UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE

“Evaluación del impacto del uso de Inteligencia Artificial para la clonación de voz en el aprendizaje, mediante el uso de la aplicación generativa de videocápsulas educacionales EduvidIA”

JOSÉ MIGUEL GONZÁLEZ BALMACEDA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Federico Meza
Profesor Correferente: Víctor Gutiérrez

Junio - 2025

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título; ☐ Tesis de Postgrado;

Título del trabajo: Evaluación del impacto del uso de Inteligencia Artificial para la clonación de voz en el aprendizaje, mediante el uso de la aplicación generativa de videocápsulas educacionales EduvidIA

Nombre del candidato(a): José Miguel González Balmaceda

Carrera / Grado: Ingeniería Civil Informática

Campus: Santiago San Joaquin **;** **Departamento**: Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Federico Meza, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

☒ El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

☐ 6 meses; ☐ 12 meses; ☐ 2 años; ☐ 3 años; ☐ 5 años; ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 28-07-2025 ; **Firma:** _____

Estudiante o Candidato(a):

Fecha: 28-07-2025 ; **Firma:** _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

A mis padres, por su apoyo a lo largo de mi formación acádemica.
A aquellos amigos que hicieron que mi estadía en la universidad fuera algo más tolerable, a pesar de que algunos no pertenecen a ella: Martín, Vicente, Anter, entre otros.
Y a todos aquellos que, de una forma u otra, me ayudaron a crecer como persona durante este período.

AGRADECIMIENTOS

Me gustaría agradecer a los profesores que, gracias a la formación que me brindaron, hicieron posible llegar hasta esta etapa y desarrollar este trabajo. En especial, al profesor guía Federico Meza, por sus consejos y la valiosa retroalimentación entregada durante el proceso.

También, agradezco a quienes colaboraron directa o indirectamente en la realización de esta memoria, incluyendo al profesor Víctor Gutiérrez y a sus estudiantes de primer año, quienes amablemente participaron como voluntarios en el estudio.

RESUMEN

Resumen— El avance de la Inteligencia Artificial ha abierto nuevas posibilidades en el ámbito educativo, especialmente frente a la necesidad de crear materiales suplementarios que refuercen el aprendizaje sin aumentar la carga laboral docente. Profesores enfrentan limitaciones de tiempo para generar recursos adicionales, lo que impacta negativamente la experiencia de los estudiantes. Este trabajo busca abordar dicha problemática mediante el uso de EduvidIA, un proyecto de la Feria de Software USM 2024, que permite generar cápsulas educativas audiovisuales a partir de documentos proporcionados por el profesor, integrando tecnologías como modelos de lenguaje y clonación de voz. El objetivo principal fue evaluar si el uso de una voz clonada, al imitar características humanas como la entonación y el ritmo, puede mejorar la comprensión y percepción del contenido educativo. Para validar esta hipótesis, se realizó un experimento controlado con estudiantes universitarios, comparando el impacto de una cápsula narrada con voz clonada versus una narrada con voz Loquendo. Los resultados mostraron mejoras significativas en la percepción de aprendizaje y naturalidad de la voz, aunque no en la comprensión objetiva. Se concluye que la clonación de voz representa una ventaja subjetiva relevante para experiencias de aprendizaje más positivas, sin perjuicio en el rendimiento académico.

Palabras Clave— Inteligencia Artificial Generativa; Aprendizaje Educativo; Clonación de voz; Cápsulas educativas; Síntesis de voz

ABSTRACT

***Abstract*—** The advancement of Artificial Intelligence has opened up new possibilities in education, especially when faced with the need to create supplementary materials that reinforce learning without increasing the teaching workload. Teachers face time constraints to generate additional resources, which negatively impacts students' experience. This work seeks to address this problem through the use of EduvidIA, a project of the USM's 2024 Feria de Software, which allows the generation of audiovisual educational capsules from documents provided by the teacher, integrating technologies such as language modeling and voice cloning. The main objective was to evaluate whether the use of a cloned voice, by imitating human characteristics such as intonation and rhythm, can improve the comprehension and perception of the educational content. To validate this hypothesis, a controlled experiment was conducted with university students, comparing the impact of a capsule narrated with a cloned voice versus one narrated with a Loquendo voice. The results showed significant improvements in the perception of learning and naturalness of the voice, although not in objective comprehension. It is concluded that voice cloning represents a relevant subjective advantage for more positive learning experiences, without detriment to academic performance.

***Keywords*—** Generative Artificial Intelligence; Educational Learning; Voice Cloning; Educational Capsules; Speech Synthesis; Speech Synthesis

ÍNDICE DE CONTENIDOS

CAPÍTULO 1: INTRODUCCIÓN

El avance de la Inteligencia Artificial y sus capacidades generativas ha generado cambios significativos en diversas áreas laborales, especialmente en la educación, que podría beneficiarse enormemente de tecnologías capaces de generar texto y material personalizado, como los LLM (Large Language Models). Estos modelos, ejemplificados por asistentes virtuales como ChatGPT, Claude y Microsoft Copilot, han demostrado ser herramientas potentes para la creación de contenido. Es relevante explorar si la Inteligencia Artificial puede facilitar el trabajo de los docentes, ya que, un estudio del Grattan Institute[1], un grupo australiano de expertos en políticas públicas, destaca que los profesores enfrentan una gran escasez de tiempo y recursos debido a la carga de tareas administrativas y pedagógicas, lo que les dificulta crear contenido adicional que permita a los estudiantes reforzar su aprendizaje de manera autónoma. Idealmente, los docentes deberían contar con herramientas que les ayuden a generar y distribuir material suplementario de manera eficiente; sin embargo, en la práctica, las limitaciones de tiempo y esfuerzo lo dificultan considerablemente.

Esta discrepancia afecta tanto a docentes como a estudiantes: los profesores no tienen los recursos necesarios para ofrecer contenido adicional, mientras que los estudiantes se ven privados de materiales de repaso fundamentales para consolidar su aprendizaje. Un estudio realizado en Arabia Saudita[2] sobre la enseñanza de la lengua inglesa revela que el uso de videos educativos, como los de plataformas tipo YouTube, puede mejorar significativamente la comprensión y el rendimiento académico de los estudiantes. Este hallazgo demuestra el impacto positivo que pueden tener los contenidos audiovisuales en la educación. Sin embargo, la creación de este tipo de materiales sigue siendo un reto para los profesores, quienes enfrentan restricciones de tiempo y recursos. En este contexto, la adopción de herramientas tecnológicas que permitan generar videos educativos de manera eficiente surge como una solución necesaria para mejorar la calidad del aprendizaje sin aumentar la carga laboral de los docentes.

El desafío de la creación de contenido adicional para el aprendizaje ha sido abordado en diversas formas en el pasado por múltiples actores, como plataformas educativas y herramientas de gestión del aprendizaje, conocidas como Learning Management Systems (LMS)[3]. Un LMS es una plataforma de software diseñada para gestionar, distribuir y evaluar actividades formativas. Estos sistemas permiten a los profesores organizar cursos en línea, asignar tareas, gestionar calificaciones y realizar un seguimiento detallado del progreso de los estudiantes. Los LMS han sido clave para facilitar el acceso al contenido educativo y, al mismo tiempo, reducir la carga administrativa de los docentes.

Entre los ejemplos más destacados de instituciones y empresas que han desarrollado recursos y herramientas para facilitar la creación y distribución de contenido educativo se encuentran Khan Academy, Coursera y Blackboard, entre otras. Estas plataformas han mejorado significativamente el acceso al aprendizaje, aunque en muchos casos lo hacen

mediante contenidos predefinidos o mecanismos generales de personalización. En el caso de Khan Academy, se ha destacado por su uso innovador de la Inteligencia Artificial para personalizar el aprendizaje de los estudiantes, ajustando el contenido y las recomendaciones a medida que los usuarios avanzan en los cursos, además de contar con el chatbot Khanmigo[4]. Este enfoque permite identificar lagunas en el conocimiento de los estudiantes y adaptar el contenido de manera dinámica, proporcionando un aprendizaje más individualizado. Coursera, por su parte, ha integrado herramientas de IA para mejorar la experiencia de evaluación automática y ofrecer comentarios en tiempo real a los estudiantes, mientras que Blackboard ha desarrollado funciones avanzadas para analizar el desempeño de los estudiantes y sugerir intervenciones educativas basadas en datos.

Para poder abordar este problema, un grupo de estudiantes de Ingeniería Civil Informática de la Universidad Técnica Federico Santa María, entre los cuales se encuentra el autor de esta memoria, han desarrollado EduvidIA en el contexto de las asignaturas Gestión de Proyectos Informáticos y Taller de Desarrollo de Proyectos Informáticos, una aplicación basada en Inteligencia Artificial que impactaría principalmente a los profesores y estudiantes. Este estudio parte de la hipótesis de que las voces clonadas, al reproducir con mayor naturalidad las características del habla humana, como el ritmo y la entonación, podrían no solo facilitar la comprensión del contenido, sino también generar una experiencia de aprendizaje más cercana y positiva para los estudiantes. EduviIA brinda un entorno de educación favorable al aprendizaje, en donde los profesores pueden generar videocápsulas de la temática que ellos deseen, tan sólo subiendo un archivo PDF, o bien, un Power Point, a partir del cual un LLM (ChatGPT 4o-mini) genera un resumen que captura los puntos clave del texto, además de un conjunto de preguntas de alternativas y verdadero o falso, que servirán para evaluar el aprendizaje de los estudiantes. Usando el resumen mencionado, se genera una videocápsula, en la cual se explican los conceptos fundamentales mediante audio narración, a la vez que visualmente se obtiene un dibujo sketch de alguna imagen relacionada con el tema, para ayudar a retener la atención de los estudiantes. En particular, la narración puede realizarse, si así el profesor lo prefiere, usando una herramienta de Inteligencia Artificial (basada en *Coqui AI*) capaz de clonar su voz, después de haber recibido una muestra por la cual guiarse, lo cual permite que los estudiantes puedan reconocer la voz ya familiar y conocida de su profesor. Los profesores se beneficiarían al tener esta herramienta facilitando la creación de contenido educativo, reduciendo su carga de trabajo y mejorando la calidad de los recursos ofrecidos. Los estudiantes se beneficiarían al tener acceso a materiales de repaso adicionales, lo que podría mejorar su comprensión y retención de la información. Además, las instituciones educativas, como beneficiarios indirectos, también experimentarían un impacto positivo a través de la mejora en la calidad del aprendizaje y la eficiencia en la creación de contenido.

