistemas de IA.

Tales imprecisiones pueden comprometer gravemente la validez de los resultados de la investigación e inducir a error en posteriores análisis y decisiones políticas.

5.2.2 Sesgos

Los sesgos en los LLM surgen de los datos con los que se entrenan estos modelos. Dado que los LLM aprenden a partir de vastos corpus de texto de Internet, pueden adquirir y perpetuar sesgos presentes en sus datos de entrenamiento. Esto plantea retos, sobre todo en aplicaciones sensibles como el análisis temático o la orientación en agricultura basada en IA, donde es crucial una representación imparcial y justa (Weidinger et al., 2022). Los estudios de Sowole et al. (2023) y otros han explorado las implicaciones de estos sesgos, haciendo hincapié en la necesidad de una monitorización continua y un reentrenamiento del modelo para mitigar sus efectos.

5.2.3 Accesibilidad

Los problemas de accesibilidad se derivan de la distribución desigual de los recursos tecnológicos, que puede impedir la adopción generalizada de las herramientas de IA. Garantizar que los LLM sean accesibles a diversas poblaciones de distintos entornos geográficos y socioeconómicos es esencial para aprovechar sus beneficios potenciales de forma universal. La mayoría de los modelos actuales de LLM no son de uso libre, sino tecnologías patentadas de grandes empresas tecnológicas como Google, Meta y OpenAI. El acceso puede estar restringido, ya sea a miembros de pago o a grandes clientes corporativos. Las consideraciones de coste también desempeñan un papel crucial, especialmente para desplegar los LLM en entornos con recursos limitados. Los recursos informáticos necesarios para ejecutar estos modelos son considerables, lo que puede limitar su accesibilidad a las instituciones más pequeñas o a los países en desarrollo. (Patil and Gudivada, 2024).

5.2.4 Limitaciones técnicas

Las restricciones de tamaño de los tokens son una limitación técnica que afecta a la profundidad y amplitud de los análisis que los LLM pueden realizar en un solo mensaje. Debido al tamaño limitado de los *prompts* de modelos como GPT-3, sólo se puede procesar una cantidad limitada de texto a la vez. Esta limitación obliga a segmentar los textos más grandes en partes más pequeñas para su análisis, lo que puede fragmentar la coherencia contextual o temática y complicar las tareas de análisis exhaustivo.

Los requisitos de diseño de los *prompts* son fundamentales para aprovechar eficazmente las capacidades del LLM. La forma en que se estructura una pregunta puede influir significativamente en la calidad y relevancia de los resultados del LLM. Un diseño eficaz de las instrucciones requiere un profundo conocimiento del funcionamiento del modelo y de las necesidades específicas del dominio para obtener respuestas precisas y útiles. Como se señala en los estudios de Dai et al. (2023), las instrucciones mal diseñadas pueden conducir a resultados menos precisos o irrelevantes, lo que subraya la necesidad de una entrada cuidadosamente seleccionada para maximizar el rendimiento del modelo.

5.2.5 Ética

El uso de LLM plantea dudas sobre la privacidad de los datos y el consentimiento. Dado que estos modelos se entrenan a partir de grandes cantidades de datos, existe el riesgo de que se incluya inadvertidamente información personal o sensible en los conjuntos de datos de entrenamiento, lo que podría suponer una violación de la privacidad. Garantizar que todos los datos utilizados en el entrenamiento y el posterior análisis de los LLM se anonimizan y se obtienen con el consentimiento adecuado es esencial para abordar estas cuestiones éticas (arXiv, 2023).

5.3 METODOLOGÍAS LLM

5.3.1 Análisis de contenidos asistido por LLMs

El análisis temático, un método de investigación cualitativa tradicionalmente basado en procesos manuales, está evolucionando mediante la integración LLMs, que ofrecen mejoras prometedoras en términos de eficiencia y profundidad. El uso metódico de los LLM en el análisis temático se ilustra en estudios como los de Dai et al. (2023) y Chew et al. (2023), que exploran formas innovadoras de combinar las habilidades analíticas humanas con las capacidades de la IA. Dai et al. (2023) introdujeron un enfoque "LLM-in-the-loop" en el que codificadores humanos y LLM colaboran en la interpretación de texto y el refinamiento de los códigos iniciales. Este proceso de colaboración no solo acelera el análisis, sino que también garantiza la fiabilidad de los temas identificados, aprovechando la capacidad del LLM para procesar grandes cantidades de datos rápidamente y, al mismo tiempo, aprovechando la visión humana para una comprensión matizada.

De forma similar, Chew et al. (2023) implementaron el Análisis de Contenido Asistido por LLM (LACA) centrándose en la codificación deductiva dentro del análisis temático. Su enfoque integra los LLM en la fase inicial de codificación, con la posterior intervención humana para refinar los códigos generados por la IA, garantizando que los constructos temáticos finales sean exhaustivos y reflejen con precisión los datos subyacentes. Estas metodologías ponen de relieve el aumento de eficiencia que supone el uso de los LLM, que pueden reducir significativamente los aspectos del análisis temático que requieren mucho trabajo.

Las primeras conclusiones de estos enfoques integradores sugieren que, aunque los LLM pueden mejorar sustancialmente el alcance y la eficacia del análisis temático, la interpretación matizada y la validación final de los temas requieren supervisión humana. La investigación futura debería centrarse en seguir perfeccionando estos modelos colaborativos, mejorar la interpretabilidad de los resultados del LLM y explorar su aplicabilidad en diversos contextos de investigación. Estos estudios subrayan el potencial transformador de los LLM a la hora de redefinir las metodologías cualitativas tradicionales, ofreciendo un enfoque más eficiente y exhaustivo del análisis temático (Dai et al., 2023; Chew et al., 2023).

2.3.2 *Fine Tuning*

El "*Fine Tuning*" o ajuste fino es un método crucial en el desarrollo de LLMs que mejora el rendimiento del modelo en tareas específicas ajustando los parámetros del modelo original a un conjunto de datos concreto. Esta técnica ha sido ampliamente adoptada para adaptar los LLM generales a dominios especializados o a necesidades específicas de los usuarios, haciéndolos más

eficaces en aplicaciones en las que el contexto y el conocimiento detallado son esenciales (Patil y Gudivada, 2024, Silva et al., 2023).

En el ámbito de las aplicaciones agrícolas, el estudio de Silva et al. (2023) muestra el impacto del ajuste fino para permitir que los LLM funcionen como asistentes de agrónomos. Al ajustar GPT-4 en un corpus de textos agrícolas, preguntas y respuestas de expertos, los investigadores pudieron mejorar significativamente la capacidad del modelo para proporcionar consejos precisos y contextualmente relevantes para las prácticas agrícolas. Este enfoque permitió al LLM no sólo entender el lenguaje técnico de la agronomía, sino también generar respuestas prácticas y aplicables a condiciones agrícolas concretas.

La aplicación del ajuste fino en este contexto demuestra cómo los LLM pueden adaptarse eficazmente para ofrecer orientación de nivel experto en campos que tradicionalmente dependen de expertos humanos. Las respuestas personalizadas generadas por estos modelos afinados se muestran capaces de respaldar decisiones que requieren altos niveles de especialización, como la gestión de cultivos y el control de plagas, lo que subraya el potencial de la IA en sectores como la agricultura (Silva et al. 2023, Sowole et al. 2023, Ibrahim et al., 2024).

5.3.2 *Embedding Based Retrieval*

La *embedding based retrieval (EBR)* es una sofisticada técnica que se utiliza junto con los grandes modelos lingüísticos (LLM) para mejorar la eficacia y precisión de la recuperación de información a partir de grandes conjuntos de datos. Este método consiste en crear representaciones vectoriales de los datos que permiten recuperar rápidamente la información pertinente basándose en la similitud semántica, en lugar de depender únicamente de la coincidencia de palabras clave. El EBR es especialmente eficaz en aplicaciones en las que la precisión de la información recuperada es fundamental para la tarea en cuestión, como en los sistemas automatizados de respuesta a preguntas o la generación de contenidos detallados (Peng et al., 2023, Silva et al., 2023).

El estudio de Peng et al. (2023) es un excelente ejemplo de recuperación basada en la incrustación aplicada en un contexto agrícola, que muestra la utilidad del método para extraer y estructurar información agrícola a partir de fuentes de datos no estructuradas. En su trabajo, Peng et al. (2023) implementaron el sistema FINDER, que integra LLM con EBR para identificar, recuperar y organizar sistemáticamente datos agrícolas en un formato estructurado. Este sistema utiliza EBR para segmentar primero los documentos en piezas más pequeñas y manejables que luego se vectorizan y almacenan en una base de datos. Cuando se realiza una consulta, el sistema recupera los segmentos de documentos más relevantes en función de su similitud vectorial con la consulta, lo que garantiza que la respuesta generada por el LLM se base en la información más pertinente.

5.3.3 *Técnicas de prompting*

En el campo de la IA, las técnicas de prompting se han convertido en un área de interés fundamental, sobre todo para mejorar el rendimiento de los LLM en tareas específicas sin necesidad de un reentrenamiento o ajuste exhaustivos. (Dai et al., 2023).

La técnica de “cadena de pensamiento” consiste en construir instrucciones que guíen al modelo a través de una secuencia lógica de pasos de pensamiento antes de llegar a una respuesta. Silva et al. (2023) exploraron este método en entornos agrícolas, incitando a los LLM a considerar diversos factores como el tipo de suelo, la estacionalidad de los cultivos y los métodos de control de plagas en secuencia para generar prácticas de gestión integrales para cultivos específicos.

Estas técnicas de *prompting* representan poderosas herramientas de la IA que permiten a investigadores y profesionales aprovechar los LLM preentrenados de forma más flexible y creativa en diversos ámbitos. Ya sea guiando al modelo a través de un proceso de razonamiento lógico, preparando el terreno con ejemplos u operando sin pistas previas, estos métodos mejoran la utilidad y adaptabilidad del LLM. Las primeras aplicaciones en campos como el análisis temático y la agricultura han arrojado resultados prometedores, lo que indica que estas técnicas no sólo mejoran la eficiencia de los LLM, sino también su eficacia a la hora de generar respuestas matizadas y adecuadas al contexto (Dai et al., 2023, Silva et al., 2023).

5.3.4 *Retrieval Augmented Generation (RAG)*

La *Retrieval Augmented Generation (RAG)* es una aplicación avanzada de los modelos lingüísticos amplios (LLM) que combina las capacidades generativas de los LLM con los sistemas de búsqueda de información para mejorar la precisión y pertinencia de los contenidos generados. Este enfoque ha tenido especial repercusión en campos en los que la información precisa y específica del contexto es crucial, como la agricultura (Silva et al. 2023).

