

INTELIGENCIA ARTIFICIAL EN EDUCACIÓN: CIBERSEGURIDAD - GROOMING

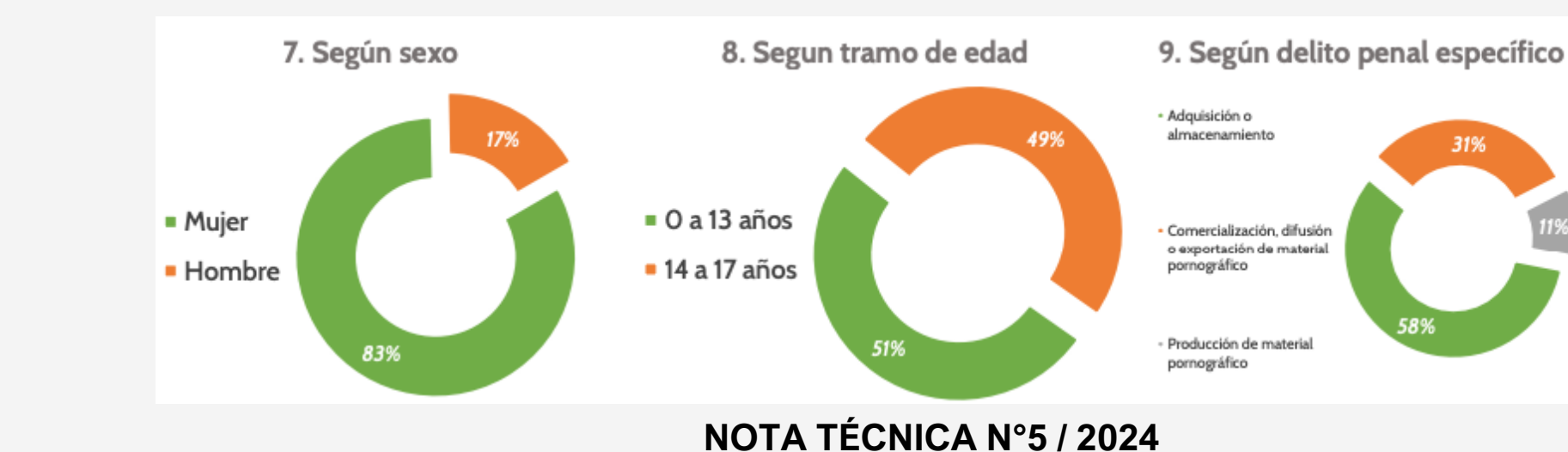
Cristian Medina Rojas. (e)

Doctorado en Inteligencia Artificial

Dra. Mabel Urrutia Martínez

INTRODUCCIÓN

En la era digital, los niños están expuestos a múltiples riesgos digitales, siendo el grooming uno de los más preocupantes. Este fenómeno implica que un adulto manipule emocionalmente a un menor para explotarlo sexualmente. La necesidad de implementar herramientas efectivas que protejan a los niños en entornos educativos es crucial (Vogt, Leser & Akbik, 2021). Se ha demostrado que es posible identificar patrones lingüísticos específicos en las conversaciones de grooming, lo que subraya la importancia de aplicar un enfoque analítico y sistemático para la detección temprana de estos comportamientos (Inches & Crestani, 2012). En este contexto, la Inteligencia Artificial (IA) ofrece soluciones prometedoras, permitiendo no solo la monitorización, sino también el análisis automatizado de las interacciones en línea, facilitando la identificación y prevención del grooming en tiempo real (Prosser & Edwards, 2024).



Datos obtenidos desde la defensorianinez.cl

Tasa de niñas, niños y adolescentes víctimas en casos ingresados al Ministerio Público durante 2022 y 2023 según materia relacionada a ESNA en medios digitales, por región

OBJETIVOS

- Desarrollar un software basado en IA que utilice modelos BERT para la detección de grooming en tiempo real en entornos educativos.
- Evaluar la precisión y efectividad del modelo BERT en la identificación de frases de grooming a partir del dataset existente.
- Integrar y promover estrategias educativas basadas en el software para aumentar la conciencia y ciberseguridad en la prevención del grooming en entornos escolares.