Como se mencionó previamente, el profesor puede elegir ocupar su propia voz para la generación de las videocápsulas. Clonar la voz del profesor para las cápsulas educativas en EduvidIA representa una innovación significativa en la personalización del aprendizaje. Al

utilizar la voz del propio profesor, el contenido adicional no solo mantiene una conexión auténtica y familiar con los estudiantes, sino que también refuerza el estilo pedagógico individual del profesor. Esta continuidad en la voz y el tono asegura que los estudiantes reciban el material en un formato que resuena con su experiencia en clase, lo que podría facilitar una mejor percepción del aprendizaje y fomentar una experiencia más cercana, aunque su impacto directo en la comprensión requiere más investigación. Además, la clonación de voz permite una producción eficiente y consistente de contenido educativo, liberando tiempo para que el profesor se concentre en la enseñanza y la interacción directa con los estudiantes. Este enfoque combina la alta calidad de los recursos educativos con una personalización que enriquece la experiencia de aprendizaje, haciendo que cada cápsula de contenido sea un reflejo auténtico del proceso educativo del profesor.

A diferencia de plataformas como Khan Academy o Blackboard, que ofrecen contenido predefinido, EduvidIA permite la creación automática de materiales educativos personalizados a partir de los documentos que los profesores ya tienen. Esto hace que la herramienta no solo sea útil para distribuir conocimiento, sino también para adaptarlo y presentarlo de manera más atractiva para los estudiantes. Además, con la opción de generar videocápsulas narradas por la voz clonada del propio profesor, EduvidIA asegura una experiencia de aprendizaje que mantiene el estilo y el tono pedagógico del docente, algo que otras soluciones no ofrecen.

La implementación de EduvidIA no solo busca resolver un problema inmediato relacionado con la creación de contenido educativo adicional, sino que tiene el potencial de cambiar el paradigma en la enseñanza. Al automatizar la producción de materiales educativos y permitir la personalización a través de la clonación de voz, EduvidIA apunta a facilitar y personalizar la producción de materiales educativos. En el largo plazo, herramientas similares podría hacer que la enseñanza personalizada y el aprendizaje autónomo sean más accesibles, asegurando que los estudiantes reciban contenido ajustado a su ritmo y estilo de aprendizaje, aunque sin reemplazar el rol activo del docente ni resolver por sí sola todas las barreras educativas actuales.

El aspecto fundamental por evaluar en este trabajo es el impacto y diferencia percibida que tiene la clonación de voz, mediante el uso de modelos de Inteligencia Artificial, en el nivel de aprendizaje. Esta misión hace que el perfil de esta memoria posea aspectos en dos perfiles posibles: el perfil científico y el perfil de diseño o experiencia de usuarios. El primero se debe a que es de interés evaluar la hipótesis que plantea que una clonación de voz adecuada generará un mejor aprendizaje, como podría parecer intuitivo pensar, mediante un estudio que permita obtener resultados del impacto que causa. En cuanto al segundo perfil, desde la perspectiva del diseño y la experiencia del usuario, se pretende garantizar que los usuarios de EduvidIA experimenten una educación más efectiva, ya que la calidad de la voz que transmite el contenido educativo es un factor clave en el proceso de aprendizaje y su mejora constituye uno de los objetivos centrales de este estudio. Para evaluar esta hipótesis, se diseñó un experimento controlado que comparó el impacto de

una cápsula educativa con voz clonada y otra con voz Loquendo, aplicadas a estudiantes de ingeniería. Se midieron tanto los resultados objetivos de comprensión como la percepción subjetiva de aprendizaje. Para llevar a cabo adecuadamente este estudio, se seguirán la metodología definida en la memoria *"Metodología del Lab Ux USM para guiar a estudiantes y memoristas en el diseño y evaluación centrada en personas"*[5], la cual brinda un marco a seguir para el proceso de diseño y evaluación de productos de software centrados en las personas.

CAPÍTULO 2: DESARROLLO DEL EXPERIMENTO

2.1 ¿Qué es EduvidIA?

EduvidIA es una plataforma desarrollada como parte de un proyecto universitario, cuyo propósito es facilitar la creación de vídeos educativos, de forma automatizada, utilizando herramientas de inteligencia artificial. Su funcionamiento se basa en la lectura de un texto base proporcionado por el usuario, que puede tratarse de un archivo PDF o Power Point, a partir del cual genera un video compuesto por narración, imágenes y subtítulos sincronizados, y de acuerdo con la temática del archivo proporcionado.

El sistema permite elegir entre distintos tipos de voz para la narración, disponiendo así de voces clonadas previamente mediante un modelo de clonación de voz, ocupando *All Talk TTS*[6], un *fork* de Coqui AI[7]. Adicionalmente, si el profesor prefiere que el vídeo lleve su propia voz para la narración, la página permite realizar una grabación de audio, la cual será ocupada para ser imitada por la narración del vídeo. Esencialmente, esta última opción permite replicar el timbre y estilo de una voz humana real, a partir de un conjunto reducido de muestras de audio. De esta forma, EduvidIA permite generar videos con una narración más natural y personalizada, en comparación con las voces artificiales más comunes.

Además de la narración, por el lado visual, la plataforma selecciona imágenes representativas para ilustrar el contenido, almacenadas en una base de datos y poseyendo metadata que permite usarla, las cuales se integran de manera automática en el video final. Para presentar estas imágenes, el sistema ocupa la foto de una mano humana, la cual va recorriendo los contornos de la imagen en cuestión, simulando así que las está dibujando, con el objetivo de así ir construyendo la imagen de a poco, tal como se realiza en un *sketch*. Se decidió hacerlo así ya que se consideró que esto podría ayudar a retener la atención visual de los estudiantes, al ir siguiendo la mano y preguntarse que terminaría dibujando. El resultado es un recurso audiovisual que puede ser utilizado como material de apoyo para el aprendizaje, sin necesidad de conocimientos técnicos en edición de video o producción multimedia.

2.2 Objetivo del experimento

El objetivo principal de este experimento fue evaluar el impacto del tipo de voz utilizada en materiales educativos audiovisuales, comparando específicamente una voz clonada mediante inteligencia artificial con una voz sintética tradicional. Para ello, se elaboraron dos versiones de un mismo video educativo, idénticas en contenido y formato visual, pero diferenciadas únicamente en la tecnología utilizada para la narración: una con una voz

clonada mediante IA, generada a través de la plataforma EduvidIA, y otra con una voz sintética del tipo Loquendo.

Este estudio buscó explorar el efecto de estas tecnologías de voz en dos dimensiones centrales. En primer lugar, se analizó la **eficacia en el aprendizaje**, entendida como la capacidad de los estudiantes para comprender, retener y procesar el contenido entregado. Esta dimensión fue evaluada mediante un test de comprensión diseñado específicamente para el video, con el fin de cuantificar el desempeño de los participantes tras la exposición al material. En segundo lugar, se investigó la **experiencia de usuario** asociada al uso de cada tipo de voz. Esta dimensión incluye percepciones subjetivas de los estudiantes sobre aspectos como la naturalidad de la voz, su claridad, el nivel de atención que permite mantener durante el video, y la comodidad general al escucharla. La evaluación se realizó a través de una encuesta que recoge impresiones cuantitativas sobre la experiencia auditiva. Esta perspectiva es especialmente relevante, ya que una voz que genere rechazo, distracción o fatiga auditiva podría afectar indirectamente la efectividad del aprendizaje, incluso si el contenido visual o pedagógico es adecuado.

La hipótesis que motivó este experimento fue que las voces clonadas, al imitar de forma más precisa el ritmo, entonación y características naturales de una voz humana real, podrían no solo mejorar la comprensión del contenido, sino también generar una experiencia de usuario más positiva. Esto, a su vez, podría favorecer una mayor atención sostenida y una actitud más receptiva frente al material educativo.

Es importante aclarar que este estudio no busca entregar conclusiones absolutas ni generalizables sobre qué tecnología es mejor en todos los contextos. Más bien, se trata de una primera aproximación realizada en un entorno controlado, que permite observar algunas tendencias y levantar ideas sobre el posible valor del uso de voz clonada en educación. Los resultados obtenidos pueden servir como punto de partida para futuras investigaciones que profundicen en esta línea.

2.3 Diseño experimental

El estudio se llevó a cabo utilizando dos grupos distintos de estudiantes, sin asignación aleatoria estricta. Esta elección se debió a las condiciones prácticas del entorno universitario, donde no era posible controlar completamente en qué grupo debía estar cada persona.

La intervención se realizó al término de una clase regular, con el apoyo del profesor titular del curso. En ese momento, se invitó a los estudiantes a participar de forma voluntaria,

informándoles que verían un video educativo y responderían una breve evaluación después. Para dividir a los participantes, se habilitaron dos salas cercanas: la misma sala donde se dictó la clase y una sala contigua. Los estudiantes podían elegir libremente en cuál quedarse, sin saber qué versión del video se mostraría en cada una. Para lograr una distribución más equilibrada, se sugirió de forma general que algunos consideraran cambiarse de sala, pero esto no fue obligatorio para nadie.

Este tipo de organización hace que no se pueda asegurar que ambos grupos fueran exactamente equivalentes desde un inicio. Factores como la motivación personal, el interés previo por la tecnología o simplemente la curiosidad por cambiar de sala podrían haber influido en la elección, afectando ligeramente los resultados. Por eso, los hallazgos deben interpretarse con cierta precaución, teniendo en cuenta que hubo variables que no se pudieron controlar del todo.

En el grupo que permaneció en la sala original, se proyectó un video con narración generada mediante una voz clonada con inteligencia artificial, basada en la voz real del mismo profesor que acababa de impartir la clase. Esta elección no fue casual: se buscó intencionadamente utilizar una voz familiar y reconocible para los estudiantes, con el objetivo de evaluar si dicha cercanía influía positivamente en la experiencia auditiva y el aprendizaje. En contraste, al grupo que se trasladó a la sala contigua se le mostró un video equivalente con narración generada mediante el motor de síntesis de voz Loquendo, conocido por su tono más robótico y genérico.

Para la evaluación se elaboraron dos versiones de un mismo video educativo centrado en la Revolución de Taiping, un conflicto político-religioso ocurrido en China durante el siglo XIX. La elección de esta temática respondió a dos criterios fundamentales. Primero, se trató de un acontecimiento históricamente relevante y llamativo, pero poco conocido por el público objetivo (estudiantes de primer año de ingeniería de una universidad chilena), lo que permitía reducir al mínimo la influencia de conocimientos previos en los resultados. Segundo, su estructura narrativa lineal y clara facilitaba la adaptación del contenido a un formato audiovisual coherente, atractivo y comprensible.

Ambas versiones del video fueron cuidadosamente diseñadas para ser equivalentes en todos sus elementos, salvo uno: la voz que explica el evento histórico. Se mantuvo constante el guion narrativo, la secuencia de imágenes y animaciones, el ritmo de exposición y la duración total del video, de aproximadamente 10 minutos. Esta estrategia de control permitió asegurar que cualquier diferencia en los resultados se debiera exclusivamente al tipo de voz utilizada.

La Versión A, mostrada al grupo que permaneció en la sala original, utilizó una voz clonada mediante inteligencia artificial generada por la plataforma EduvidIA. Esta voz fue modelada a partir del timbre y estilo del mismo docente que había impartido la clase inmediatamente anterior, con el fin de aprovechar la familiaridad auditiva y generar una sensación de cercanía, naturalidad y fluidez expresiva.