Balaguer et al. (2024) exploraron la eficacia de la RAG comparándola con los enfoques tradicionales de ajuste fino en el contexto de la agricultura, centrándose en cómo influyen estas metodologías en el rendimiento de los LLM a la hora de generar respuestas a consultas específicas del dominio. El método RAG implica la recuperación dinámica de información relevante de un conjunto de datos durante el proceso de generación, lo que permite al LLM producir respuestas que no sólo son contextualmente apropiadas, sino que también están profundamente informadas por el conocimiento existente. Esto es especialmente beneficioso en agricultura, donde el asesoramiento debe ser científicamente preciso y adaptarse a las condiciones locales. El estudio subraya el potencial de la GAR para mejorar la calidad de la información proporcionada por los sistemas de IA, lo que la convierte en una herramienta valiosa para las industrias que buscan implantar soluciones de IA que dependen de grandes volúmenes de conocimientos especializados.

5.3.5 *Ingeniero Agrónomo Artificial*

La funcionalidad interactiva de LLMs como ChatGPT ha fomentado la investigación en la creación de asistentes agrícolas artificiales que sirven para asesorar a agricultores y dar consejos y información científica. Esta integración pretende hacer frente a diversos retos, como la escasez de servicios humanos de extensión agrícola y la creciente complejidad de la gestión de las explotaciones modernas. Estudios como los de Sowole et al. (2023), Ibrahim et al. (2024) y Silva et al. (2023) han explorado la aplicación y eficacia de la IA para proporcionar asesoramiento y soluciones agrícolas críticas a los agricultores, especialmente en zonas con pocos recursos.

Sowole et al. (2023) implementaron un modelo en Nigeria que aprovechaba la GPT-3.5 de OpenAI para actuar como agente de extensión agrícola virtual. Este sistema de IA se perfeccionó con un amplio conjunto de datos adaptados a las prácticas agrícolas locales y se integró a través

de un interfaz de programación de aplicaciones (API) que los agricultores podían consultar para recibir asesoramiento sobre diversas cuestiones agrícolas. La capacidad del sistema para ofrecer respuestas accesibles e inmediatas a los agricultores demostró su potencial para mejorar significativamente la productividad agrícola y la toma de decisiones.

Para analizar la eficacia de este método, Ibrahim et al. (2024) realizaron un estudio para evaluar la calidad de las respuestas proporcionadas por ChatGPT a las preguntas de los agricultores sobre el cultivo de arroz de regadío en Nigeria, comparándolas con las dadas por agentes de extensión humanos. Los resultados mostraron que ChatGPT proporcionaba a menudo respuestas de mayor calidad, gracias a su completa base de datos y a su capacidad para generar respuestas detalladas. Sin embargo, el estudio también detectó áreas en las que los agentes humanos superaban a la IA, sobre todo en conocimientos específicos locales como los tiempos de siembra y las dosis de semillas, lo que subraya la importancia de integrar los conocimientos locales en los modelos de IA.

Además, Silva et al. (2023) ampliaron el alcance de los asistentes de IA en agricultura probando si GPT-4 podía funcionar eficazmente como asistente de agrónomo. Su enfoque utilizó otras metodologías como RAG y EBR para responder a preguntas de exámenes de certificación agrícola y proporcionar directrices de gestión de cultivos, demostrando la capacidad del modelo para dar soporte a tareas de conocimiento agrícola más complejas.

En conjunto, estos estudios ilustran el prometedor potencial de los asistentes agrícolas con IA para mejorar la accesibilidad y la calidad de los servicios de extensión agraria. También ponen de relieve la necesidad de que estos sistemas se actualicen continuamente y se integren con los conocimientos de los expertos locales para garantizar su pertinencia y eficacia. La integración de la IA en la agricultura está llamada a transformar las prácticas agrícolas tradicionales proporcionando un apoyo escalable, eficiente y preciso a los agricultores de todo el mundo, mejorando así la productividad y la sostenibilidad del sector agrícola.

6. METODOLOGÍA

6.1 ELECCIÓN DE LA METODOLOGÍA DE IA

Aunque todas las metodologías mencionadas anteriormente demuestran mucha utilidad, este estudio se enfoca en LACA, basado principalmente en la investigación de Dai et al. (2023) and Chew et al. (2023). Dado el limitado alcance de este proyecto y la gran cantidad de datos cualitativos necesarios para utilizar cualquiera de estas herramientas de evaluación, se optó por realizar el análisis a partir de un conjunto de datos preexistente de transcripciones de entrevistas (Jones, 2014).

6.2 PREPARACIÓN DE DATOS

Se encontró una tesis de máster en el *Qualitative Data Repository* (<https://qdr.syr.edu/>). La tesis de Kristal Jones, titulada "*Emerging Seed Markets, Substantive Seed Economies and Integrated Seed Systems in West Africa: A Mixed Methods Analysis*", explora la compleja

interacción entre el desarrollo agrícola orientado al mercado y los sistemas agrícolas tradicionales en la región saheliana de África Occidental, centrándose especialmente en Malí, Burkina Faso y Níger (Jones, 2014). Esta tesis utilizó una metodología de entrevistas semiestructuradas con agricultores, abordando temas relacionadas a la compraventa y utilización de semillas.

Se extrajo el texto de 496 transcripciones de entrevistas en PDF y se formateó para su procesamiento en LLM como archivos de texto sin formato, garantizando que se preservaba toda la integridad contextual de los datos.

Se seleccionaron 15 campos de datos del libro de códigos que Jones utilizó para llevar a cabo su análisis temático de las entrevistas. Estos campos fueron seleccionados por su carácter analítico, requiriendo una comprensión más matizada de las entrevistas para evaluar con éxito. Los campos consisten en las respuestas de las preguntas de entrevista, resumido en forma de un número que sirve como código para categorizar las respuestas. Los campos elegidos son:

<i><u>Anim in Vil</u></i>	Hay un <i>animateur</i> (agrónomo) en el pueblo?
<i>Agro in Vil</i>	Hay un agrocomerciante en el pueblo?
<i>Initial Info</i>	Dónde se encontró la información inicial sobre variedades modernas?
<i>Yrs saved</i>	Durante cuántos años han guardado y reutilizado semillas?
<i>Sharing</i>	El agricultor comparte semillas?
<i>Access Place</i>	Dónde el agricultor pudo acceder a semillas de variedades modernas?
<i>Price</i>	Precio pagado por semillas
<i>Person</i>	Persona que abasteció las semillas
<i>Ident Var</i>	El agricultor es capaz de identificar las variedades usadas?
<i>Packaging</i>	Con qué tipo de empaquetado llegaron las semillas?
<i>Saving</i>	El agricultor guarda semillas?
<i>Why Used</i>	Por qué eligió usar variedades modernas?
<i>Saw MVs</i>	El agricultor ha visto variedades modernas en el campo antes de usarlas?
<i>Union</i>	El agricultor pertenece a un sindicato?

Más información sobre la metodología de Jones (2014) está colocada en *Anexo 1*, junto con un ejemplo de una transcripción de entrevista.

6.3 ANÁLISIS LLM

6.3.1 LLMs utilizados

El proceso de análisis se llevó a cabo dos veces para considerar la diferencia de resultados entre dos LLM, GPT-3.5 y GPT-4, ambos de OpenAI. Se eligieron estos dos modelos por su facilidad de acceso, la facilidad de integración de la API para la automatización del proceso de análisis y su uso en estudios previos sobre el tema (Dai et al. 2023, Chew et al. 2023).

6.3.2 *Prompt*

Se escribió un programa en Python para interactuar con la API de OpenAI, lo que permitió el procesamiento masivo de los archivos de texto de las entrevistas. Se proporcionó una versión reformateada del libro de códigos de Jones como *prompt* del sistema, mientras que el texto de la entrevista se proporcionó como *prompt* del usuario.

Basándose en Silva (2023), el *prompt* se diseñó para emplear un razonamiento de cadena de pensamiento, en el que se pedía a la respuesta que, antes de seleccionar el código, mostrara cuál era el significado de ese código específico basándose en el libro de códigos y, a continuación, pusiera también una línea explicando su razonamiento para elegir el código. Un ejemplo del output del LLM está disponible en el anexo 2.

La temperatura se fijó en 0 para minimizar creatividad en el modelo (Dai et al, 2023).

El texto completo del *prompt* está ubicado en el anexo 3.

6.3.3 Extracción de datos

Para simular el proceso de análisis temático manual en el que varios codificadores evalúan el mismo texto y se ponen de acuerdo, y para minimizar el impacto de las alucinaciones en el conjunto de datos, cada entrevista fue procesada tres veces por el LLM y, a continuación, se eligió la respuesta más común de las tres muestras. Si no había respuesta más común, se consideraba que la respuesta final se había dejado en blanco. Esto pretende ser una versión simplificada de la metodología de Tai et al. (2024), en la cual los mismos textos fueron procesados por un LLM hasta 160 veces para ver las diferencias entre los resultados. Debido a los rendimientos decrecientes del procesamiento repetido demostrado por Tai et al. y motivos presupuestarios, se eligió usar solo tres muestras.

6.4 ANÁLISIS DE DATOS

6.4.1 Tasa de clasificación

Dado que el *prompt* no solicita que el modelo proporcione una respuesta el 100% de las veces, sino que deja la oportunidad de dejar la respuesta en blanco para el código dado si el modelo considera que no hay suficiente información basada en la transcripción de la entrevista para poder proporcionar un código preciso. Por ello, el índice de respuesta de ambos modelos se calculó junto con el índice de respuesta de los codificadores humanos, como un simple porcentaje del número de respuestas válidas dividido por el número total de preguntas.

6.4.2 Kappa de Cohen

La Kappa de Cohen (κ) es una medida estadística utilizada para evaluar el nivel de acuerdo entre dos o más evaluadores que clasifican ítems en categorías mutuamente excluyentes (Cohen, 1960). Resulta especialmente útil para evaluar la fiabilidad entre evaluadores cuando se codifican datos cualitativos, ya que tiene en cuenta el acuerdo que se produce por azar. En este estudio, se emplea el Kappa de Cohen para comparar la precisión y coherencia de la codificación entre analistas humanos y grandes modelos lingüísticos (LLM), como GPT-3.5 y GPT-4.

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

donde:

Po es el acuerdo observado entre los evaluadores y Pe es el acuerdo esperado por azar.