METODOLOGÍA

- Se empleó un modelo de algoritmo para detectar patrones lingüísticos de grooming con Dataset(PAN12), mediante IA.
- Se utilizó el Dataset con frases de grooming para entrenar el modelo de IA.
- Actualmente se ajusta el algoritmo de detección y entrenamiento, con el modelo BERT, lo que facilita la integración del análisis lingüístico e integración de nuevos corpus.
- Actualmente, se está desarrollando plataforma grafica para un ambiente educativo que integra los modelos entrenados, permitiendo la monitorización en tiempo real de interacciones para detectar grooming, con sistema de alerta tipo semáforo integrado con Arduino.
- Se validará el software en entornos simulados y se comparará el rendimiento de BERT con otros modelos de lenguaje a gran escala (LLMs), evaluando precisión y fiabilidad.
- Se integrará didáctica y capacitará a los educadores en el uso del software y se llevará a cabo un monitoreo continuo para mejorar el modelo en función de la retroalimentación y las nuevas tácticas de grooming.

Referencias

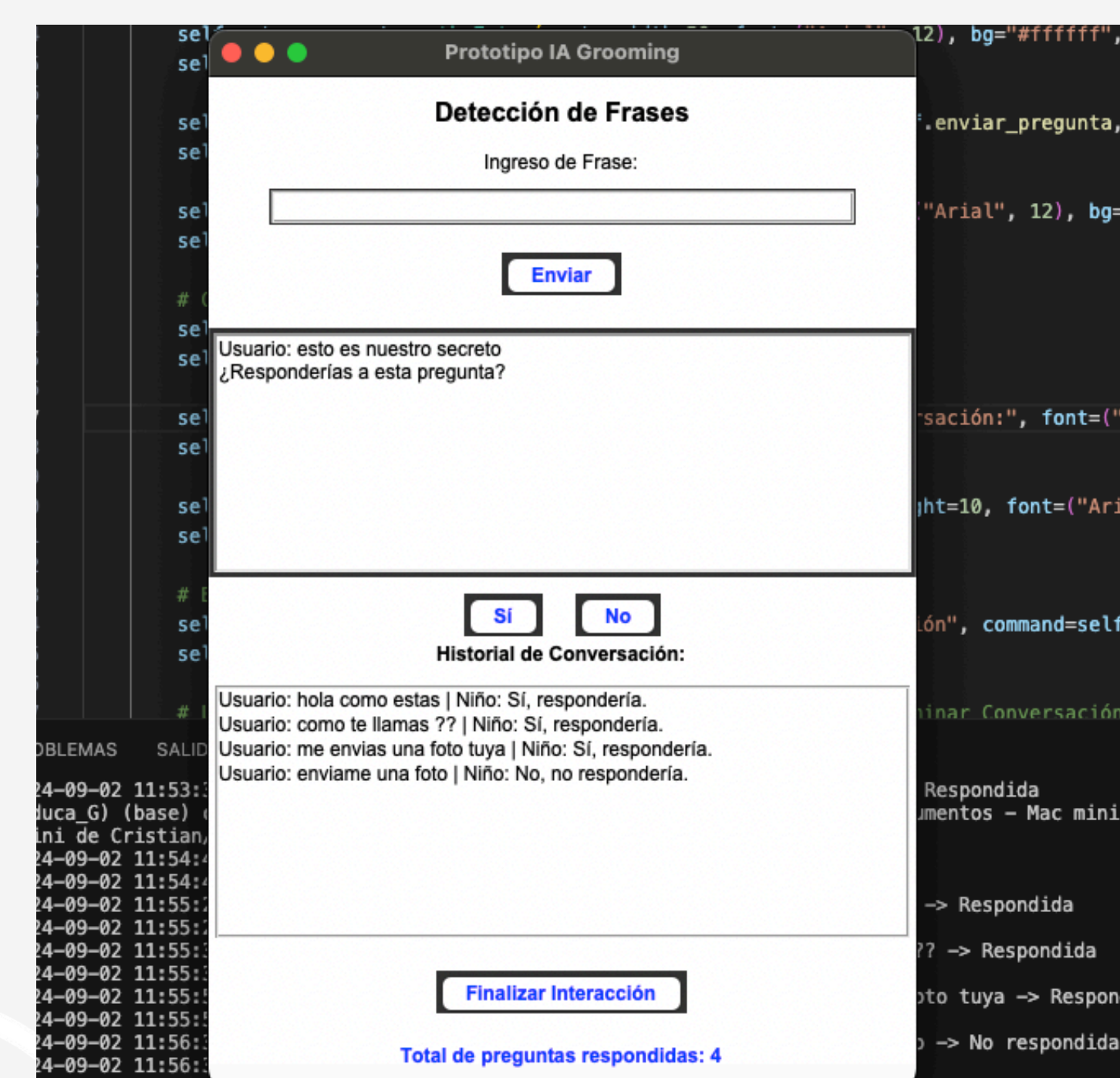
- Inches, G., & Crestani, F. (2012). Overview of the International Sexual Predator Identification Competition at PAN-2012. *En Conference and Labs of the Evaluation Forum (CLEF 2012)*
- Vogt, M., Leser, U., & Akbik, A. (2021). Early detection of sexual predators in chats. *En Zong, C., Xia, F., Li, W., & Navigli, R. (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 4985-4999). Association for Computational Linguistics.*
- Prosser, E., & Edwards, M. (2024). Helpful or harmful? Exploring the efficacy of large language models for online grooming prevention. *En Proceedings of the European Interdisciplinary Cybersecurity Conference (EICC 2024) (pp. 1-10). ACM.*