La Versión B, en cambio, fue presentada al grupo que se trasladó a una sala contigua. En este caso, se utilizó una narración sintetizada con el motor Loquendo, caracterizado por una entonación robótica, baja modulación emocional y una dicción mecánica. Aunque funcional desde una perspectiva informativa, este tipo de voz suele percibirse como monótono y artificial, lo que podría incidir en la experiencia del espectador.

Ambas versiones se encuentran disponibles en Youtube, los enlaces respectivos para verlos se adjuntan en los Anexos.

Esta diferenciación buscó simular dos escenarios contrastantes en el uso de tecnologías de voz, permitiendo evaluar su influencia tanto en el aprendizaje como en la percepción de los estudiantes frente a un mismo contenido.

El diseño experimental de este estudio tuvo como objetivo principal evaluar el efecto del tipo de voz empleada en un video educativo sobre dos dimensiones clave: el aprendizaje (desempeño cognitivo) y la experiencia del usuario (percepción subjetiva). Para ello, se definieron claramente una variable independiente y dos variables dependientes, las cuales estructuran el análisis de los datos obtenidos.

La variable independiente correspondió al tipo de voz utilizada en la narración del video, con dos posibilidades. La primera corresponde a una voz clonada con inteligencia artificial (EduvidIA), que buscaba replicar la entonación natural y el timbre del docente, generando un efecto de cercanía y realismo. Por otro lado, está la voz sintetizada tradicional (Loquendo), caracterizada por un tono robótica y una calidad menos expresiva y monótona.

Estas voces fueron aplicadas sobre un mismo video base, cuyo contenido, duración y elementos visuales se mantuvieron constantes. Esta estrategia de control permitió aislar la influencia del tipo de voz sobre las variables dependientes, asegurando la validez interna del experimento.

Por otro lado, las variables dependientes fueron las siguientes:

1. Comprensión del contenido: evaluada a través de un test de opción múltiple, diseñado para medir el grado de retención y entendimiento de la información presentada. El instrumento incluyó preguntas sobre hechos históricos relevantes, relaciones causales, actores principales, cronología y consecuencias del conflicto. Esta medición objetiva permitió cuantificar el aprendizaje efectivo generado por cada versión del video.

2. Percepción de la experiencia auditiva: medida mediante una encuesta de carácter subjetivo, orientada a recoger la opinión de los participantes sobre la calidad sonora del material. Se consideraron dimensiones como la naturalidad y claridad de la voz, el nivel de concentración que permite la voz en cuestión, y el agrado general percibido en la presentación.

Ambas dimensiones fueron seleccionadas por su pertinencia en el contexto educativo: mientras que la comprensión permite evaluar el impacto cognitivo del contenido, la percepción auditiva aporta información sobre cómo la presentación influye en la atención, motivación e involucramiento de los estudiantes. En conjunto, estos indicadores permiten estimar de forma más completa el valor educativo de las tecnologías de síntesis y clonación de voz aplicadas a contextos de enseñanza-aprendizaje.

Como se mencionó previamente, la población objetivo del presente estudio correspondió a estudiantes de primer año de ingeniería de una universidad chilena, específicamente aquellos inscritos en una asignatura común del plan de estudios. Se trató de un grupo con formación técnica, familiarizado con entornos digitales, pero sin conocimientos previos específicos sobre el tema abordado en el video educativo. Esta elección respondió tanto a criterios de accesibilidad como a la intención de evaluar el impacto de las tecnologías de voz en un público universitario con alto nivel de alfabetización digital, pero sin formación especializada en historia.

La muestra estuvo compuesta por un total de 37 estudiantes que participaron de forma voluntaria en la actividad. De estos, 21 visualizaron la versión del video narrada con voz clonada mediante inteligencia artificial, mientras que los 16 restantes vieron la versión con voz sintetizada tradicional estilo Loquendo. La asignación a cada grupo fue determinada por la sala en la que se encontraban al momento de realizarse la intervención, lo cual, si bien no constituyó un muestreo aleatorio estricto, permitió mantener condiciones naturales de aula y minimizar el sesgo de selección. No se registraron casos de estudiantes que abandonaran la actividad antes de completarla.

La intervención se llevó a cabo durante una clase presencial, en coordinación con el profesor titular de la asignatura. Previo a la actividad, se informó a los estudiantes que participarían en una experiencia de aprendizaje experimental, sin especificar el objetivo exacto del estudio, con el fin de evitar sesgos en sus respuestas. La clase se dividió en dos grupos: uno permaneció en la sala original, mientras que el otro fue conducido a una sala contigua, ambas equipadas con condiciones audiovisuales equivalentes.

Para cada grupo se proyectó una de las dos versiones del video educativo sobre la Revolución de Taiping, asegurándose que el entorno fuera propicio para la concentración (sin interrupciones externas y con buena calidad de audio e imagen). Inmediatamente después de ver el video, los estudiantes respondieron dos instrumentos, que venían unidos como una sola encuesta de evaluación: un test de opción múltiple diseñado para medir la comprensión del contenido, y una encuesta de percepción orientada a recoger sus impresiones respecto a la experiencia auditiva. Las preguntas realizadas a los participantes y su respectivo formato se pueden encontrar en los Anexos.

Ambos instrumentos mencionados fueron completados de forma individual y anónima, en papel y en el mismo espacio donde se visualizó el video. El tiempo total destinado a la actividad, incluyendo la visualización, la evaluación y la encuesta, fue de aproximadamente 25 minutos. Los formularios fueron recogidos al finalizar, y los datos

fueron posteriormente digitalizados para su análisis cuantitativo, almacenando así las respuestas en un Excel, con el cual, utilizando un *script* de Python, se realizó un análisis y visualización de los datos reunidos.

CAPÍTULO 3: RESULTADOS Y ANÁLISIS

3.1 Desempeño general en la evaluación

Con el objetivo de evaluar el impacto del tipo de voz utilizada en videos educativos, ya sea clonada mediante IA o sintetizada con tecnología tradicional (Loquendo), se diseñó una experiencia controlada en la que dos grupos de estudiantes visualizaron distintas versiones de una misma cápsula informativa, para luego responder una evaluación. Este capítulo presenta los principales resultados obtenidos a partir de los instrumentos aplicados, organizados en función del tipo de información recogida: primero se revisan los puntajes de la evaluación objetiva (preguntas 1 a 6), seguidos por las percepciones y valoraciones subjetivas expresadas por los participantes (preguntas 7 a 12). Las visualizaciones de datos que acompañan este análisis permiten ilustrar las diferencias y similitudes entre ambos grupos a lo largo de las distintas dimensiones estudiadas.

Con el objetivo de observar si el tipo de voz influía en el desempeño global de los estudiantes, se compararon los puntajes totales obtenidos por cada grupo en la evaluación de comprensión. En la Figura 1 se presenta un boxplot que ilustra la distribución de dichos puntajes para ambos grupos.

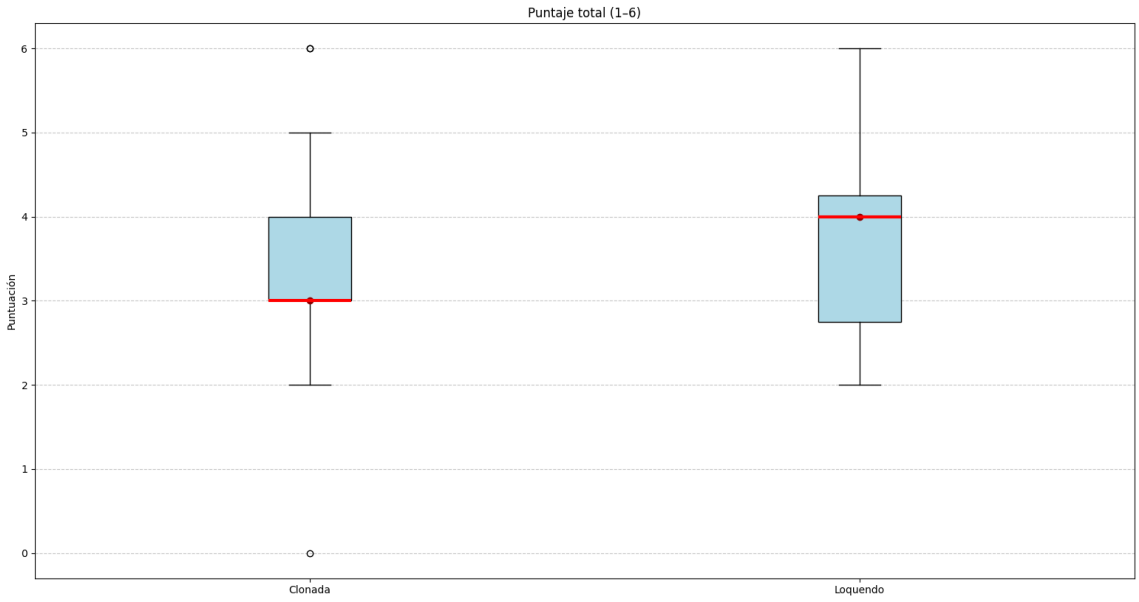

Figura 1. Diagrama de caja de los puntajes obtenidos por ambos grupos.

Fuente: Elaboración propia

El grupo que visualizó el video con voz clonada obtuvo una media de 3,29 puntos (desviación estándar = 1,42), mientras que el grupo que visualizó la versión con voz

Loquendo alcanzó una media ligeramente superior de 3,63 puntos (desviación estándar = 1,26).

Para evaluar si esta diferencia era estadísticamente significativa, se aplicó una prueba *t* para muestras independientes. Esta prueba se utiliza comúnmente para comparar las medias de dos grupos distintos y determinar si la diferencia observada entre ellos puede atribuirse a algo más que al azar. El estadístico *t* obtenido en este caso fue -0,77, con 35 grados de libertad.

El resultado de la prueba incluye también un valor *p*, que representa la probabilidad de obtener una diferencia igual o mayor a la observada si en realidad no existiera una diferencia real entre los grupos, es decir, si la hipótesis nula fuera cierta. En este análisis, el *p*-valor fue de 0,447, lo cual está muy por encima del umbral convencional de 0,05. Esto indica que la diferencia observada no es estadísticamente significativa, y, por tanto, no se puede afirmar que el tipo de voz haya generado un impacto real en el desempeño global de los estudiantes.

Estos resultados sugieren que, en términos generales, el tipo de voz utilizada en la narración del video no tuvo un impacto significativo en la comprensión global del contenido. No obstante, esto no descarta la posibilidad de que existan diferencias relevantes a nivel de preguntas específicas o en la percepción subjetiva de la experiencia, lo cual será explorado a continuación.