En este estudio, se utiliza Kappa de Cohen para evaluar la fiabilidad de la codificación temática realizada por los LLM en comparación con los codificadores humanos, similar a la metodología de Tai et al (2024) y Xiao et al (2023). Calculando κ para diferentes conjuntos de datos codificados tanto por LLMs como por humanos, podemos cuantificar el nivel de acuerdo y determinar hasta qué punto los LLMs replican los estándares de codificación humana. Un valor κ alto indica una gran concordancia, lo que sugiere que los LLM son eficaces en el análisis temático, mientras que un valor κ bajo indicaría la necesidad de perfeccionar los modelos o el proceso de codificación (Cohen, 1960).

Figura 1 sirve como resumen de toda la metodología implementada en este estudio.

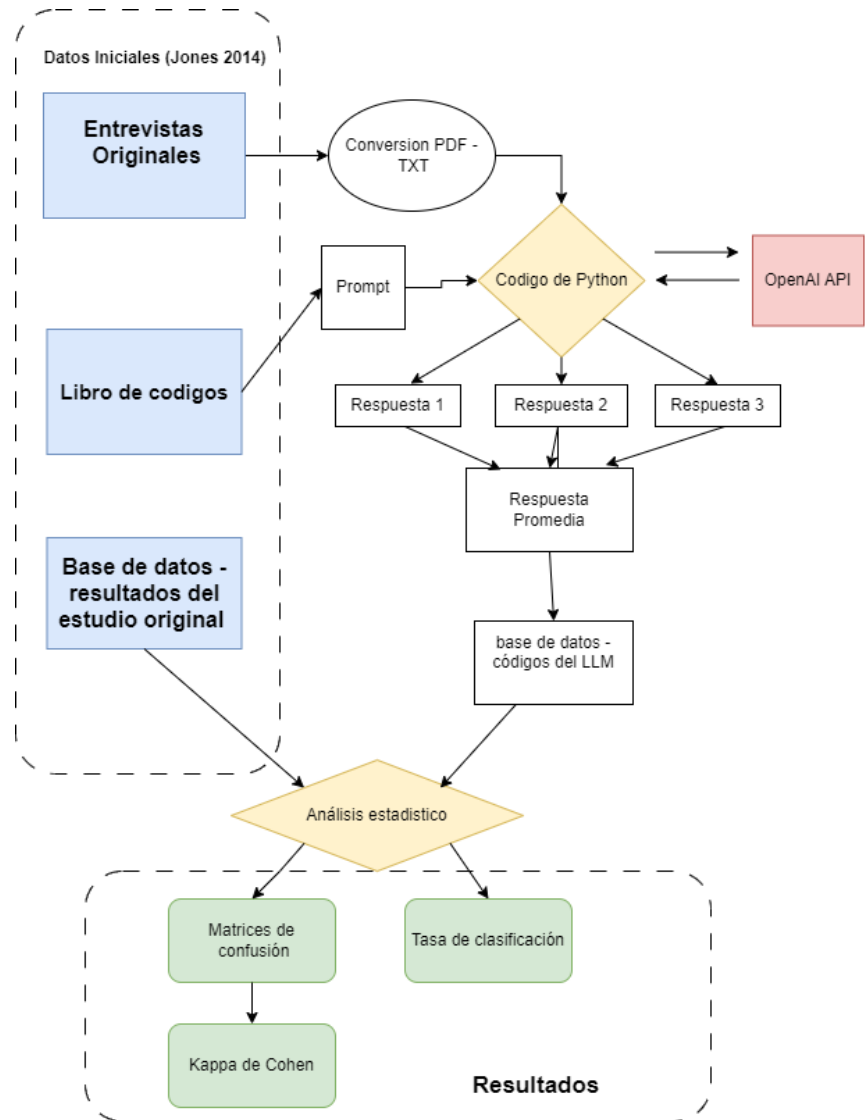


Figura 1 - Estructura de metodología. Elaboración propia.

7. RESULTADOS

7.1.1 Tasa de clasificación

La Figura 2 muestra la tasa de clasificación de a las preguntas de la entrevista seleccionadas de los codificadores humanos en el estudio de Jones (2014), contrastada con la proporción de respuestas proporcionadas por GPT-4 y GPT-3.5. La tasa de respuesta humana es relativamente alta, pero no alcanza el 100%, lo que indica que hay entrevistas que no fueron capaces de cubrir la variedad de preguntas de investigación planteadas por Jones. El modelo GPT-4 respondió a una tasa comparativamente alta, aproximándose a la de los evaluadores humanos. En cambio, el modelo GPT-3.5 se mostró mucho más indeciso a la hora de dar respuestas a las preguntas, respondiendo más a menudo que la información era insuficiente para poder proporcionar un código.

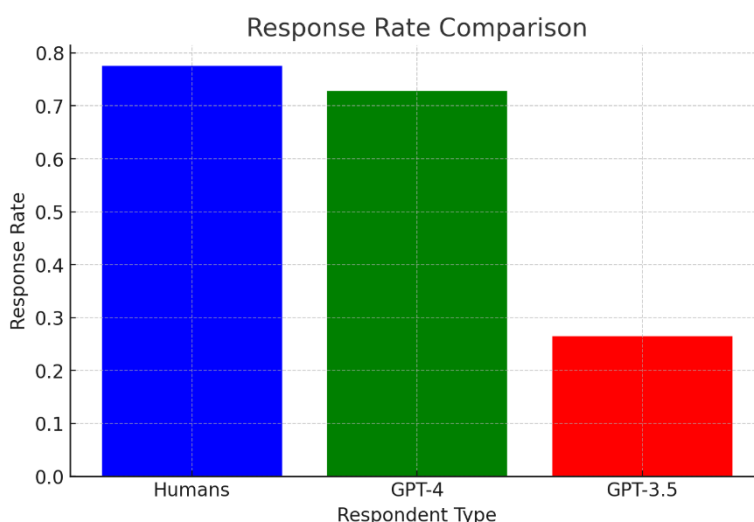


Figura 2 - Tasa de respuesta según tipo de participante. Elaboración propia.

7.1.2 Concordancia Simple

La Figura 3 muestra la tasa de concordancia simple entre codificadores humanos y codificadores LLM para cada una de las preguntas de investigación seleccionadas en el cuestionario, separadas por las dos opciones LLM evaluadas en este estudio. Las barras más claras indican la tasa de concordancia simple cuando se filtran las respuestas "sin respuesta", de modo que sólo se evalúan las respuestas en las que el LLM se sentía lo suficientemente "seguro" para responder. Los mayores índices de acuerdo simple se dieron en las preguntas "union", en la que se preguntaba si el entrevistado pertenecía o no a un sindicato o cooperativa de agricultores, y "saw MV", en la que se preguntaba si el agricultor había visto antes variedades modernas de semillas. Ambas preguntas se aproximan o superan el 80% de concordancia simple con las respuestas humanas en los conjuntos de datos filtrados. Las preguntas "Anim in vil" y "Agro in vil", relativas a

la presencia de un agrónomo o de un vendedor de semillas, respectivamente, tuvieron la concordancia simple más baja sin filtrar, pero ambas recibieron un gran aumento del proceso de filtrado, lo que sugiere que la principal fuente de error en ambas categorías procedía de preguntas en las que los LLM no pudieron dar una respuesta a partir del texto de la entrevista. En la mayoría de las preguntas, la GPT-4 demostró una mayor tasa de acuerdo simple que la GPT-3.5 en el conjunto de datos sin filtrar, mientras que en el conjunto de datos filtrado hubo más variabilidad entre ambas.

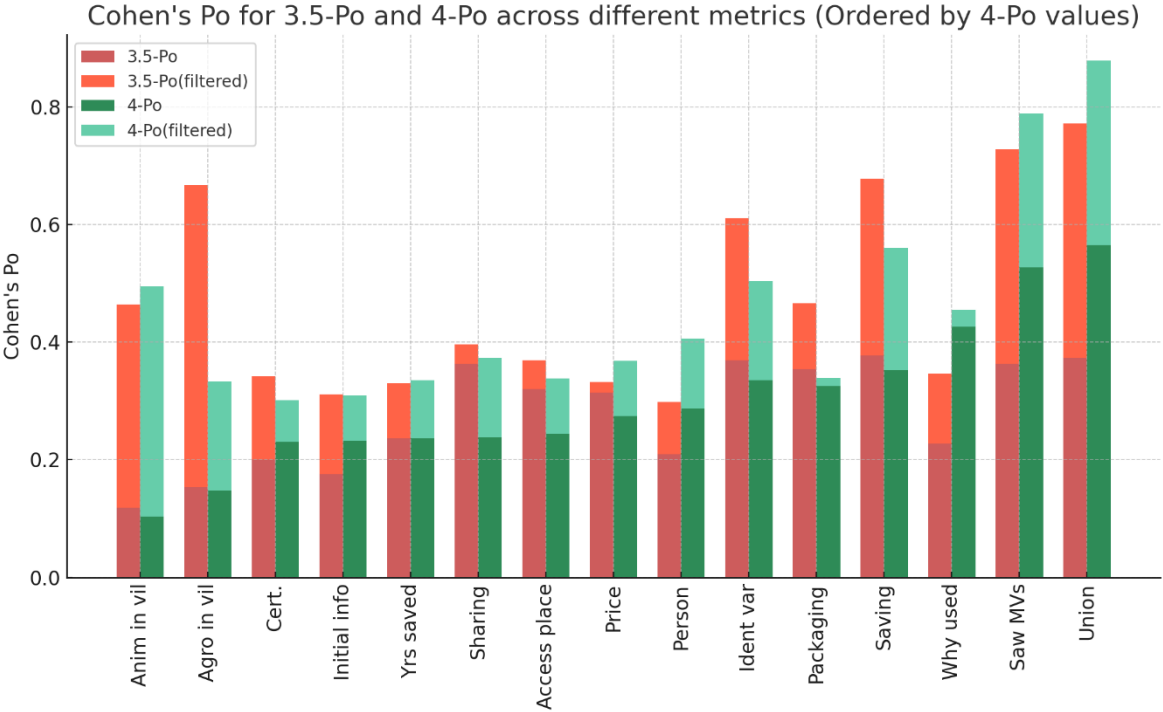


Figura 3- Concordancia simple de cada pregunta del prompt. Elaboración propia.