ANID/Proyecto Fondecyt Exploración 13220040

DESARROLLO DEL CASO

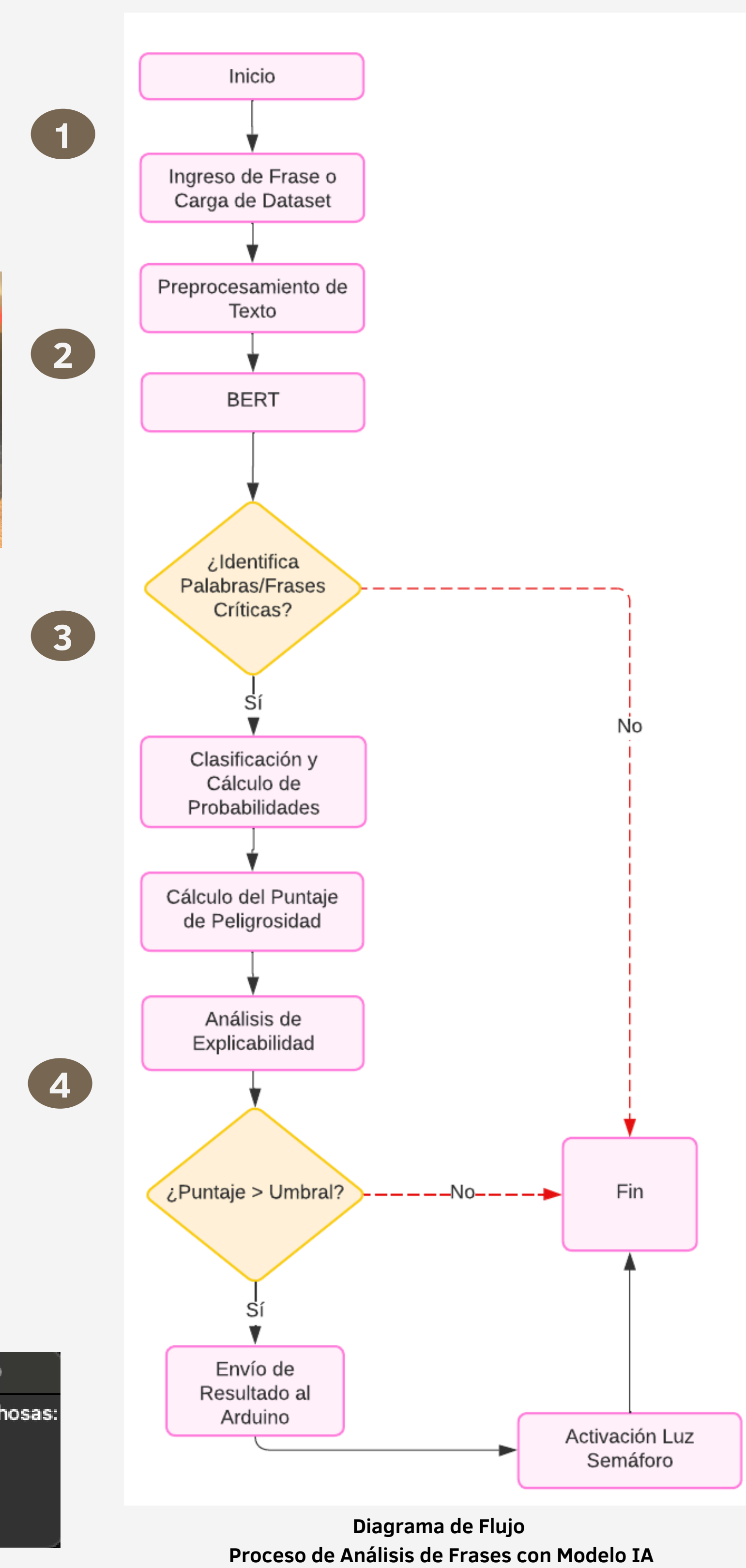
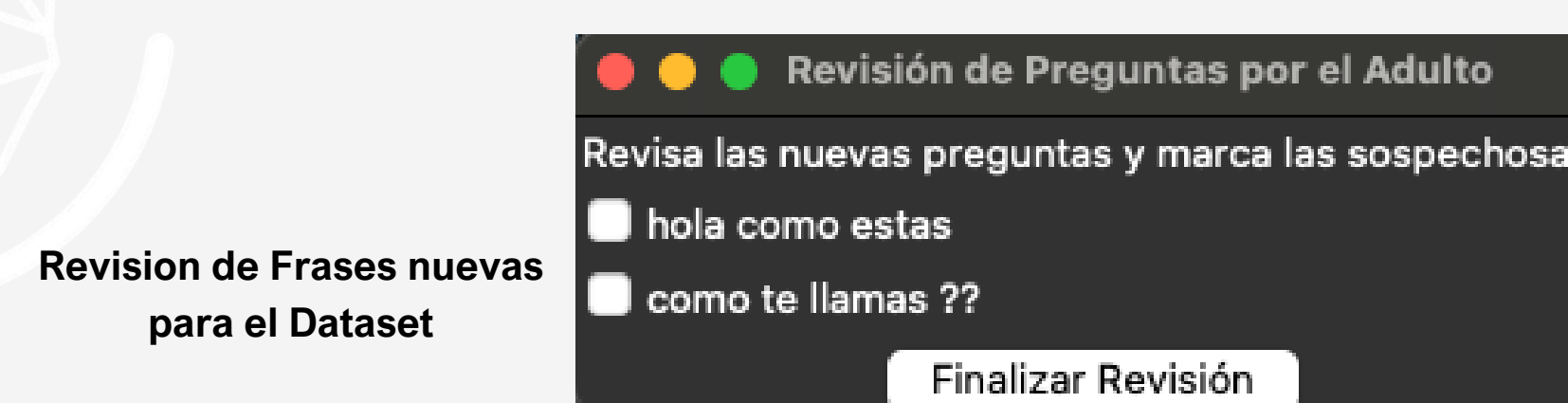
El sistema desarrollado integra técnicas avanzadas de IA para la detección automatizada de grooming en entornos educativos, siguiendo el proceso descrito a continuación:

1. Ingreso y Preprocesamiento de Datos
2. Análisis modelo IA BERT
3. Clasificación y Evaluación
4. Generación de Alertas



Plataforma Gráfica de interacción

Ingreso de Frases



RESULTADOS y DISCUSIÓN

- Implementación de un sistema de IA diseñado para detectar patrones de grooming en línea.
 - Integración de análisis lingüístico avanzado para identificar comportamientos sospechosos.
 - Creación de un entorno controlado y educativo para supervisar y operar el sistema.
 - El sistema identifica con alta precisión comportamientos de grooming que podrían pasar desapercibidos con métodos tradicionales.
 - Prepara a los menores para reconocer y evitar situaciones peligrosas, fomentando una cultura de ciberseguridad que va más allá de medidas reactivas.
- Aspectos por Desarrollar:
- Optimización de la interfaz de usuario para facilitar su uso.
 - Coordinación para realizar pruebas con estudiantes, educadores para refinar la experiencia de usuario
 - Implementación y pruebas del modelo BERT, mediante patrones lingüísticos asociados a grooming.

CONCLUSIÓN

El sistema de IA detectó con éxito patrones de grooming a través de la interacción entre el usuario y el dataset. El entorno gráfico, diseñado específicamente para un contexto educativo controlado, actúa como una solución proactiva para reforzar la ciberseguridad de los menores desde las aulas. Actualmente, se están realizando mejoras en la integración del modelo BERT, ajustándolo a diversos contextos lingüísticos, lo que facilitará la incorporación de nuevos corpus al dataset, haciendo que el modelo sea escalable. Además, se han tomado en cuenta consideraciones éticas en el desarrollo del sistema, garantizando que las decisiones automatizadas no perjudiquen a los usuarios en las pruebas que se realizarán en la comunidad educativa para la implementación de mejoras para su optimización.