La ya mencionada Figura 1 permite observar con mayor detalle la distribución de puntajes dentro de cada grupo. Se aprecia que ambos grupos presentan promedios similares, aunque las medianas difieren: el grupo con voz Loquendo alcanza una mediana de 4, mientras que el grupo con voz clonada presenta una mediana de 3, en una escala de 0 a 6. La dispersión de los datos es algo mayor en el grupo con voz clonada, lo que se refleja en una caja más ancha y en la presencia de un valor atípico (outlier) en el extremo inferior, correspondiente a un puntaje de 0. Por el contrario, el grupo Loquendo muestra una distribución más concentrada, aunque también presenta un valor extremo alto (puntaje 6). Estas observaciones refuerzan el resultado estadístico previamente señalado: si bien existen algunas diferencias en la forma de la distribución, no se evidencia una diferencia sustantiva sobre el rendimiento objetivo de los estudiantes entre ambos grupos, que pueda atribuirse con significancia al tipo de voz utilizada.

Luego de presentar los puntajes totales obtenidos por cada grupo, resulta útil examinar el desempeño en cada una de las preguntas de la evaluación. De esta manera, se busca identificar no solo el nivel general de comprensión, sino también los aspectos específicos del contenido que fueron mejor o peor entendidos, así como posibles diferencias entre los grupos que podrían asociarse con el tipo de voz utilizada en el video. A continuación, se presenta la distribución de respuestas para las preguntas 1 a 6, las cuales abordan conceptos centrales de la cápsula educativa.

Pregunta 1: ¿Cuál fue una de las principales causas de la Rebelión Taiping?

- a) La expansión territorial de los Qing.
- b) La mejora de las relaciones comerciales con Europa.
- c) La invasión de potencias extranjeras.
- d) La sobrepoblación y la pobreza extrema en el sur de China.

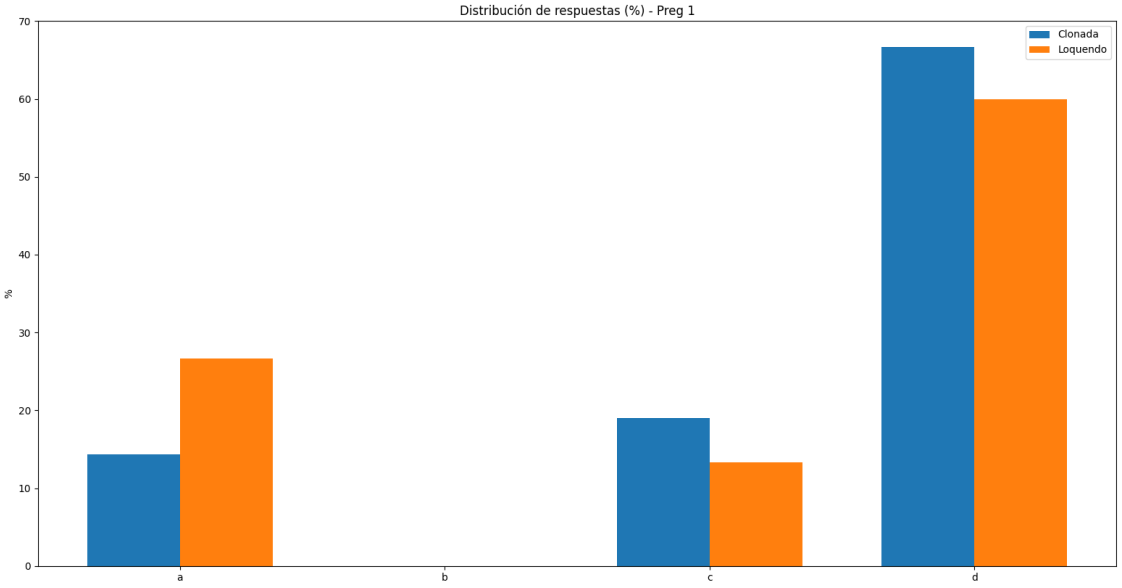

Figura 2. Distribución porcentual de respuestas por grupo – Pregunta 1.

Fuente: Elaboración propia

En esta pregunta, la alternativa correcta era la d), la cual fue seleccionada por el 66,7% del grupo con voz clonada y el 60,0% del grupo Loquendo. Si bien ambos grupos identificaron mayoritariamente la respuesta correcta, el grupo con voz clonada presentó un desempeño ligeramente superior.

Las respuestas incorrectas se distribuyeron de forma distinta entre ambos grupos. La opción a), “expansión territorial”, un distractor plausible desde el punto de vista histórico, fue seleccionada por el 26,7% del grupo Loquendo, frente a un 14,3% del grupo con voz clonada. La opción c), “invasión extranjera”, fue elegida en menor proporción, con un 19,0% en el grupo clonada y un 13,3% en el grupo Loquendo. Ningún participante eligió la opción b).

Estas diferencias sugieren que, aunque el nivel de comprensión general fue bueno en ambos grupos, el grupo con voz clonada mostró una menor tendencia a confundirse con distractores, lo que podría estar relacionado con una mayor claridad o énfasis en la narración de ciertos detalles relevantes del contenido.

Pregunta 2: ¿Qué proclamó Hong Xiuquan tras sus visiones divinas?

- a) Que él era el líder de una nueva dinastía Qing.
- b) Que se convertiría en el emperador de China.
- c) Que era el hermano menor de Jesucristo y debía derrocar a la dinastía Qing.
- d) Que China debería aliarse con las potencias extranjeras.

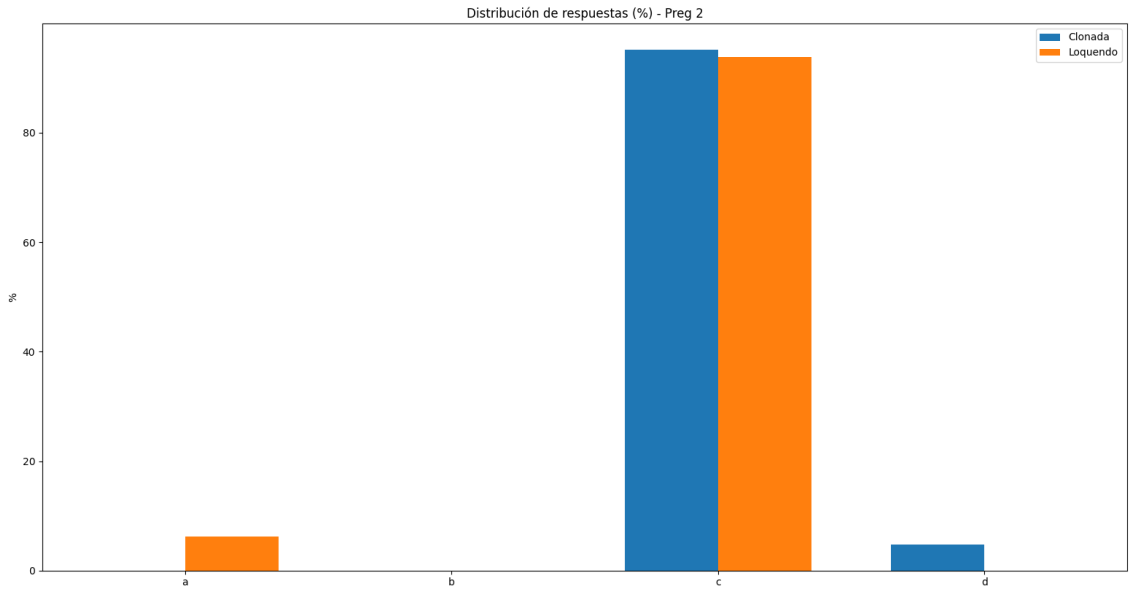

Figura 3. Distribución porcentual de respuestas por grupo – Pregunta 2.

Fuente: Elaboración propia

Esta fue la pregunta con mayor nivel de aciertos en toda la evaluación. La alternativa correcta, c), fue seleccionada por el 95,2% del grupo con voz clonada y el 93,8% del grupo Loquendo, mostrando un desempeño casi idéntico y ampliamente satisfactorio.

Las respuestas incorrectas fueron prácticamente inexistentes. Solo un participante del grupo Loquendo eligió la opción a), y uno del grupo clonada seleccionó la opción d). La opción b) no fue seleccionada por ningún estudiante.

Estos resultados indican que el contenido relacionado con la proclamación de Hong Xiuquan fue comprendido de forma clara por la gran mayoría de los participantes, sin diferencias relevantes entre los grupos. Es posible que la naturaleza llamativa o única del contenido, una proclamación religiosa radical, haya contribuido a su fácil recordación, independientemente del tipo de voz utilizada.

Pregunta 3: ¿Qué ideología social propuso el Reino Celestial Taiping?

- a) Una sociedad jerárquica basada en el confucianismo.
- b) Una sociedad basada en el orden y la tradición militar.
- c) Una sociedad influenciada por el socialismo europeo.
- d) Una sociedad igualitaria con redistribución de tierras.

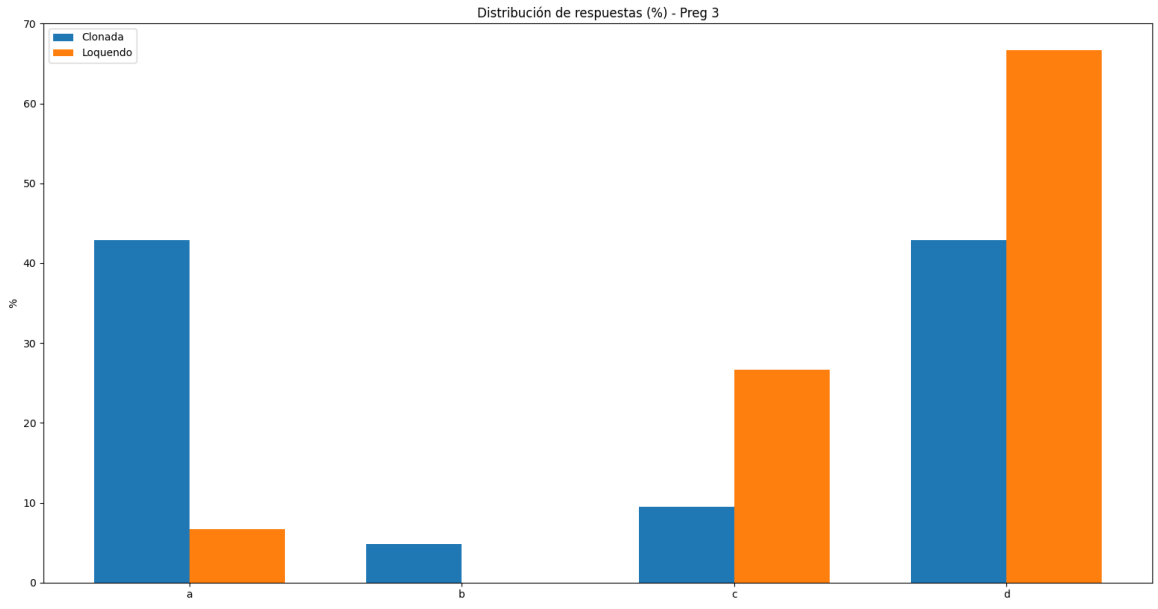

Figura 4. Distribución porcentual de respuestas por grupo – Pregunta 3.