Kappa de Cohen

La figura 4 muestra el Kappa de Cohen para las respuestas de ambos modelos a cada una de las preguntas de la investigación en comparación con las respuestas humanas. Al igual que en la figura 3, las barras de color más claro indican el Kappa de Cohen cuando se eliminan los datos "sin respuesta". En comparación con los datos de concordancia simple, el estadístico Kappa de Cohen es mucho más bajo, lo cual es lógico si se tiene en cuenta que los factores estáticos de concordancia esperada entre los dos evaluadores se basan en el azar. La mayoría de las preguntas tienen un Kappa de Cohen inferior a 0,1, ya sea filtrado o sin filtrar, y dado que un Kappa de Cohen de 0 indica que las respuestas no son diferentes de lo que cabría esperar del azar, esto demuestra un rendimiento muy pobre de los LLM. Al igual que los estadísticos de concordancia simple, dentro

de los datos sin filtrar, GPT-4 obtuvo mejores resultados que GPT-3.5 en casi todas las categorías.

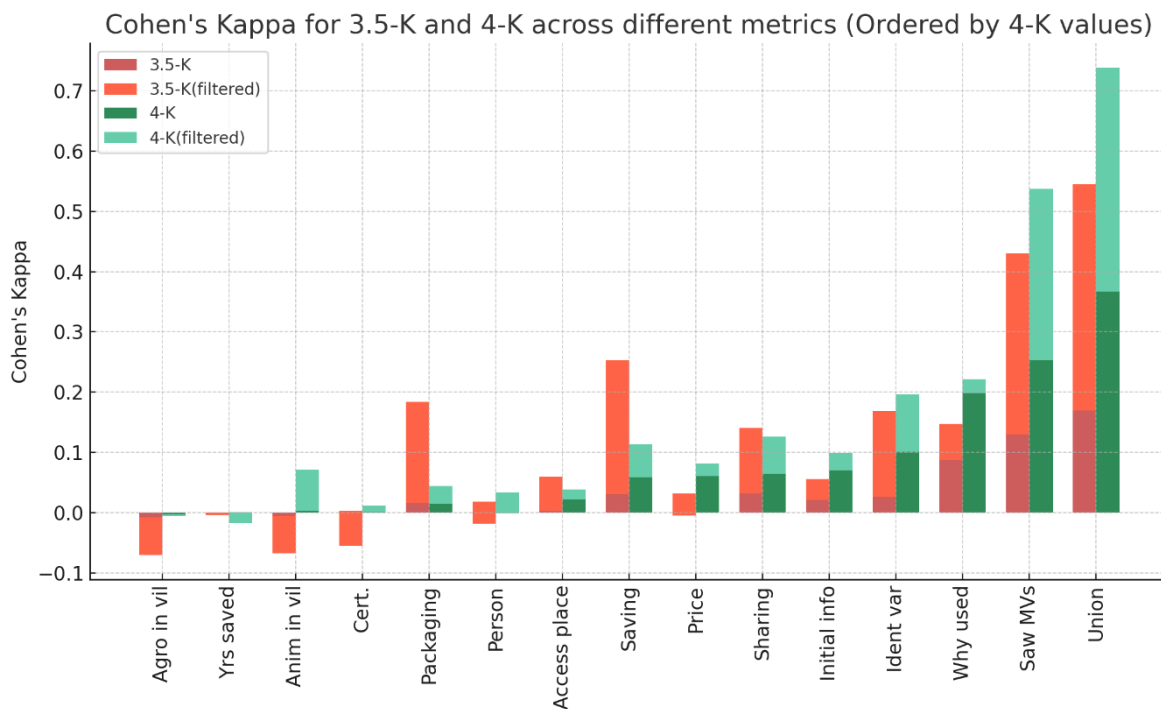


Figura 4 - Kappa de Cohen por cada pregunta del prompt. Elaboración propia.

Sin embargo, hay dos preguntas en las que los LLM obtuvieron puntuaciones Kappa relativamente altas, especialmente en los conjuntos de datos filtrados: "union", en la que se preguntaba si el entrevistado pertenecía o no a un sindicato o cooperativa de agricultores, y "saw MV", en la que se preguntaba si el agricultor había visto antes variedades modernas de semillas. En estos dos casos, los valores kappa alcanzan rangos de concordancia significativa entre humanos y LLM, especialmente en los conjuntos de datos filtrados.

7.1.3 Matrices de confusión

En las figuras 5 a 9 se muestran las matrices de confusión utilizadas para calcular el estadístico Kappa de Cohen de cuatro de las preguntas de investigación con peores resultados, como se indica en sus respectivos títulos. Cada celda indica el número de entrevistas que se evaluaron basándose en las respuestas del GPT-4 en las columnas y de los analistas humanos en las filas, y están coloreadas para ofrecer una mejor indicación de la cantidad a simple vista. Estas matrices se utilizan más adelante en la discusión para matizar las fuentes de desacuerdo y comprender mejor qué errores cometió el LLM al analizar las entrevistas.

En la figura 5 se nota una tasa bastante alta de no respuesta para los humanos, mayor de la del los LLM.

En la figura 6 también se nota una tasa de no respuesta mucho más alta para los humanos que los LLM. Otro resultado curioso es que los humanos casi nunca eligieron uno de los códigos, a pesar

"Person"- ¿Quién es la persona que proporcionó semillas de variedades modernas?

Respuestas - GPT-4

	No respuesta	Agrocomerciante	Autoabastecido	Productor de semillas	Otro agricultor	No utilizó variedades modernas
No respuesta	25	98	14	4	28	4
Agrocomerciante	14	108	3	4	18	1
Autoabastecido	7	73	2	4	12	2
Productor de semillas	1	0	0	0	0	0
Otro agricultor	0	6	0	1	1	0
No utilizó variedades modernas	6	31	3	0	4	0

Figura 5 – Matriz de confusión de la pregunta "person". Elaboración propia.

de haberlo incluido en el diseño inicial

"Price" - ¿Cómo el agricultor pagó por semillas de variedades modernas?

Respuestas - GPT-4

Respuestas - Humanos

	No Respuesta	Monetario, estandarizado (venta formal)	Autoabastecido - sin coste	Monetario, no estandarizado (venta informal)	No monetario (intercambio por otros bienes)	No utilizó variedades modernas
No Respuesta	36	88	17	23	6	3
Monetario, estandarizado (venta formal)	18	78	2	28	0	3
Autoabastecido - sin coste	13	52	12	22	0	1
Monetario, no estandarizado (venta informal)	0	0	0	1	0	0
No monetario (intercambio por otros bienes)	6	9	8	1	3	0
No utilizó variedades modernas	9	22	3	10	0	0

Figura 6 - Matriz de confusión para la pregunta "precio". Elaboración propia.

Por otro lado, hay una tasa casi zero de no respuesta en los humanos en la figura 7.

En el código "para probar y aprender" hay una tasa de acuerdo muy alta entre los humanos y los LLM.

No hay ninguna entrevista analizada por humanos que diga que la falta de lluvia influye su decisión, pero sí que hay algunas evaluadas así por el GPT-4.

"Why Used" - ¿Por qué el agricultor decidió probar variedades modernas?

Respuestas - GPT-4

Respuestas - Humanos		No respuesta	Maduración temprana	Falta de lluvia	Para probar y aprender	Mejorar la cosecha	Experiencia previa
	No respuesta	0	1	0	0	0	0
	Maduración temprana	2	29	5	8	5	0
	Falta de lluvia	0	0	0	0	0	0
	Para probar y aprender	19	19	9	148	60	4
	Mejorar la cosecha	2	1	0	8	17	1
	Experiencia previa	6	14	5	48	55	8

Figura 7 - Matriz de confusión de la pregunta "Why Used". Elaboración propia.

La figura 8 también demuestra una gran diferencia entre las interpretaciones de los LLM y los humanos. En general, parece que, si no ha mencionado un agrocomerciante, los humanos daban una respuesta de No, mientras GPT-4 no contestaba.

"Agro in vil" - ¿Hay un agrocomerciante en el pueblo?

		Respuestas - GPT-4		
		No respuesta	Sí	No
Respuestas - Humanos	No respuesta	0	0	0
	Sí	13	5	3
	No	251	137	65

Figura 8 - Matriz de confusión de la pregunta "Agro in vil". Elaboración propia.

8. DISCUSIÓN

8.1 INTERPRETACIÓN DE LOS RESULTADOS

8.1.1 Tasa de clasificación

La tasa de respuesta de la GPT-3.5, tal y como se observa en la figura 2, fue muy baja (27%), lo que sugiere que en la gran mayoría de los casos no fue capaz de obtener suficiente información para dar una respuesta y, por tanto, parece inadecuada para este tipo de aplicación. En cambio, la tasa de respuesta de GPT-4 se aproxima a la de los evaluadores humanos, lo que sugiere que recibía información suficiente para, como mínimo, aventurar la respuesta correcta. Cabe mencionar que el porcentaje de respuestas humanas no fue del 100%, lo que indica que la metodología de entrevistas de la investigación de Jones no fue capaz de responder a todas las preguntas de la investigación (Jones, 2014).

Otros estudios como los de Xiao et al (2023), y Tai et al (2024) no analizaron la tasa de respuesta porque estructuraron sus *prompts* de manera diferente para no darle al LLM la oportunidad para no responder. Valdría la pena investigar como varían los resultados con los dos estilos de *prompt*, si sirve bien para reducir la tasa de alucinaciones, o si provoca un decremento en la calidad de respuestas.

8.1.2 Concordancia simple y Kappa de Cohen

Los resultados Kappa de Cohen en la Figura 4 son muy diferentes para las distintas preguntas, lo que sugiere que hay una gran importancia en el planteamiento de la propia pregunta y la calidad del conjunto de datos en los intentos de utilizar LLM para analizar las respuestas de las entrevistas. En cualquier caso, en este estudio ni siquiera la pregunta con mejor rendimiento llegó a tener una concordancia suficiente para considerarse para la investigación cualitativa científica. Esto no indica necesariamente que los LLM no sean adecuados para esta aplicación, sino más bien que su aplicabilidad depende del contexto y requiere un mayor desarrollo metodológico. Estos resultados son comparables con el análisis de Xiao et al. (2023) en lo cual los valores de Kappa de Cohen variaban entre 0,1 y 0,4 usando GPT-3.

Los valores kappa de Cohen de este estudio sugieren que hay limitaciones en la implementación de los LLM para realizar una investigación cualitativa en un análisis posterior. La metodología del estudio de Jones no tenía en cuenta la posibilidad de análisis con LLM y, como tal, presentaba algunas ambigüedades en el conjunto de datos que contribuyeron a unos valores de Kappa de Cohen bajos. No obstante, la existencia de varias preguntas que demostraron un acuerdo alto entre los codificadores humanos y los LLM respalda las conclusiones de Chew et al (2023) de que se trata de una metodología viable si la metodología está diseñada para acomodar este tipo de análisis.