Fuente: Elaboración propia

Esta pregunta mostró una distribución de respuestas mucho más dispersa que las anteriores, evidenciando mayor dificultad entre los participantes, especialmente en el

grupo con voz clonada. La alternativa correcta, d), fue seleccionada por el 42,9% de este grupo, mientras que en el grupo Loquendo alcanzó un 66,7%, lo que representa una diferencia de más de 20 puntos porcentuales en favor de este último.

Curiosamente, se puede observar que exactamente el mismo porcentaje del grupo clonada (42,9%) eligió la opción a), que es incorrecta pero posiblemente confundida por su asociación con estructuras sociales tradicionales en China. Esta coincidencia refleja una división equitativa entre la respuesta correcta y un distractor verosímil dentro de este grupo, lo que podría indicar ambigüedad en la comprensión del contenido. En contraste, la opción a) fue casi inexistente en el grupo Loquendo, con un 6,7% de preferencia.

Las opciones b) y c) recibieron pocos votos en ambos grupos, aunque un 26,7% del grupo Loquendo eligió la opción c), lo cual podría sugerir una confusión con ideologías modernas de igualdad social. En conjunto, estos resultados indican que la formulación de esta pregunta generó ambigüedad en una parte considerable del grupo con voz clonada, donde hubo una distribución equitativa entre la respuesta correcta y una distracción plausible. Esto sugiere que, en este ítem específico, el tipo de voz pudo haber tenido un leve impacto en la claridad con la que se transmitió el contenido ideológico del movimiento Taiping.

Pregunta 4: ¿Qué grupo militar fue clave para la derrota de los Taiping?

- a) El Ejército Xiang, liderado por Zeng Guofan.
- b) El Ejército Mandarín.
- c) El Ejército Nacional Chino.
- d) El Ejército de la Reforma.

En esta pregunta, la mayoría de los participantes identificó correctamente “El Ejército Xiang, liderado por Zeng Guofan”, como el grupo militar clave en la derrota de los Taiping. La opción correcta a) fue seleccionada por el 76,2% del grupo con voz clonada y el 87,5% del grupo Loquendo, siendo esta una de las preguntas con mayor tasa de aciertos en ambas grupos, similar al caso de la pregunta 2.

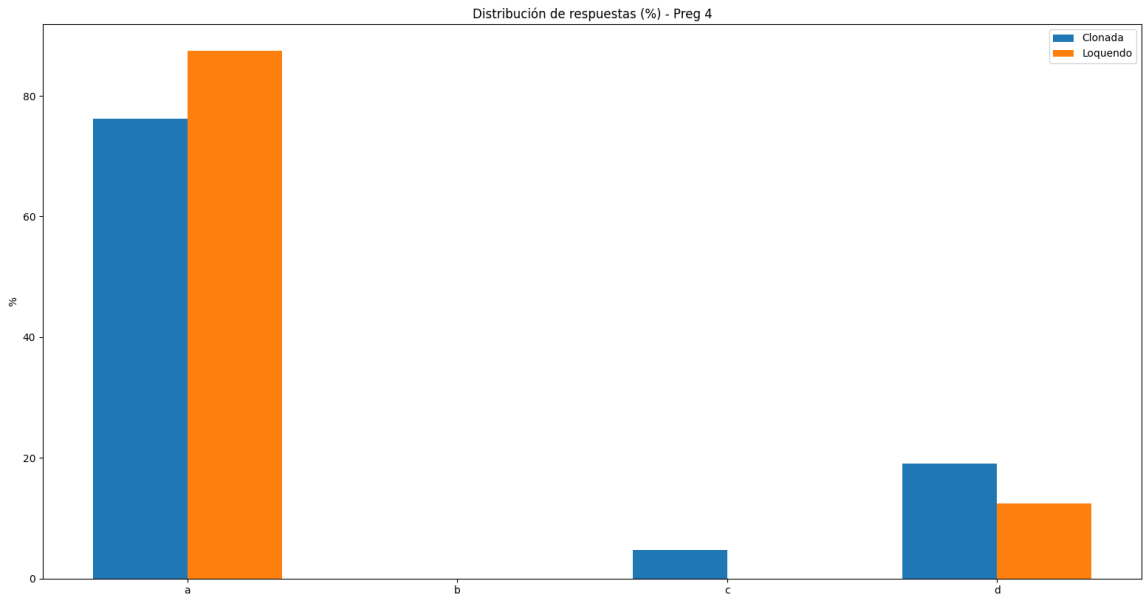

Figura 5. Distribución porcentual de respuestas por grupo – Pregunta 4.

Fuente: Elaboración propia

Las opciones incorrectas recibieron una proporción menor de respuestas. La alternativa d), “Ejército de la Reforma”, fue la segunda más elegida, con un 19,0% en el grupo clonada y un 12,5% en Loquendo, lo que sugiere que algunos estudiantes pudieron haberla confundido por su tono aparentemente verosímil, aunque no corresponde al contexto histórico de la Rebelión Taiping.

Finalemente, las opciones intermedias, b) y c) prácticamente no fueron seleccionadas, lo cual indica que los participantes lograron descartar aquellas respuestas con menor plausibilidad. En resumen, esta pregunta evidenció un buen nivel de comprensión por parte de ambos grupos, aunque con un desempeño ligeramente superior en el grupo Loquendo. Aun así, la diferencia entre ambos grupos no es lo suficientemente pronunciada como para sugerir una influencia clara del tipo de voz utilizada.

Pregunta 5: ¿Qué motivó la intervención de las potencias extranjeras en la Rebelión Taiping?

- a) La preocupación por la inestabilidad de China y el impacto en los intereses comerciales.
- b) El interés en las riquezas que los Taiping podrían ofrecer.
- c) El apoyo al cristianismo promovido por los Taiping.
- d) El deseo de controlar el comercio de opio.

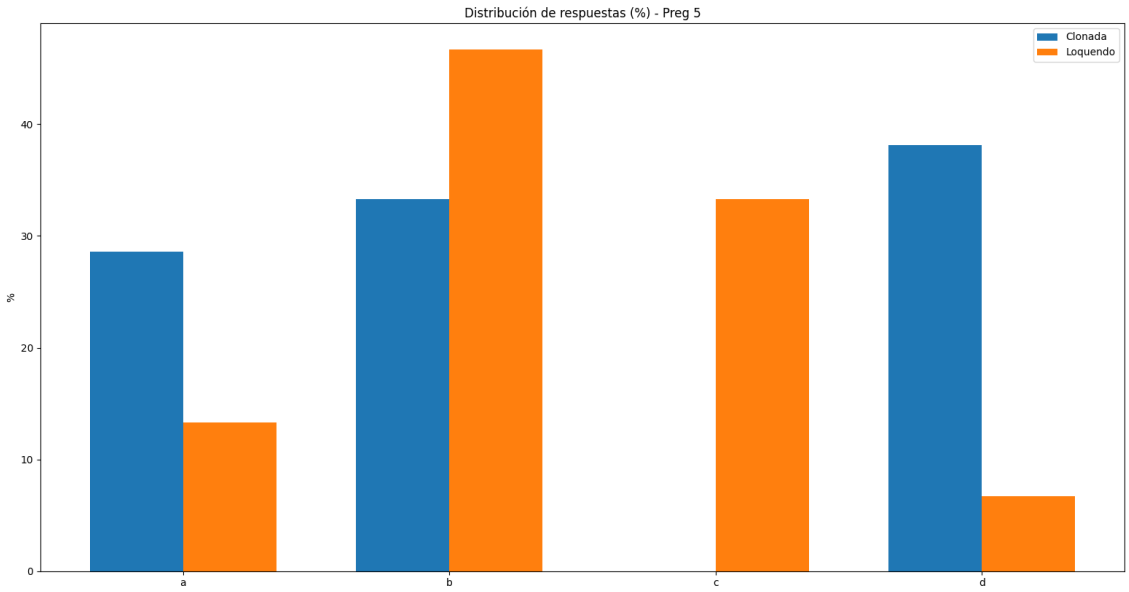

Figura 6. Distribución porcentual de respuestas por grupo – Pregunta 5.

Fuente: Elaboración propia

Esta fue una de las preguntas con mayor dispersión en las respuestas, lo que indica que generó confusión en ambos grupos. La alternativa correcta, b), “la preocupación por la inestabilidad de China y el impacto en los intereses comerciales”, fue seleccionada solo por el 33,3% del grupo con voz clonada y el 46,7% del grupo Loquendo.

Llama la atención que un número significativo de participantes eligió respuestas incorrectas pero plausibles. En el grupo con voz clonada, la opción d), “el deseo de controlar el comercio de opio”, fue la más seleccionada (38,1%), lo que sugiere una posible interferencia con conocimientos previos sobre otras intervenciones extranjeras en

China, como las Guerras del Opio, lo cual resulta plausible ya que, como se mostrará más adelante en la pregunta 7, el grupo con voz clonada reportó ser más familiar al tema presentado. Por otro lado, este patrón no se repitió con la misma intensidad en el grupo Loquendo, donde solo el 6,7% eligió esta opción.

Por otro lado, un 33,3% del grupo Loquendo optó por la opción c), que hacía referencia al cristianismo promovido por los Taiping, posiblemente influenciados por la figura religiosa de Hong Xiuquan. Sin embargo, históricamente, las potencias occidentales desconfiaban del movimiento Taiping pese a su uso de símbolos cristianos.

En suma, esta pregunta evidenció una comprensión más débil del contexto geopolítico de la intervención extranjera, con una alta tasa de error en ambos grupos. No obstante, el grupo Loquendo mostró una mayor proporción de respuestas correctas, aunque también una tendencia a elegir una opción incorrecta que es distinta a la proporción obtenida del grupo clonada, lo que podría reflejar diferencias en cómo fue interpretado el contenido según el tipo de voz.

Pregunta 6: ¿Qué fue uno de los legados sociales de la Rebelión Taiping?

- a) La consolidación del confucianismo como doctrina estatal.
- b) La reforma agraria y la redistribución de tierras.
- c) La completa adopción del cristianismo en todo China.
- d) La expansión del comercio internacional.

Esta fue una de las preguntas con mayor tasa de error en ambos grupos. La opción correcta, b), “la reforma agraria y la redistribución de tierras”, fue seleccionada por apenas un 14,3% de los participantes con voz clonada y un 20,0% de los del grupo Loquendo. Esto sugiere una débil retención o comprensión del contenido vinculado a los aspectos sociales del Reino Celestial Taiping.

La mayoría de los participantes optó por la alternativa a), “la consolidación del confucianismo como doctrina estatal”, con un 57,1% en el grupo clonada y 33,3% en el grupo Loquendo. Esta elección es llamativa, dado que va en dirección opuesta a lo explicado en el video: el movimiento Taiping se oponía activamente al confucianismo tradicional, promoviendo en su lugar reformas igualitarias inspiradas en su propia interpretación religiosa.