Los mejores resultados se obtuvieron con preguntas binarias con respuesta sí-no. Prueba de ello es la respuesta “*Union*”, en la que se preguntaba simplemente si el participante formaba parte de un sindicato o no. Esto es consistente con la metodología de Tai et al (2024) que obtuvo mejores resultados limitando los códigos a preguntas binarias.

La diferencia entre los valores de acuerdo simple y los valores Kappa de Cohen pone de manifiesto la importancia de utilizar medidas estadísticas sólidas en la evaluación de los métodos LLM, ya que el acuerdo simple puede dar una puntuación injustamente alta a los resultados.

Aunque las puntuaciones Kappa en muchos casos son bajas, esto no sugiere necesariamente una diferencia en las interpretaciones de los datos originales. En muchos casos de investigación cualitativa como éste, la idea es que se obtenga una impresión general de lo que ocurre en el campo para informar el proceso de decisión política, y que no necesariamente existe una cifra correcta (Propoky, 2011). Vale la pena señalar que no hubo casos en los que el Kappa fuera fuertemente negativo, lo que indicaría un desacuerdo en la interpretación entre el LLM y los codificadores humanos (Cohen, 1960).

Al evaluar la pregunta "*agro in vil*" mediante Kappa de Cohen, se hizo evidente que había muy poca concordancia entre los LLM y los codificadores humanos. El objetivo de esta pregunta era averiguar si había un comerciante de productos agrícolas en el pueblo. Sin embargo, un examen más detallado de la matriz de confusión reveló que esta información específica no se extraía sistemáticamente de todas las entrevistas. Hay cero respuestas "sin respuesta" evaluado por los codificadores humanos, pero al leer unas entrevistas específicas, se nota que ni siquiera se preguntaba a cada agricultor. Esto sugiere que este dato se extrapoló de otras entrevistas realizadas en el mismo pueblo, lo que provocaba incoherencias en la codificación.

8.2 FUENTES DE ERROR

8.2.1 Ambiguidad de códigos

En la pregunta "*price*", la puntuación Kappa fue relativamente baja. Esta pregunta requería distinguir entre "monetario pero no estandarizado (venta informal)" y "monetario y estandarizado (venta formal)". La baja puntuación Kappa sugiere que el LLM tuvo dificultades con esta distinción, probablemente debido a la falta de definiciones claras de lo que constituye una venta formal o informal. Dado que estos términos no están estandarizados universalmente, al LLM le resultó difícil diferenciarlos con precisión. Ante múltiples opciones de respuesta, la kappa de Cohen no diferencia entre respuestas "más erróneas" y "menos erróneas" (Cohen, 1960). Por ejemplo, en la figura 5 sobre el precio, yo consideraría que "monetario no estandarizado" y "monetario estandarizado", aunque no son iguales, están parecidas, mientras que "monetario no estandarizado" y "propio (sin coste)" son muy diferentes - independientemente de ello, el LLM recibió la misma penalización por la selección incorrecta.

La pregunta "*why used*" recibió una puntuación Kappa algo aceptable, pero seguía siendo relativamente baja en comparación con otras preguntas. Esta pregunta tenía múltiples opciones de respuesta que se solapaban: "experiencia previa", "para aprender/probar/ver" y "para tener un mejor rendimiento". La matriz de confusión reveló que los evaluadores humanos a menudo clasificaban el motivo como "experiencia previa", mientras que el LLM lo clasificaba como "para aprender/probar/ver" o "para tener un mejor rendimiento". En realidad, estos razonamientos no son mutuamente excluyentes, ya que la decisión de un agricultor podría entrar fácilmente en varias categorías simultáneamente. Este solapamiento de posibilidades probablemente contribuyó a la menor puntuación Kappa.

8.2.2 Falta de contexto

Parece que hay algunos casos en los que una respuesta de los codificadores humanos no puede justificarse a partir de la propia transcripción de la entrevista. Dado que Jones llevó a cabo su investigación a escala comunitaria, es probable que esté incorporando más información rica en

contexto en la codificación de las respuestas, lo que puede no ser posible con un LLM que carezca de contexto. En un flujo de trabajo ideal centrado en el LLM, las entrevistas intentarían tratar toda la información relevante.

Del mismo modo, en figura 5 la pregunta "*person*" tenía por objeto identificar quién proporcionó las semillas de variedades modernas al agricultor, si se las proporcionó él mismo, otro agricultor, un productor de semillas o un agrodistribuidor/"*animateur*". Las transcripciones de las muestras indicaban que los agricultores solían referirse al proveedor por su nombre, que los investigadores humanos podían interpretar basándose en sus conocimientos locales. En cambio, el LLM asumía con frecuencia que la persona mencionada era otro agricultor y no un comerciante agrícola, lo que demuestra las limitaciones del LLM en cuanto a la comprensión contextual ante la falta de contexto.

Como posible forma de aumentar esto, puede merecer la pena considerar la posibilidad de desarrollar un documento que incluya detalles adicionales esenciales para interpretar las entrevistas, o emplear un método que utilice la RAG como sugiere Silva et al (2023) y Balaguer et al (2024), en lo cual se proporciona una biblioteca de estudios previos o información complementaria que pueda aportar un contexto más rico al LLM. Por ejemplo, en el ejemplo anterior del error con el nombre del "*animateur*" que se confundió con otro agricultor, si hubiera habido un documento complementario que explicara más sobre los participantes y otros nombres a los que se hace referencia en las entrevistas, quizás el LLM no habría confundido entre las respuestas.

8.2.3 Errores heredados del estudio original

Dado que Jones (2014) utilizó una metodología mixta para la recogida de datos en su tesis, la calidad de cada entrevista es variable, según el deseo del participante. Entonces, los textos de las entrevistas no siguen un guion y, como tal, cada uno de ellos contiene información diversa. En muchos casos es de suponer que algunos datos se recogieron a través de la observación personal, o de conversaciones informales fuera de la propia entrevista.

Las transcripciones de las entrevistas de Jones fueron traducidas del francés al inglés por el propio Jones y, en varios casos, la entrevista se realizó en una lengua local y se anotó en francés,

I: Ok. And did you save part of your harvest in 2011 for next year?
T: Just a little bit.
I: Ok. And do you think that you will buy again next year?
T; He would like to do a field, just this variety.
I: Ok.
T: One hectare.
I: and were there people who asked you for this variety?
T: There were people who asked and he even gave.
I: Ah ok. After 2010 or 2011?
T: This year.
I: Can you give me the names of those who got seeds with you?

lo que da lugar a múltiples niveles de abstracción de lo que realmente dijo el agricultor. Por ello, es posible que en las transcripciones de las entrevistas se pasen por alto sutilezas o detalles fundamentales debido al proceso de traducción.

Todos los análisis estadísticos presuponen que las respuestas de Jones en el libro de códigos son "correctas". Hay varias razones por las que esto puede no ser así. Por ejemplo, en la figura 8 se muestra un segmento de una entrevista.

En estas entrevistas, en el campo *Sharing*, Jones dijo que el agricultor no había participado en actividades de intercambio de semillas, mientras que GPT-4 indica que el agricultor sí había participado en el intercambio de semillas. Al leer el texto, se ve claramente que el agricultor sí compartió semillas, por lo que la evaluación de GPT-4 es más "correcta". Esto sirve para recordar que los índices de similitud son sólo eso: indicadores de similitud, no necesariamente de corrección. Esto sugiere una implementación de los LLM en el análisis cualitativo para filtrar los datos y destacar los posibles errores.

8.3 LIMITACIONES DEL ESTUDIO

Los resultados diferentes entre técnicas de *prompting* no se consideraron en este estudio, por lo que es posible que existan otras que produzcan mejores resultados que las mías. La optimización de la ingeniería de las preguntas habría sido costosa.

Uno de los objetivos del estudio de Jones (2014) era poder realizar un seguimiento entre años de los mismos agricultores. Esto significa que, en la tabla de datos de Jones, es probable que alguna información general sobre el propio agricultor pueda duplicarse de un año a otro, a pesar de no mencionarse explícitamente en la transcripción de la entrevista. Esto provocaría una reducción del rendimiento del LLM.

Las puntuaciones kappa pueden parecer bajas debido al hecho de que sólo se seleccionaron las respuestas que requerían interpretación, es decir, no son simples hechos que se extraen de la entrevista como simples cifras o palabras, sino más bien información sobre las experiencias o pensamientos del agricultor que deben "entenderse" desde un contexto más amplio. Es probable que se hubieran obtenido mejores resultados si hubiéramos analizado datos más fáciles de interpretar – por ejemplo, Jones (2014) preguntó la edad de los agricultores, lo que sería un proceso de extraer un número en lugar de interpretar el sentimiento de una conversación.

Debido a las limitaciones de tiempo y presupuesto, este análisis se realizó sobre entrevistas que originalmente se habían planificado con otras metodologías. Esto es consistente con estudios anteriores de LACA como Dai et al. (2023), Tai et al (2024), y Chew et al. (2023). Sin embargo, ambos estudios usaron códigos inventados para el propósito de análisis con LLM en lugar de códigos de un análisis anterior. Dado que no se diseñaron para esta metodología, los códigos

usados por Jones (2014) perjudicaron el rendimiento del LLM. En el futuro, sería interesante llevar a cabo las entrevistas considerando que se van a analizar por LLM, y así dirigir la conversación para abordar todos los detalles en una manera un poco más explícita.

Dado que Jones (2014) había diseñado las preguntas de la entrevista y el libro de códigos teniendo en mente preguntas de investigación específicas, y había realizado una codificación inductiva de las respuestas, quedaban menos oportunidades para debatir otros temas de los deseos de los agricultores. Uno de los problemas inherentes a la investigación cualitativa, especialmente en conceptos transculturales, es que los prejuicios del investigador sobre lo que es importante y relevante pueden ofuscar la información y los conocimientos especializados que los sujetos son capaces de proporcionar (Propoky, 2011).