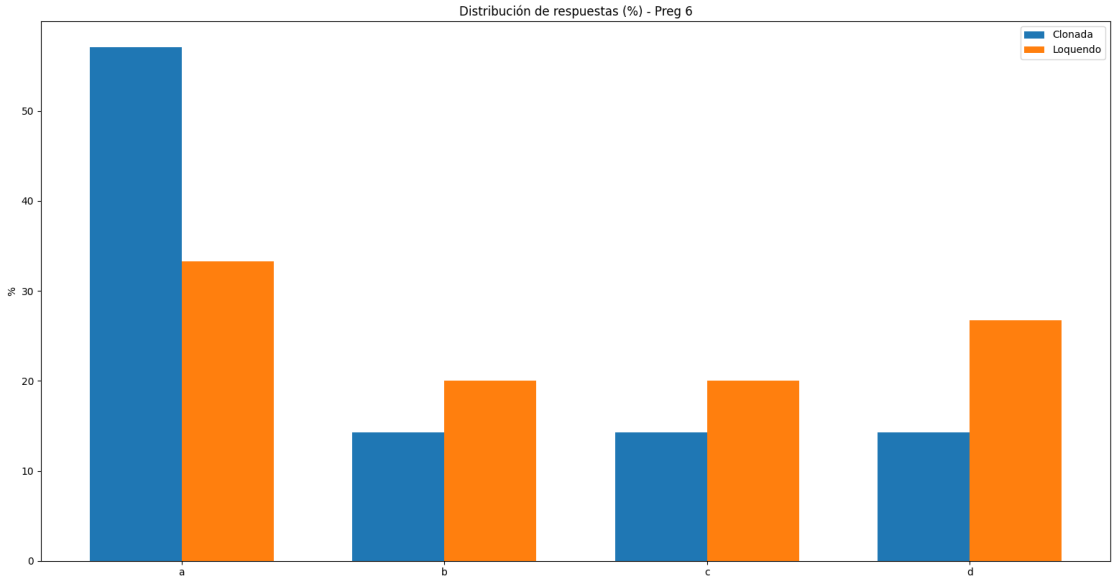

Figura 7. Distribución porcentual de respuestas por grupo – Pregunta 6.

Fuente: Elaboración propia

Este hecho podría indicar que, en ausencia de una asociación clara con un legado social alternativo, los participantes tendieron a elegir la opción que les parecía más tradicional o familiar en el contexto chino. La selección casi uniforme del resto de las opciones incorrectas (opciones c y d) también sugiere una distribución bastante aleatoria, sin una respuesta dominante clara aparte de la errónea opción a. En conclusión, esta pregunta no sólo fue mal respondida por una mayoría, sino que mostró un sesgo hacia conceptos que contradicen el contenido del video. Esta tendencia fue más marcada en el grupo clonada, lo que podría indicar que el tipo de voz no logró transmitir con suficiente énfasis este punto clave del material.

3.2 Percepción subjetiva de la experiencia de aprendizaje

Si bien las primeras seis preguntas se centraron en medir la comprensión del contenido presentado en el video, el resto de la encuesta se orientó a explorar aspectos subjetivos relacionados con la experiencia de aprendizaje. En las preguntas 7 a 12 se abordan elementos como el grado de familiaridad previa con el tema, la percepción del aprendizaje logrado, la valoración del tipo de voz utilizada y la disposición a utilizar este tipo de recursos educativos en el futuro. A continuación, se presentan y analizan los resultados

obtenidos en estas dimensiones, junto con las diferencias observadas entre ambos grupos.

Pregunta 7: ¿Qué tan familiarizado/a estabas con el tema antes de ver el video?

- a) Nada familiarizado/a
- b) Algo familiarizado/a
- c) Bastante familiarizado/a

Esta pregunta buscaba establecer el grado de conocimiento previo que los participantes tenían sobre la Rebelión Taiping antes de exponerse al contenido audiovisual. En ambos grupos, la opción a), "nada familiarizado/a", fue la más seleccionada, alcanzando un 71,4% en el grupo con voz clonada y un 93,3% en el grupo Loquendo. Esto indica que la gran mayoría de los estudiantes no poseía conocimientos previos significativos sobre el tema.

Sin embargo, se observan diferencias interesantes entre los grupos. En el grupo con voz clonada, un 19,0% declaró estar "algo familiarizado/a" y un 9,5% "bastante familiarizado/a", mientras que en el grupo Loquendo estas dos opciones fueron prácticamente inexistentes, con solo un 6,7% seleccionando "bastante familiarizado/a" y ningún estudiante optando por la alternativa intermedia. Esto podría reflejar una mayor diversidad de niveles de conocimiento previo en el grupo clonada, o bien una percepción más crítica o conservadora del propio nivel de familiaridad en el grupo Loquendo.

En conjunto, los resultados reafirman que la temática del video era poco conocida por la mayoría del público objetivo, lo cual refuerza la pertinencia de haber elegido un tema histórico poco tratado, en el contexto de conocimiento general sobre historia universal en Chile. También sugiere que las diferencias de rendimiento observadas en las preguntas de contenido no pueden atribuirse al conocimiento previo, dado que ambas condiciones partieron, en general, desde una base similarmente baja de familiaridad con el tema.

Pregunta 8: En una escala del 1 al 10, ¿qué tanto sientes que aprendiste sobre el tema? (Nada → 1 | Mucho → 10)

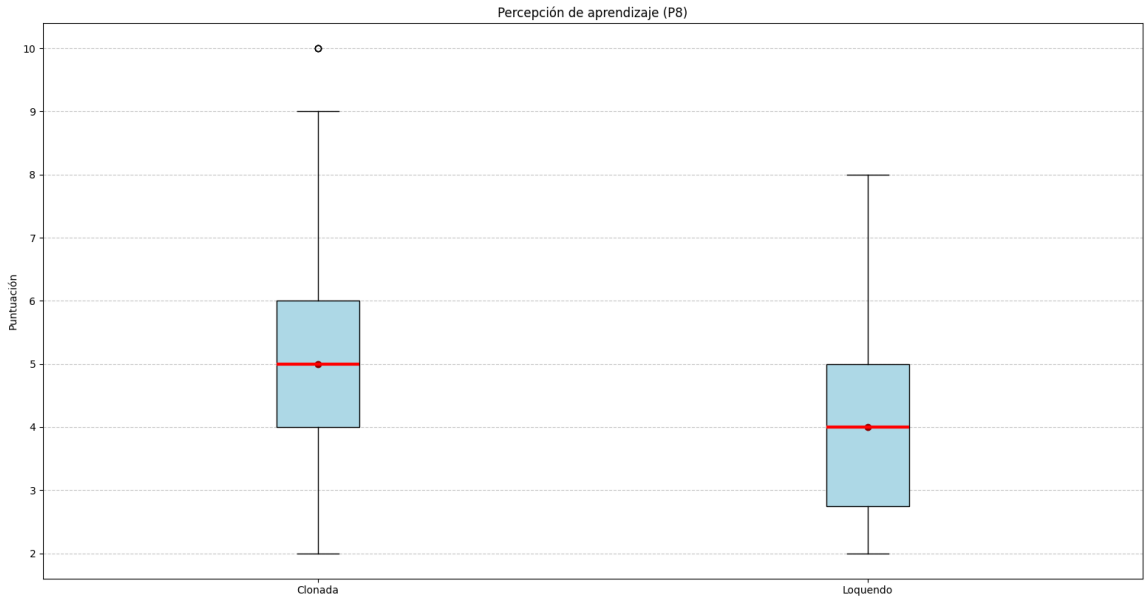

Figura 8. Distribución de autoevaluación del aprendizaje percibido – Pregunta 8.

Fuente: Elaboración propia

Esta pregunta tenía como objetivo recoger la percepción subjetiva de aprendizaje que los participantes experimentaron tras ver el video. A diferencia de las preguntas anteriores, que evaluaban conocimientos específicos, esta pregunta y las siguientes buscaban capturar la impresión general de cuánto se sintieron beneficiados por el material, en términos de comprensión y adquisición de contenido.

Los resultados muestran una diferencia clara entre los grupos. El grupo con voz clonada obtuvo una media de 5,57 (desviación estándar = 2,58), mientras que el grupo Loquendo promedió 4,00 (desviación estándar = 1,83). Esta diferencia fue estadísticamente significativa (t = 2,1681; p = 0,0371), lo que indica que no se trata de una variación aleatoria, sino de un efecto atribuible con alta probabilidad a la variabilidad de la voz usada en el vídeo.

El boxplot incluido, correspondiente a la Figura 8, permite visualizar mejor la dispersión y distribución de las respuestas. Se observa que, aunque ambos grupos presentan respuestas variadas, el grupo clonada tiende a concentrarse más hacia valores intermedios y altos, mientras que el grupo Loquendo muestra una mayor proporción de evaluaciones bajas, reflejando una percepción más modesta del aprendizaje alcanzado.

Este resultado es especialmente relevante, ya que sugiere que el uso de una voz clonada puede tener un efecto positivo en la percepción de aprendizaje del estudiante, aunque no necesariamente en el rendimiento objetivo, como se observó en algunas preguntas de contenido y en el promedio de puntaje obtenido por cada grupo. Esta distinción entre aprendizaje percibido y aprendizaje real es importante en el diseño de materiales educativos, dado que la motivación, el interés y la sensación de progreso subjetivo pueden influir en la disposición del estudiante a continuar aprendiendo.

Pregunta 9: ¿Te gustaría que tus profesores crearan más videocápsulas de este estilo para repasar contenidos?

- a) Sí
- b) No
- c) Indiferente

Esta pregunta apuntaba a medir el grado de aceptación de este tipo de recurso educativo entre los estudiantes, considerando tanto el formato audiovisual como el estilo narrativo y visual del video. En otras palabras, buscaba identificar si los estudiantes verían con buenos ojos una mayor implementación de este tipo de cápsulas por parte del cuerpo docente.

En ambos grupos, la respuesta más común fue “Sí”, con un 42,9% en el grupo clonada y un 43,8% en el grupo Loquendo, lo que muestra una apertura general a este tipo de herramientas. No obstante, en el grupo Loquendo también un 43,8% eligió “No”, evidenciando una polarización equilibrada entre aceptación y rechazo en este grupo de participantes. En contraste, sólo un 28,6% del grupo clonada se mostró contrario a la idea, lo que sugiere una mayor disposición hacia este tipo de formato cuando se utiliza voz clonada.

Llama la atención, además, que la opción “Indiferente” fue elegida por el 28,6% del grupo clonada, pero sólo por el 12,5% del grupo Loquendo. Esto sugiere que, si bien ambos grupos presentan divisiones similares entre quienes están a favor o en contra, el grupo clonada tiende más a una postura neutral que el grupo Loquendo, el cual se inclina más hacia posiciones firmes, ya sea a favor o en contra.

En conjunto, los datos muestran que las videocápsulas tienen un grado moderado de aceptación entre los participantes, con una leve mayor apertura en el grupo que utilizó voz clonada. Si bien no existe un consenso rotundo, el interés mostrado por casi la mitad de los estudiantes en volver a consumir videocápsulas similares por parte de sus profesores, indica que este formato puede ser una herramienta valiosa para reforzamiento de contenidos, siempre que se ajuste a los estilos de aprendizaje y preferencias del público objetivo.