8.4 IMPLICACIONES

Estos resultados indican una utilidad más limitada de las metodologías LLM como medio de análisis posterior. Los resultados podrían ser mejores si la metodología completa se desarrollara teniendo en cuenta el procesamiento basado en LLM. Si la intención es utilizar los LLM para el análisis de transcripciones de entrevistas, sería prudente adoptar una metodología de entrevista que refleje esa intención. En la medida de lo posible, se podrían formular preguntas aclaratorias de seguimiento para proporcionar detalles clave del contexto que no se incluyan en la propia entrevista

Los posibles beneficios de eficiencia pueden ahorrar horas de tiempo a los investigadores en la codificación (Hamilton et al, 2023). Sin embargo, este beneficio de eficiencia ha sido criticado por algunos, que afirman que las horas dedicadas a analizar los datos les han ayudado a comprender de forma más intuitiva las tendencias de los datos y a detectar algunos temas y detalles que, de otro modo, no se habrían tenido en cuenta en la generación del libro de códigos (Hamilton et al, 2023, Morgan, 2023)

En resumen, la evaluación de estas preguntas específicas pone de manifiesto varios retos en el uso de los LLM para el análisis cualitativo. Las incoherencias en la extracción de información matizada, la dificultad para interpretar la terminología específica del contexto y la naturaleza solapada de las categorías de respuesta contribuyen a las variaciones en el acuerdo entre los LLM y los codificadores humanos. Estos resultados subrayan la importancia de perfeccionar las metodologías LLM y mejorar su comprensión contextual para mejorar su fiabilidad en la investigación cualitativa.

8.5 ORIENTACIÓN FUTURA DE LA INVESTIGACIÓN

Dada la amplitud de las diferentes metodologías que se han explorado en torno al uso de los LLMs en la investigación cualitativa en el sector agrícola, futuras investigaciones podrían explorar la combinación de estas metodologías para aprovechar los beneficios de los LLM en varios pasos del proceso de investigación. Los avances rápidos sugieren un modelo de metodología de investigación comprensivo que incluye elementos de todos los métodos abarcados en los estudios anteriores. Un modelo de implementación de este sistema está explicado en la figura 10.

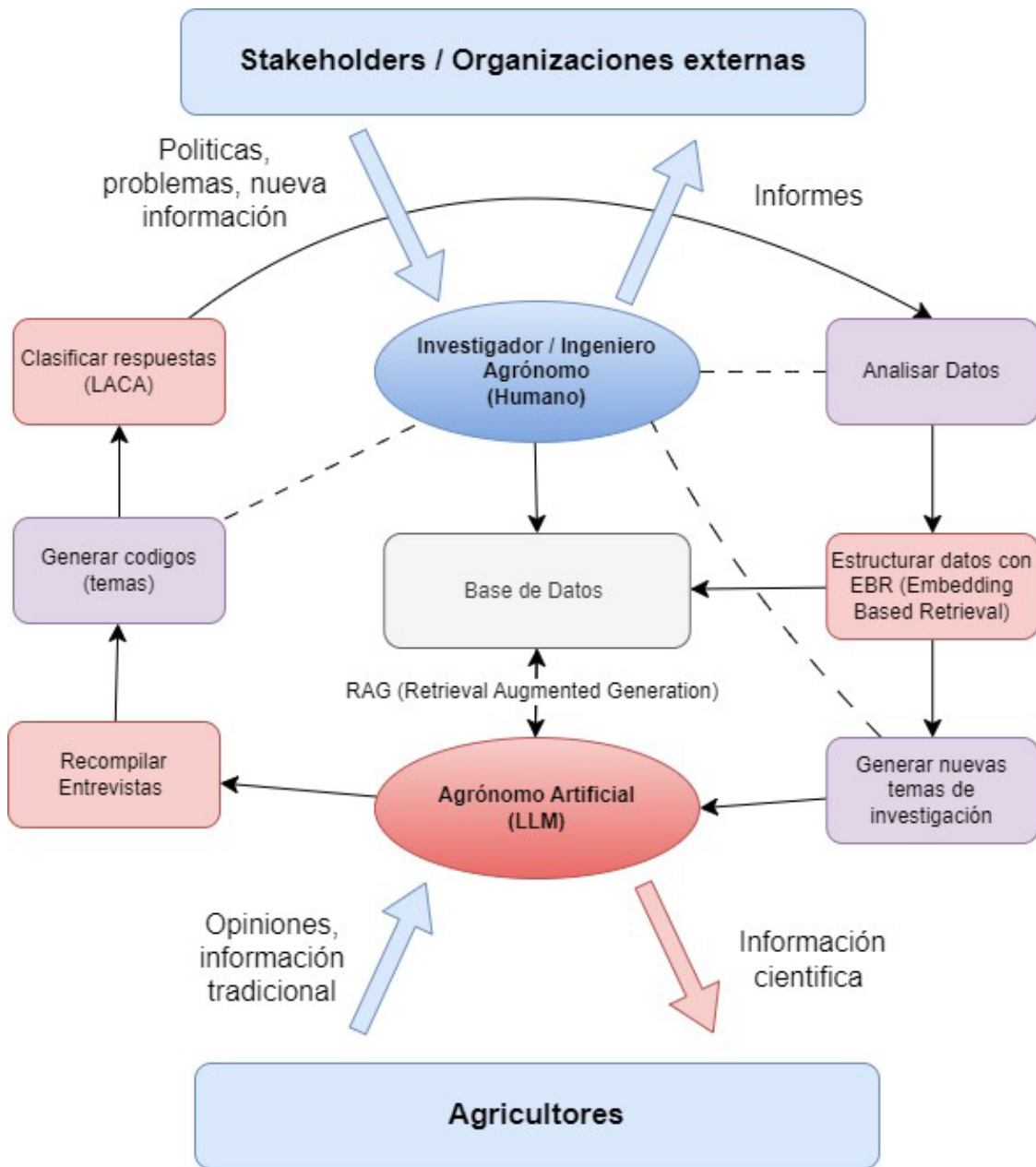


Figura 10. Modelo propuesto de sistema de investigación continua basado en metodologías con LLMs. Elaboración propia.

Este modelo se basa en un sistema de colaboración multiagente entre un investigador humano y un LLM como planteó por Dai et al (2023), combinado con un ingeniero agrónomo artificial como implementaron Sowole et al (2023) and Ibrahim et al. (2024). En este modelo, el LLM sirve como punto de contacto con un grupo de agricultores, proporcionando información científica mediante el uso de RAG como herramienta para asegurar la precisión de la información (Silva et al, 2024).

Mientras el LLM proporciona esta información, también se puede encargarlo con la responsabilidad de llevar a cabo una entrevista con el agricultor para coleccionar información sobre un tema de investigación. Las entrevistas pueden realizarse mediante una IA conversacional de última generación, minimizando el coste de la mano de obra humana y el transporte. Utilizar la IA para las entrevistas puede automatizar el proceso, haciéndolo más escalable y menos intensivo en recursos. Este método no solo ahorra tiempo, sino que también garantiza la coherencia en la recopilación de datos entre distintos encuestados.

Luego, las transcripciones de las conversaciones de estas interacciones se tratarían de forma similar a las entrevistas tradicionales para realizar la identificación de temas por parte del LLM con colaboración del investigador, similar a Hamilton et al. (2023).

El siguiente paso es llevar a cabo el proceso de LACA usando los temas identificados como libro de códigos, en la misma manera en que se implementó en este estudio y los del Dai et al (2023) y Chew et al (2023). A través de este análisis, se cuantifica las ideas y los sentimientos de los agricultores en cuanto a los temas de investigación, destacando los retos más importantes en su día cotidiana, y compilando los conocimientos tradicionales que suelen ser olvidados por el mundo de investigación científico (Propoky et al. 2011).

Los resultados se compilan mediante el EBR usando las metodologías de Peng et al., (2023) y Silva et al. (2023), cambiando la estructura de los datos para facilitar su acceso por parte del LLM. Los temas y consejos generados a partir de estas interacciones se añadirían a la biblioteca del RAG para su uso posterior, o bien para los agricultores o para el investigador. El investigador es capaz de interpretar estos resultados para los stakeholders u organizaciones externas que buscan la participación de los agricultores. El agrónomo o investigador también utilizaría el asesor digital como herramienta para difundir nueva información y llevar a cabo actividades de divulgación. Al añadir preguntas de investigación e interpretar datos dentro del sistema, el agrónomo se asegura de que la IA siga siendo un recurso valioso para la comunidad agrícola. Este bucle de retroalimentación permite que el sistema crezca en conocimientos especializados, abarcando tanto los aspectos científicos como los socioeconómicos de la agricultura.

Con estos resultados, el LLM y el investigador colaboran en la generación de nuevas preguntas de investigación. Así se completa el bucle de retroalimentación, planteando nuevas preguntas para conversaciones futuras entre el LLM y los agricultores. Este proceso podría hacerse de forma continua, generando un sistema resiliente que responda a las necesidades de los

agricultores y los investigadores.

La metodología holística de investigación propuesta ofrece un marco sólido para integrar los LLM en la investigación y los servicios de asesoramiento agrícolas. Al combinar los puntos fuertes de la IA y la experiencia humana, este enfoque puede mejorar la eficiencia, la profundidad y la aplicabilidad de la investigación cualitativa. Esta estructura innovadora no sólo mejora la recopilación y el análisis de datos, sino que también proporciona una solución escalable para ofrecer asesoramiento agrícola a medida. A medida que el sistema evoluciona, tiene el potencial de influir significativamente en el desarrollo agrícola sostenible y la resiliencia de los agricultores.

9. CONCLUSIONES

Este trabajo pretende analizar la posibilidad de la utilización de los LLM en el análisis de entrevistas en la investigación cualitativa en el sector agrario. Usando una selección de métodos desarrollados en la investigación actual, se ha podido realizar este objetivo en una manera eficaz.

Basándonos en los resultados de este estudio, la aplicación de los LLM para el posterior análisis de contenido temático de investigaciones previas es limitada; sin embargo, son muchas las utilidades que esta tecnología puede tener en este campo, y una trayectoria positiva para futuros desarrollos. Los investigadores deben ser cautelosos a la hora de aplicar esta tecnología sin diseñar todo el estudio para adaptarlo a las necesidades del LLM.

El uso de LLM parece magnificar cualquier práctica de diseño metodológico deficiente en la investigación cualitativa. Un LLM, aunque tiene una comprensión más rica del contexto que las formas anteriores de inteligencia artificial, sigue funcionando de una manera muy literal, por lo que cualquier fallo de diseño en la metodología en la que la interpretación se deja en manos del codificador puede malinterpretarse sistemáticamente.

Los valores Kappa de Cohen para la mayoría de las preguntas de la entrevista fueron muy próximos a cero, independientemente del modelo utilizado, lo que indica un nivel de significación no diferente del azar. Sin embargo, hubo dos preguntas que demostraron una concordancia muy alta, especialmente cuando se eliminan las situaciones de ausencia de respuesta, lo que indica que hay situaciones en las que un LLM puede realizar esta tarea con suficiente precisión.

Comparando GPT-3.5 y GPT-4, los resultados sugieren que existen diferencias sustanciales entre los dos modelos y que cualquier metodología que utilice LLM se beneficiaría del uso del modelo más avanzado, a pesar del mayor coste. Este aumento del rendimiento da motivos para ser optimistas sobre el potencial que pueden ofrecer los modelos futuros.

El uso de los LLM en la investigación cualitativa en el sector agrícola muestra un gran potencial, aunque queda mucho trabajo por hacer para llegar a implementaciones que puedan competir con el análisis humano en calidad suficiente para ser incluidas en la investigación académica de grado profesional. Uniendo las variadas metodologías que están siendo exploradas por los investigadores sobre cómo aprovechar mejor esta tecnología naciente, podemos ir más allá

de la lente de adaptar el uso de los LLMs a metodologías existentes y empezar a considerar la posibilidad de metodologías centradas en los LLMs que permitan la evaluación de una cantidad mucho mayor de datos cualitativos, de forma similar a como la computación ha sido capaz de realizar para datos cuantitativos en décadas anteriores.

10. BIBLIOGRAFÍA

Dai, Shih-Chieh, Aiping Xiong, and Lun-Wei Ku. "LLM-in-the-loop: Leveraging Large Language Model for Thematic Analysis." arXiv, 23 Oct. 2023, arXiv:2310.15100v1 [cs.CL].

Chew, Robert, John Bollenbacher, Michael Wenger, Jessica Speer, and Annice Kim. "LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding." arXiv, 23 Jun. 2023, arXiv:2306.14924v1 [cs.CL].

Torii, Maya Grace, Takahito Murakami, and Yoichi Ochiai. "Expanding Horizons in HCI Research Through LLM-Driven Qualitative Analysis." arXiv, 7 Jan. 2024, arXiv:2401.04138v1 [cs.HC].

Tai, Robert H., Lillian R. Bentley, Xin Xia, Jason M. Sitt, Sarah C. Fankhauser, Ana M. Chicas-Mosier, and Barnas G. Monteith. "An Examination of the Use of Large Language Models to Aid Analysis of Textual Data." *International Journal of Qualitative Methods*, vol. 23, 2024, pp. 1-14, DOI: 10.1177/16094069241231168.

Xiao, Ziang, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. "Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding." In 28th International Conference on Intelligent User Interfaces (IUI '23 Companion), March 27–31, 2023, Sydney, NSW, Australia. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3581754.3584136>.

Balaguer, Angels, Vinamra Benara, Renato Cunha, Roberto Estevão, Todd Hendry, Daniel Holstein, Jennifer Marsman, Nick Mecklenburg, Sara Malvar, Leonardo O. Nunes, Rafael Padilha, Morris Sharp, Bruno Silva, Swati Sharma, Vijay Aski, and Ranveer Chandra. "RAG vs Fine Tuning: Pipelines Tradeoffs and A Case Study on Agriculture." arXiv, 30 Jan. 2024, arXiv:2401.08406v3 [cs.CL].

Peng, Ruoling, Kang Liu, Po Yang, Zhipeng Yuan, and Shunbao Li. "Embedding-Based Retrieval with LLM for Effective Agriculture Information Extracting from Unstructured Data." arXiv, 6 Aug. 2023, arXiv:2308.03107v1 [cs.AI].

Silva, Bruno, Leonardo Nunes, Roberto Estevão, Vijay Aski, and Ranveer Chandra. "GPT-4 as an Agronomist Assistant? Answering Agriculture Questions Using Large Language Models." arXiv, 12 Oct. 2023, arXiv:2310.06225v2 [cs.AI].

Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., and Mordatch, I. Improving factuality and reasoning in language models through multiagent debate. arXiv preprint arXiv:2305.14325, 2023.

Sowole, Oladimeji Samuel, and Jonathan Enudeme. "Leveraging Large Language Models for Improving Agricultural Extension in Nigeria." 5th Deep Learning Indaba Conference (DLI 2023), 2023.

Ibrahim, Ali, Kalimuthu Senthilkumar, and Kazuki Saito. "Evaluating Responses by ChatGPT to Farmers' Questions on Irrigated Lowland Rice Cultivation in Nigeria." *Scientific Reports*, vol. 14, no. 3407, 2024, <https://doi.org/10.1038/s41598-024-53916-1>.

William, P., Ahmad Y. A. Bani Ahmad, A. Deepak, Rashmi Gupta, K. K. Bajaj, and Dr. Renuka Deshmukh. "Sustainable Implementation of Artificial Intelligence Based Decision Support System for Irrigation Projects in the Development of Rural Settlements." *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3s, 2024, pp. 48–56.

Hamilton, Leah, Desha Elliott, Aaron Quick, Simone Smith, and Victoria Choplin. "Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data." *International Journal of Qualitative Methods*, vol. 22, 2023, pp. 1–13, DOI: 10.1177/16094069231201504.

Morgan, David L. "Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT." *International Journal of Qualitative Methods*, vol. 22, 2023, pp. 1–10, DOI: 10.1177/16094069231211248.

Patil, Rajvardhan, and Venkat Gudivada. "A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)." *Applied Sciences*, vol. 14, no. 2074, 2024, <https://doi.org/10.3390/app14052074>. Accessed 1 Mar. 2024.

Makridakis, Spyros, Fotios Petropoulos, and Yanfei Kang. "Large Language Models: Their Success and Impact." *Forecasting*, vol. 5, no. 3, 2023, pp. 536-549, <https://doi.org/10.3390/forecast5030030>. Accessed 25 Aug. 2023.

Prokopy, Linda S. "Agricultural Human Dimensions Research: The Role of Qualitative Research Methods." *Journal of Soil and Water Conservation*, vol. 66, no. 1, 2011, pp. 9A-12A, <https://doi.org/10.2489/jswc.66.1.9A>. Accessed 10 Feb. 2016.

Vaswani, Ashish, et al. "Attention Is All You Need." 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017, arXiv:1706.03762v7. Accessed 2 Aug. 2023.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. doi:10.1177/001316446002000104

Pretty, Jules N., and Rachel Hine. "The Sustainable Intensification of Agriculture." *World Development*, vol. 23, no. 4, 1995, pp. 501-510, [https://doi.org/10.1016/0305-750X\(95\)00125-J](https://doi.org/10.1016/0305-750X(95)00125-J).

European Commission, Joint Research Centre, Millard, Jeremy, Manzoni, Marco, and Schade, Sven. *Impact of Digital Transformation on Public Governance – New Forms of Policy-Making and the Provision of Innovative, People-Centric and Inclusive Public Services*. Publications Office of the European Union, 2023, <https://doi.org/10.2760/204686>. Accessed 2023.

Weidinger, Laura, et al. "Taxonomy of Risks Posed by Language Models." *FAcCT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 214-229. Accessed 26 Apr. 2023.

Zhang, Zhengyu, et al. "Towards Adaptive Prefix Tuning for Parameter-Efficient Language Model Fine-Tuning." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, arXiv:2305.15212. Accessed 2023.

Bano, Muneera, Didar Zowghi, and Jon Whittle. "Exploring Qualitative Research Using LLMs." *CSRIO's Data61*, 2023, arXiv:2306.13298v1. Accessed 2023.

Jones, K. "A Comprehensive Study on the Impact of Digital Transformation in Education." PhD dissertation, University of XYZ, 2023.

Chu, Zhibo, et al. "History, Development, and Principles of Large Language Models—An Introductory Survey." *Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences*, 2024, arXiv:2402.06853v1. Accessed 2024.

Smith, John, et al. "Artificial Intelligence in Healthcare: The Role of Large Language Models." *BMJ*, vol. 380, no. 074256, 2022, pp. 1-10. Accessed 2022.

11. ANEXOS

ANEXO 1: METODOLOGÍA DE JONES (2014)

El estudio examina cómo la reciente introducción de mercados de insumos para variedades de semillas mejoradas repercute en las prácticas agrícolas locales y en el panorama agrícola en general, concretamente en lo que respecta al sorgo y el mijo perla, cereales autóctonos en los que hasta ahora no se había centrado la investigación agrícola privada.

Jones emplea un enfoque de métodos mixtos, integrando el análisis temático de datos cualitativos, el modelado de la curva de crecimiento de un conjunto de datos de panel y la representación visual de la difusión de semillas para analizar el acceso a las semillas a varias escalas, desde las acciones individuales hasta los sistemas comunitarios y regionales. La tesis se apoya en los marcos teóricos de las economías sustantivas y los hábitos económicos, que conceptualizan las acciones e instituciones económicas como moldeadas por sus motivaciones sociales y contextos naturales específicos.

Las principales conclusiones del estudio sugieren que en la región coexisten sistemas de semillas basados tanto en el intercambio como en el aprovisionamiento, cada uno de ellos sustentado por principios organizativos distintos pero interconectados a través de acciones individuales de acceso a las semillas. El estudio revela que las características del entorno social y natural influyen significativamente en las decisiones individuales de acceso a las semillas a lo largo del tiempo, observándose notables diferencias entre hombres y mujeres y entre individuos con distinta proximidad a infraestructuras sociales como los mercados.

La metodología empleada por Kristal Jones en su tesis doctoral para realizar las entrevistas y codificar los datos cualitativos es detallada y meticulosa, y refleja un enfoque sólido para comprender los complejos sistemas agrícolas de África Occidental. He aquí un resumen de cómo se llevaron a cabo las entrevistas y la posterior codificación de los datos cualitativos:

Jones realizó sus entrevistas principalmente en comunidades rurales de Malí, Burkina Faso y Níger, centrándose en zonas muy dedicadas al cultivo de sorgo y mijo perla. La selección de los participantes en las entrevistas fue estratégica, con el objetivo de abarcar un amplio espectro de perspectivas dentro de la comunidad agrícola. Entre ellos había agricultores, distribuidores de semillas y funcionarios agrícolas locales, lo que garantizaba una representación diversa de los sistemas de semillas de estas regiones.

Las entrevistas fueron semiestructuradas, con preguntas específicas para obtener información detallada sobre el uso de las semillas y otras más abiertas que animaban a los participantes a hablar de sus experiencias y percepciones de los cambios agrícolas. Este formato flexible permitió

a Jones profundizar en los temas en función de las aportaciones de los entrevistados, proporcionando datos cualitativos de gran riqueza.