Pregunta 10: ¿Qué tan natural te pareció la voz utilizada en el video?
(Muy artificial → 1 | Muy natural → 10)

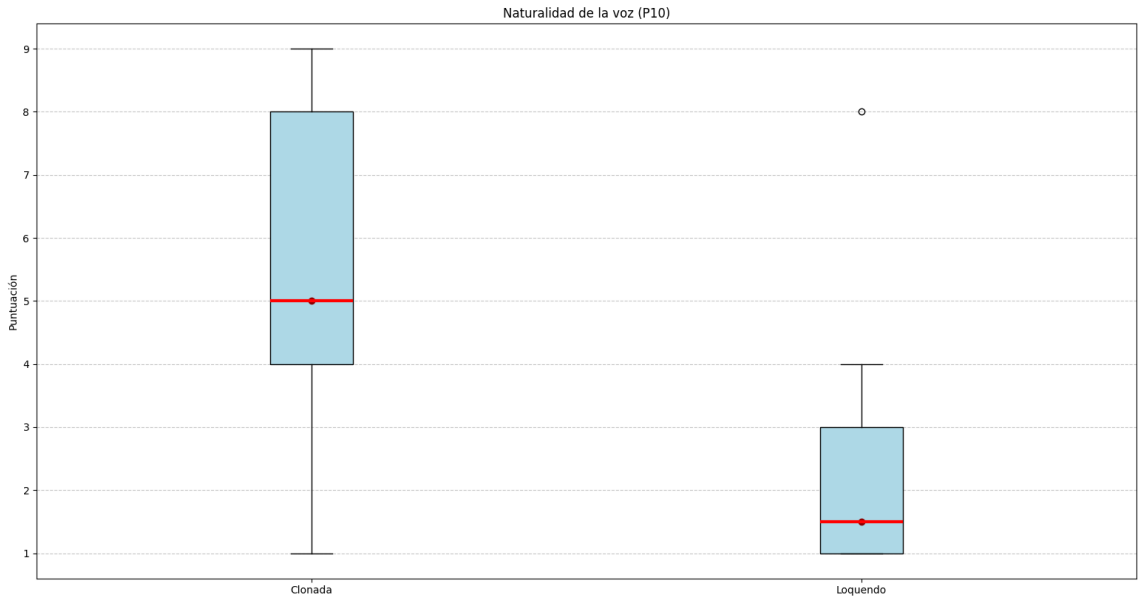

Figura 9. Distribución de la percepción de la naturalidad de la voz – Pregunta 10.

Fuente: Elaboración propia

Esta pregunta tenía como objetivo explorar la percepción subjetiva de naturalidad de la voz empleada en cada versión del video. Dado que una de las principales motivaciones del uso de clonación de voz es justamente acercarse a la calidad expresiva de una voz humana real, este ítem resulta clave para evaluar el impacto del tipo de voz en la experiencia del usuario.

Los resultados muestran una diferencia considerable entre los grupos. El grupo que visualizó la versión con voz clonada otorgó una media de 5,38 puntos (con desviación estándar = 2,58), mientras que el grupo Loquendo promedió sólo 2,19 puntos (desviación estándar = 1,83). Esta diferencia fue estadísticamente significativa, y dispone de un valor $t(35) = 4,40$ y un *p-valor* = 0,0001.

La Figura X, que presenta un boxplot de la distribución de puntajes, refuerza esta diferencia al evidenciar no sólo una mediana más alta para el grupo clonada, sino también una mayor dispersión, lo cual indica que algunos participantes percibieron la voz como notablemente más natural, aunque con cierta variabilidad en la evaluación. En contraste, las respuestas del grupo Loquendo se agrupan de forma más estrecha hacia los valores bajos de la escala, reflejando una percepción casi unánime de artificialidad.

Este resultado corrobora la hipótesis de que las voces clonadas pueden ofrecer una experiencia auditiva percibida como más natural que las voces sintéticas tradicionales. Si bien no todos los estudiantes del grupo clonada evaluaron la voz como completamente natural, sí se evidencia una mejora significativa en la percepción general. Esto sugiere que la tecnología de clonación de voz tiene un potencial claro para mejorar la calidad subjetiva de materiales audiovisuales educativos, especialmente cuando se busca generar una experiencia más cercana a la interacción humana real.

Pregunta 11: ¿La voz utilizada en el video te ayudó a mantenerte concentrado/a en el contenido?

(Nada → 1 | Mucho → 10)

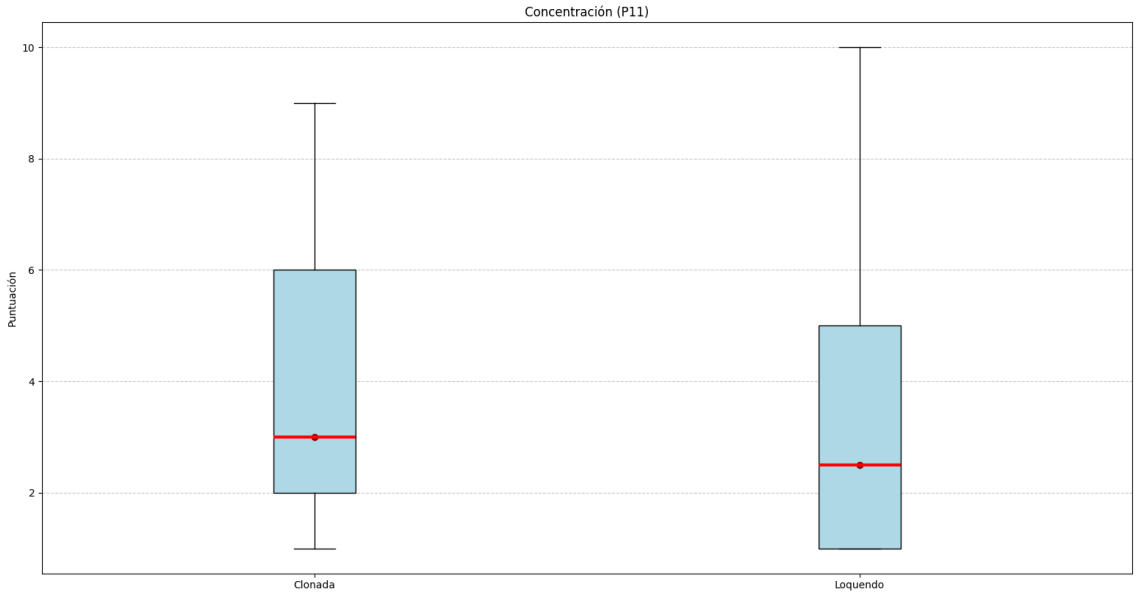

Figura 10. Distribución de autoevaluación de la concentración – Pregunta 11.

Fuente: Elaboración propia

Esta pregunta tenía como propósito evaluar si el tipo de voz influía en la capacidad de los estudiantes para mantener la atención durante el visionado del video. Aunque la percepción de naturalidad puede afectar la inmersión en un recurso audiovisual, no necesariamente implica una mejora directa en la concentración, por lo que este ítem resulta útil para distinguir entre ambos aspectos.

El grupo que visualizó el video con voz clonada otorgó un puntaje promedio de 3,95 (desviación estándar = 2,40), mientras que el grupo Loquendo promedió 3,19 puntos (desviación estándar = 2,54). Si bien se observa una leve diferencia a favor del grupo clonada, esta no fue estadísticamente significativa, pues se obtuvieron valores $t(35) = 0,93$

y p = 0,3591, lo que sugiere que el tipo de voz utilizada no tuvo un efecto claro sobre la concentración de los participantes.

La Figura 10 muestra un boxplot con la distribución de puntajes en ambos grupos. Puede observarse que ambos presentan una mediana similar y un rango intercuartílico amplio, indicando una alta variabilidad en las respuestas. Esto refuerza la idea de que la concentración durante el visionado estuvo influida por factores diversos más allá del tipo de voz, como el interés personal por el tema, el entorno o el estado anímico del participante.

En conclusión, aunque la voz clonada fue evaluada como más natural en preguntas anteriores, este efecto positivo no se tradujo en una mejora clara en la capacidad de mantener la concentración durante el video. Esto sugiere que la naturalidad de la voz puede mejorar la experiencia auditiva, pero no necesariamente garantiza por sí sola un mayor enfoque atencional sobre el contenido.

Pregunta 12: ¿Te resultó útil este formato de video para tu aprendizaje en comparación con otros métodos (como lecturas o clases tradicionales)?

- a) Más útil
- b) Igual de útil
- c) Menos útil

Esta pregunta buscaba captar la percepción comparativa de los estudiantes respecto a la utilidad del video frente a otros métodos tradicionales de aprendizaje, como lecturas o clases expositivas. Las respuestas entregan una visión general sobre cómo se posiciona el formato de videocápsula dentro del repertorio pedagógico habitual.

La opción más seleccionada en ambos grupos fue la b), “igual de útil”, con un 47,6% en el grupo con voz clonada y un 37,5% en el grupo Loquendo. Esto indica que una porción significativa del estudiantado no percibió el video como una mejora clara, pero sí como un recurso al menos equiparable en utilidad a métodos tradicionales.

Sin embargo, destaca también el alto porcentaje de estudiantes que consideró el video “menos útil” (opción c): 42,9% en el grupo clonada y 43,8% en el grupo Loquendo. Este resultado sugiere que, para una parte importante de los participantes, el formato audiovisual no logró superar ni igualar en efectividad a enfoques convencionales, ya sea por estilo de aprendizaje personal, nivel de profundidad percibido, o factores relacionados con la calidad de la experiencia de visionado.

La opción a), que representa una valoración positiva, “más útil”, fue la menos seleccionada en ambos casos: apenas un 9,5% en el grupo clonada y un 18,8% en el grupo Loquendo. Esto puede interpretarse como una señal de que, aunque el formato fue funcional, no fue considerado por la mayoría como una herramienta superior para el aprendizaje. En conjunto, estos resultados reflejan una recepción relativamente moderada del formato. La mayoría lo percibió como igual o menos útil que otras metodologías, con una valoración positiva solo marginal. Esto subraya la importancia de considerar el uso de videos como complemento y no como reemplazo directo de otras estrategias pedagógicas, al menos según las percepciones actuales de los estudiantes, frente a un formato como el que les fue presentado para este experimento.

En conjunto, los resultados obtenidos ofrecen una visión matizada sobre el impacto del tipo de voz utilizada en las videocápsulas, tanto en el desempeño objetivo de los participantes como en sus percepciones subjetivas. Si bien no se identificaron diferencias estadísticamente significativas en el puntaje total de la evaluación, sí emergieron variaciones relevantes en ítems individuales y, especialmente, en ciertas dimensiones perceptuales como la naturalidad de la voz o la sensación de aprendizaje logrado. Asimismo, las respuestas a preguntas abiertas y de valoración permitieron identificar patrones de preferencia y escepticismo hacia este formato, los cuales serán discutidos en mayor profundidad en el capítulo siguiente, dedicado a las conclusiones generales extraídas del estudio.

CAPÍTULO 4: CONCLUSIONES Y TRABAJO FUTURO

La hipótesis que motivó este estudio sostenía que las voces clonadas, al reproducir de manera más realista las cualidades del habla y narración de una voz humana, como el ritmo, la entonación y la fluidez, podrían no solo facilitar la comprensión del contenido, sino también mejorar la experiencia subjetiva del usuario, promoviendo una mayor atención y receptividad hacia el material educativo.

A la luz de los resultados obtenidos, esta hipótesis se valida de manera parcial. Aunque no se observaron diferencias estadísticamente significativas en el puntaje global de comprensión, sí emergieron diferencias relevantes en otras dimensiones: la percepción de aprendizaje, la naturalidad atribuida a la voz, y la disposición general hacia el uso futuro de cápsulas educativas con esta tecnología. Estos hallazgos sugieren que, si bien la mejora en comprensión objetiva no fue concluyente, el impacto subjetivo positivo de las voces clonadas representa una ventaja concreta que merece ser considerada en futuros desarrollos pedagógicos basados en Inteligencia Artificial.

Con esto en mente, los resultados permiten evaluar de manera integral el cumplimiento del objetivo central de esta investigación: determinar si el uso de voces clonadas mediante inteligencia artificial, en comparación con voces sintéticas tradicionales como Loquendo, tiene un impacto significativo en la comprensión y percepción del contenido presentado en cápsulas educativas. A partir del análisis realizado, se pueden extraer conclusiones relevantes tanto desde el punto de vista del aprendizaje objetivo como de la experiencia subjetiva del usuario.

Aunque el análisis de los puntajes totales no arrojó diferencias estadísticamente significativas entre los grupos, el examen detallado de cada ítem reveló algunos matices importantes. En algunas preguntas específicas, como la 3 y la 5, se observaron diferencias notables en la distribución de respuestas, lo que sugiere que el tipo de voz pudo influir en el modo en que ciertos contenidos fueron comprendidos o recordados. Sin embargo, en la mayoría de los casos, el rendimiento fue comparable entre ambas condiciones, lo que refuerza la idea de que la voz clonada puede utilizarse sin detrimento de la comprensión objetiva, al menos en este tipo de evaluaciones breves y temáticas acotadas.

Más reveladoras resultaron las percepciones subjetivas reportadas por los participantes. La voz clonada fue evaluada de forma significativamente más positiva en cuanto a su naturalidad, lo que indica una mejora perceptible en la experiencia auditiva. Asimismo, los estudiantes que vieron el video con voz clonada manifestaron una mayor sensación de aprendizaje logrado, en comparación con aquellos que vieron la versión con voz

Loquendo. Estos hallazgos, respaldados por diferencias estadísticamente significativas, sugieren que el uso de tecnologías avanzadas de síntesis de voz no sólo impacta la forma en que se recibe el contenido, sino también la forma en que este es valorado por los usuarios.

Por otro lado, la concentración durante el visionado no pareció verse afectada significativamente por el tipo de voz, lo que indica que otros factores, como el interés personal en el tema o el diseño visual del video, podrían desempeñar un papel más decisivo en este aspecto. En cuanto a la utilidad percibida del formato en relación con métodos tradicionales, la mayoría de los estudiantes lo consideró “igual de útil” o “menos útil”, con una baja proporción señalándolo como claramente superior. Este resultado sugiere que, aunque las videocápsulas con voz clonada pueden mejorar ciertos aspectos de la experiencia, aún existen desafíos para posicionarlas como una herramienta pedagógica central o preferida.

En conjunto, el análisis permite concluir que la voz clonada ofrece ventajas perceptivas claras, sin comprometer el rendimiento objetivo de los estudiantes. No obstante, su efecto no es suficiente, por sí solo, para generar un cambio sustantivo en la percepción global del formato educativo. En este contexto, plataformas como EduvidIA podrían desempeñar un papel útil al facilitar la creación de contenidos más cercanos y personalizados, pero su impacto dependerá en gran medida de cómo se integren dentro de una estrategia pedagógica más amplia, y no de la tecnología por sí sola.

No obstante, también es necesario reconocer las limitaciones actuales tanto del experimento como de la tecnología utilizada. En primer lugar, el tamaño muestral fue reducido, lo cual restringe la generalización de los resultados. Del mismo modo, el contenido evaluado correspondía a una cápsula breve y de carácter informativo, por lo que los efectos observados podrían diferir en formatos más largos o con una carga cognitiva mayor. A nivel tecnológico, si bien las voces clonadas fueron evaluadas como más naturales, aún persisten desafíos en la expresividad emocional, la entonación contextual y la modulación de ciertas palabras, lo que podría afectar su desempeño en videos de tipo narrativo. A esto se suma la necesidad de contar con múltiples muestras de voz originales, potencialmente horas de audio de buena calidad, para realizar una clonación eficaz que genere un modelo fidedigno a la voz que se busca replicar, lo que plantea cuestiones éticas y desafíos logísticos sobre consentimiento, privacidad y obtención de registros vocales útiles.

Dado este panorama, se proponen tres líneas para continuar o escalar el proyecto. En primer lugar, ampliar el alcance de las pruebas a diferentes asignaturas, niveles educativos y tipos de contenido. Por ejemplo, explicativos, argumentativos o prácticos, con el fin de

identificar en qué contextos específicos las voces clonadas generan el mayor valor agregado. En segundo lugar, explorar mejoras técnicas que permitan dotar a las voces clonadas de mayor expresividad y control emocional, acercándolas aún más a la experiencia de una voz humana real, potencialmente ocupando software de pago para obtener una mejor replicación. Y en tercer lugar, incorporar una mayor variedad de instrumentos de mediciones, como preguntas abiertas, tareas aplicadas o evaluaciones a largo plazo, entre otros, idealmente con una población de participantes más grande, permitiría obtener una visión más completa sobre el impacto educativo.

En suma, las voces clonadas representan una innovación con alto potencial educativo, cuya efectividad depende no solo de la calidad técnica de la síntesis, sino también de su integración cuidadosa en estrategias didácticas coherentes. Aunque aún hay aspectos por mejorar, los hallazgos de este estudio respaldan su uso como herramienta complementaria en la creación de materiales audiovisuales accesibles, atractivos y personalizados, especialmente en un contexto donde la educación digital y los recursos automatizados seguirán cobrando relevancia.

REFERENCIAS BIBLIOGRÁFICAS

[1] Hunter, J., Sonnemann, J., & Joiner, R. (2022). *Making time for great teaching: How better government policy can help*. Grattan Institute. **https://grattan.edu.au/report/making-time-for-great-teaching-how-better-government-policy-can-help/**

[2] Almurashi, W. A. (2016). *The effective use of YouTube videos for teaching English language in classrooms as supplementary material at Taibah University in Alula*. International Journal of English Language and Linguistics Research, 4(3), 32-47. https://eajournals.org/ijellr/vol-4-issue-3-april-2016/the-effective-use-of-youtube-videos-for-teaching-english-language-in-classrooms-as-supplementary-material-at-taibah-university-in-alula/

[3] TechTarget. (n.d.). *Learning management system (LMS)*. https://www.techtarget.com/searchcio/definition/learning-management-system

[4] Khan Academy. (n.d.). *Khanmigo: AI-powered teaching assistant & tutor*. https://www.khanmigo.ai/

[5] Reyes Gómez, F. J. (2022). *Metodología del Lab UX USM para guiar a estudiantes y memoristas en el diseño y evaluación centrada en personas*. Universidad Técnica Federico Santa María.

[6] erew123. (2024). *alltalk_tts* [Repositorio GitHub]. GitHub. https://github.com/erew123/alltalk_tts

[7] Gölge, E., & The Coqui TTS Team. (2021). *Coqui TTS* (versión 1.4) [Repositorio GitHub]. Zenodo. https://doi.org/10.5281/zenodo.6334862

ANEXOS

Vídeos del experimento:

- Versión A, voz clonada mediante IA: https://www.youtube.com/watch?v=v-OBC6EbpyQ
- Versión B, voz sintetizada con Loquendo: https://www.youtube.com/watch?v=sB6-yv0ne0E

Cuestionario que respondieron los participantes:

Marque con una X la opción que considera correcta.

1.- ¿Cuál fue una de las principales causas de la Rebelión Taiping?

- a) La expansión territorial de los Qing.
- b) La mejora de las relaciones comerciales con Europa.
- c) La invasión de potencias extranjeras.
- d) La sobrepoblación y la pobreza extrema en el sur de China.

2.- ¿Qué proclamó Hong Xiuquan tras sus visiones divinas?

- a) Que él era el líder de una nueva dinastía Qing.
- b) Que se convertiría en el emperador de China.
- c) Que era el hermano menor de Jesucristo y debía derrocar a la dinastía Qing.
- d) Que China debería aliarse con las potencias extranjeras.

3.- ¿Qué ideología social propuso el Reino Celestial Taiping?

- a) Una sociedad jerárquica basada en el confucianismo.
- b) Una sociedad basada en el orden y la tradición militar.
- c) Una sociedad influenciada por el socialismo europeo.
- d) Una sociedad igualitaria con redistribución de tierras.

4.- ¿Qué grupo militar fue clave para la derrota de los Taiping?

- a) El Ejército Xiang, liderado por Zeng Guofan.
- b) El Ejército Mandarín.
- c) El Ejército Nacional Chino.
- d) El Ejército de la Reforma.

5.- ¿Qué motivó la intervención de las potencias extranjeras en la Rebelión Taiping?

- a) El interés en las riquezas que los Taiping podrían ofrecer.
- b) La preocupación por la inestabilidad de China y el impacto en los intereses comerciales.
- c) El apoyo al cristianismo promovido por los Taiping.
- d) El deseo de controlar el comercio de opio.

6.- ¿Qué fue uno de los legados sociales de la Rebelión Taiping?

- a) La consolidación del confucianismo como doctrina estatal.
- b) La reforma agraria y la redistribución de tierras.
- c) La completa adopción del cristianismo en todo China.
- d) La expansión del comercio internacional.

7.- ¿Qué tan familiarizado/a estabas con el tema antes de ver el video?

a) Nada familiarizado/a b) Algo familiarizado/a c) Bastante familiarizado/a

8.- En una escala del 1 al 10, ¿qué tanto sientes que aprendiste sobre el tema?

Nada Mucho

1	2	3	4	5	6	7	8	9	10

9.- ¿Te gustaría que tus profesores crearán más videocápsulas de este estilo para repasar contenidos?

a) Sí b) No c) Indiferente

10.- ¿Qué tan natural te pareció la voz utilizada en el video?

Muy artificial Muy natural

1	2	3	4	5	6	7	8	9	10

11.- ¿La voz utilizada en el video te ayudó a mantenerte concentrado/a en el contenido?

Nada Mucho

1	2	3	4	5	6	7	8	9	10

12.- ¿Te resultó útil este formato de video para tu aprendizaje en comparación con otros métodos (como lecturas o clases tradicionales)?

a) Más útil b) Igual de útil c) Menos útil