Tras la recopilación de datos, Jones empleó el análisis temático para identificar, analizar e informar sobre patrones dentro de los datos cualitativos. Este proceso iterativo consistió en leer detenidamente las transcripciones de las entrevistas, anotar las ideas iniciales, codificar sistemáticamente los datos de todo el conjunto de datos y cotejar estos códigos en temas potenciales. La estrategia de codificación fue tanto deductiva como inductiva; se utilizaron códigos predefinidos para etiquetar los datos basándose en el marco teórico de las economías sustantivas, mientras que de los datos surgieron nuevos códigos que reflejaban temas inesperados relacionados con el acceso a las semillas y los comportamientos económicos.

Para garantizar la fiabilidad de las conclusiones, el proceso de codificación incluyó varias rondas de revisión y ajustes. Inicialmente, Jones aplicó los códigos, que luego fueron revisados por un investigador independiente para validar la coherencia y la interpretación de la codificación. Cualquier discrepancia se debatió hasta que se alcanzó un consenso, reforzando así la validez de los constructos temáticos derivados de los datos. Este riguroso proceso garantizó que el análisis cualitativo fuera sólido y reflejara con precisión las complejidades de los sistemas de semillas en África Occidental.

Una transcripción de una entrevista de Jones (2014) está adjunto abajo:

4472012

Transcribed 2/26/2013

I: Your name?

R: ----

I: And you are how old?

T: 60.

I: And there are how many people in your household?

T: 6.

I: And you yourself, you planted how much space for millet last year?

T: Four hectares.

I: Ok. And you bought how many packets of seeds last year?

T: One packet of one kilo.

I: And it was which variety?

T: He said he used another variety, but with the seeds he bought, eh got them with el Hadj, he really liked them.

I: And was it the first time that you'd used this variety?

T: He said that he gets each time, there were seeds...he tried, what he'd like...partner, he's a member of the union, so he does observations. So these seeds, he's used to using them.

I: Ok. And each year, do you save part of your harvest to replant the next year?

T: Yes.

I: Mmm hmm?

T: He is familiar with these seeds, and he's used to bringing these seeds.

I: And do you notice a difference between the first time that you buy these seeds, and the second

time that you replant them?

T: The seeds that he got with El Hadj, they...

I: but with these seeds, if you save and replant the next year, are there changes that come?

T: He said that the seeds give a good yield, and then next year the yield is even better.

I: Ok. And are there people who ask you for this variety?

T: Yes, and he gave to those who asked, and they also tried and saw the advantages of these seeds.

I: Ok. Can you give me the names of those who got seeds with you?

T: There's a list.

I; Ok.

R: ----

I: And it was after which year that you gave to them?

T: 2010.

I; Ok. And this year, did you save a part to replant next year?

T: He said that with these seeds, those that don't have seeds, he will buy them at the input shop.

He will buy them.

I: How many this year?

T: Four packets.

I: And do you have suggestions in terms of seed sales?

T: He said that the seeds, finding them at the input shop, it's a good thing, because if people want them they can buy them.

I: Ok. And have you noticed changes in terms of how people find and get all kinds of seeds?

T: He said that the seeds of improved varieties, you can find them with the input shop. HE said that allows them to have a good yield, and at the same time, you use...it allows them to work, because you can pass there on time.

I: Ok. And do you listen to the radio?

R: Yes.

I: Ok.

R: Binta.

I: And do you have equipment or cows?

T: No.

I: Ok. And are there other things that you can tell me in terms of seeds?

T: He said about the seeds, to have other varieties, of millet, vouandzou, peanuts. Bring them also with inputs.

I: Ok.

ANEXO 2: *PROMPT*

You are an agricultural researcher conducting qualitative research in West Africa. Your job is to code a series of interview transcripts according to the master codebook below. Each code definition consists of a number in [square brackets] that organizes the heading of each code, some clarifying text that explains the information that is sought, and then some coding information to explain the format of the desired response.

Master codebook

[1]Farmer ID

Unique ID assigned to each individual farmer interviewed

[2]Saw MVs

Did the farmer see modern varieties in a field prior to using them?

1=Yes

0=No

[3]Initial info

Where did the farmer get initial information about modern varieties? 'Modern varieties' refer to those that have been developed through scientific plant breeding techniques, and are maintained as durable, uniform and standard under a specific name.

1=other farmer

2=animateur

3=agrodealers

4=radio

[4]Why used

Why did the farmer decide to try modern varieties?

1=to learn/try/see

2=previous experience

3=to have a better yield

4=to deal with lack of rainfall

5=to have an early maturing variety

[5]Union

Is the farmer a member of a farmers/producers union?

1=Yes

0=No

[6]Anim in vil

Is there an animateur in the village? An 'animateur' is something like a local agricultural trainer or extension agent. They can be employed by the local government agricultural office, or by farmers' unions.

1=Yes

0=No

[7]Agro in vil

Is there an agrodealer in the village? An 'agrodealer' is a small merchant who sells seeds, pesticides, fertilizer and other agricultural inputs. They are

often employed by either an input company (like Cignone, the French imprint of Monsanto) or by development

organizations and development dollars. In this study, there were agrodealers only in Mali, and almost all were

funded through a Bill and Melinda Gates Foundation project.

1=Yes

0=No

[8]Access place

Location that modern variety seeds were accessed

0=Did not use

1=Self (saved from previous use)

2=Someone else's home or field

3=Market from informal seller

4=Agrodealer shop

[9]Price

Price paid for modern variety seeds

0=Did not use

1=Self (no cost)

2=Non-monetary (exchange of other goods for seeds)

3=Monetary but not standardized (informal sale)

4=Monetary and standardized (formal sale)

[10]Person

Person who provided modern variety seeds

0=Did not use

1=Self

2=Another farmer (other than a seed producer)

3=A seed producer

4=Agrodealer or animateur

[11]Certification

Information about the quality of the modern variety seeds provided when seeds were acquired

0=Did not use

1=No information or certification

2=Social certification (seed provider gave information about type and quality of seed)

3=Formal certification (seeds were scientifically certified for type and quality)

[12]Yrs saved

Number of years a farmer has saved and re-used the same modern variety seeds since initial acquisition

0=Did not use

1=3 years

2=2 years

3=1 year

4=Has not saved

[13]Packaging

Packaging of modern variety seeds when initially acquired

0=Did not use

1=No packaging

2=Plastic bag but not sealed and with no information

3=Sealed plastic bag with information

[14]Ident var

Ability of farmer to identify the modern varieties they have used

0=Did not use

1=Did not know anything about them

2=Identification based on characteristics (e.g. 'long red panicles')

3=Identification by name

[15]Saving

Seed saving actions with modern varieties taken by the farmer

0=Did not use

1=Saved seeds and used following year

2=Saved or wanted to save but didn't use the next year

3=No saving

[16]Sharing

Seed sharing actions with modern varieties taken by the farmer

0=Did not use

1=Gifted seeds to another farmer

2=Exchanged seeds with another farmer

3=Gave information about seeds to another farmer

4=No sharing

For the interview transcript, you must analyze the content and assign a value for each and every one of the codes according to the system specified by the codebook. Any line starting with I: is a question of the interviewer - for coding we are interested in the responses of the interviewee.

As you go one by one, respond as such:

[number] codename

(meaning) meaning of code

(Rationale): Rationale for selecting your code (could be a quote from the interview or interpreted response)

{final code choice} make sure you are providing the code in the format suggested in the codebook above, as a numerical response where appropriate

If you are not able to provide a code for a given field in the interview, you may answer blank as { }, but you must provide an entry for all 16 of the questions.

ANEXO 3: EJEMPLO DE RESPUESTA GPT

[1] Farmer ID

Unique ID assigned to each individual farmer interviewed

(Rationale): The unique ID seems to be the identifier used to name the file: 447.

{final code choice} {447}

[2] Saw MVs

Did the farmer see modern varieties in a field prior to using them?

(Rationale): The text does not explicitly state whether the farmer saw modern varieties in the field prior to using them, only that he has experience with them.

{final code choice} { }

[3] Initial info

Where did the farmer get initial information about modern varieties?

(Rationale): The farmer mentions getting seeds with "el Hadj," suggesting interaction with another farmer, but also mentions being part of a union and making observations, which could imply different sources of information.

{final code choice} {1} (assuming el Hadj is another farmer)

[4] Why used

Why did the farmer decide to try modern varieties?

(Rationale): The farmer states he used these seeds before, tries them, and observes their performance, indicating an experimental approach.

{final code choice} {1}

[5] Union

Is the farmer a member of a farmers/producers union?

(Rationale): The farmer confirms he is a member of the union.

{final code choice} {1}

[6] Anim in vil

Is there an animateur in the village?

(Rationale): There is no specific mention of the presence of an animateur in the village.

{final code choice} { }

[7] Agro in vil

Is there an agrodealer in the village?

(Rationale): The farmer mentions buying seeds at the input shop, suggesting the presence of an agrodealer.

{final code choice} {1}

[8] Access place

Location that modern variety seeds were accessed

(Rationale): The farmer mentions buying seeds at the input shop.

{final code choice} {4}

[9] Price

Price paid for modern variety seeds

(Rationale): The farmer purchased seeds at an input shop, likely implying a monetary and standardized transaction.

{final code choice} {4}

[10] Person

Person who provided modern variety seeds

(Rationale): The farmer obtained seeds from an input shop, which is run by an agrodealer.

{final code choice} {4}

[11] Certification

Information about the quality of the modern variety seeds provided when seeds were acquired

(Rationale): The text does not specify the type of certification information provided when acquiring the seeds.

{final code choice} { }

[12] Yrs saved

Number of years a farmer has saved and re-used the same modern variety seeds since initial acquisition

(Rationale): The farmer saves seeds every year but it is not clear how many years he has been reusing them.

{final code choice} { }

[13] Packaging

Packaging of modern variety seeds when initially acquired

(Rationale): No specific information about the initial packaging of the seeds.

{final code choice} { }

[14] Ident var

Ability of farmer to identify the modern varieties they have used

(Rationale): The farmer is familiar with the seeds and their performance, indicating knowledge of the variety based on characteristics.

{final code choice} {2}

[15] Saving

Seed saving actions with modern varieties taken by the farmer

(Rationale): The farmer saves part of his harvest to replant each year.

{final code choice} {1}

[16] Sharing

Seed sharing actions with modern varieties taken by the farmer

(Rationale): The farmer has given seeds to others who asked for them.

{final code choice} {1}