



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



BLINDWIKI 2.0: AN INTELLIGENT VERSION

ADRIA LISA BOU

Thesis supervisor

KARINA GIBERT OLIVERAS (Department of Statistics and Operations Research)

Thesis co-supervisor

XAVIER ANGERRI TORREDEFLOT (Department of Statistics and Operations Research)

Degree

Master's Degree in Innovation and Research in Informatics (Advanced Computing)

Master's thesis

Facultat d'Informàtica de Barcelona (FIB)

Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

TO DO.

Abstract

[1] TO DO.

Keywords

TO DO.

Contents

1	Introducción	4
1.1	Motivación y Objetivos	4
1.2	Objetivos	5
2	Estado del Arte	5
2.1	Desarrollo de Aplicaciones Móviles Multiplataforma	5
2.1.1	Frameworks Considerados	5
2.1.2	Análisis Comparativo	5
2.2	Modelos Clave en Traducción de Voz	6
2.2.1	El Panorama Actual de la Traducción Automatizada	6
2.2.2	Análisis Comparativo de Modelos	6
2.2.3	SeamlessM4T (Meta AI)	7
2.2.4	Whisper (OpenAI)	7
2.2.5	Análisis Comparativo de Robustez	8
2.2.6	Evaluación en Benchmarks Estándar	8
2.2.7	Ventajas Prácticas y Consideraciones de Implementación	8
2.2.8	Comparación Detallada de Modelos de Traducción	9
2.2.9	Análisis Multidimensional	9
2.2.10	Análisis por Casos de Uso	9
2.2.11	Benchmarks: Metodologías y Significancia	10
2.2.12	Fleurs: Evaluación Multilingüe Robusta	10
2.2.13	CVSS: Enfoque en Traducción Voz a Voz	10
2.2.14	CoVoST 2: Diversidad de Hablantes	10
2.2.15	Flores: Enfoque en Idiomas de Bajos Recursos	11
2.2.16	Blaser 2.0: Revolución en Métricas de Evaluación	11
2.2.17	Aplicaciones Industriales y Casos de Éxito	11
2.2.18	Desafíos Éticos y Sociales	12
2.2.19	Recomendaciones para Desarrollo Responsable	12
2.2.20	Convergencias y Tendencias Futuras	13
2.2.21	Tendencias Tecnológicas Emergentes	13
2.2.22	El Horizonte de Investigación	13
2.2.23	Implicaciones Socioeconómicas y Culturales	13
2.2.24	Consideraciones de Equidad y Acceso	14
3	Specification and design of the solution	14
3.1	Overview	14
4	Development of the work	14

5	Experimentation and evaluation of the work	14
6	Sustainability analysis and ethical implications	15
7	Conclusion	15

1. Introducción

BlindWiki es una red de audio geolocalizada que permite a personas con discapacidad visual total o parcial compartir sus experiencias mediante grabaciones sonoras utilizando teléfonos inteligentes. Creada en 2014 por Antoni Abad, la plataforma no se limita a documentar dificultades y barreras urbanas, sino que constituye un repositorio de experiencias, opiniones e historias que genera una cartografía colaborativa y creativa de lo invisible.

El proyecto tuvo su origen como una iniciativa para dar voz a las personas con discapacidad visual, permitiéndoles documentar y compartir su percepción única del entorno urbano. Desde su creación, BlindWiki ha experimentado una notable expansión internacional, desarrollándose en ciudades como Roma (2014-2015), Sydney (2015), Berlín y Wrocław (2016), Venecia (2017), Valencia (2020) y São Paulo (2022), adaptándose a cada contexto cultural y lingüístico.

La aplicación móvil de BlindWiki, disponible tanto para Android como iOS, permite a los participantes grabar audio específico del lugar y publicarlo inmediatamente en la plataforma. Los usuarios pueden desplazarse por sus ciudades mientras publican y reciben descripciones de audio geolocalizadas, historias, obstáculos o crónicas previamente contribuidas a través de la app. Esta funcionalidad facilita la creación de un mapa sensorial colectivo que enriquece la experiencia de navegación urbana para personas con discapacidad visual.

1.1 Motivación y Objetivos

La necesidad de rediseñar la aplicación surge de los avances tecnológicos en los dispositivos móviles y sus sistemas operativos, así como de la comunidad internacional de usuarios que demanda mejoras en la accesibilidad y funcionalidad.

El stack tecnológico de la aplicación original se compone de PhoneGap/Cordova como framework base para el empaquetado, e Ionic 1 con Angular.js para la interfaz de usuario y lógica de la aplicación. Esta arquitectura, aunque permitió un desarrollo multiplataforma eficiente en su momento, ha quedado obsoleta frente a los estándares actuales de desarrollo móvil, lo que dificulta la implementación de nuevas funcionalidades y afecta al rendimiento en dispositivos modernos.

Para abordar estos desafíos, se ha desarrollado BlindWiki 2.0, una aplicación completamente rediseñada utilizando tecnologías modernas y eficientes. El nuevo stack tecnológico se basa en **React Native** y **Expo**.

Además, la expansión internacional del proyecto ha llevado a la necesidad de superar las barreras lingüísticas que limitan la comunicación entre participantes de diferentes países. Para lograr este objetivo, es preciso implementar una capa de inteligencia artificial basada en modelos avanzados de procesamiento de lenguaje natural y audio.

Específicamente, se utilizan tres componentes principales:

- **Identificación automática de idioma:** Mediante el modelo **TO DO.**, que permite detectar el idioma original de cada grabación de audio.
- **Transcripción de audio a texto:** Utilizando el modelo **TO DO.** para convertir las grabaciones de audio en texto en su idioma original.
- **Traducción de audio a audio:** **TO DO.**

1.2 Objetivos

2. Estado del Arte

En esta sección discutiremos los últimos avances en tecnologías de desarrollo de aplicaciones móviles multiplataforma, y el estado del arte en la traducción automática Speech-to-Speech (S2ST) y el reconocimiento automático del habla (ASR).

2.1 Desarrollo de Aplicaciones Móviles Multiplataforma

El desarrollo de aplicaciones multiplataforma ha evolucionado significativamente durante la última década como respuesta a la fragmentación del mercado móvil entre Android (71.9%) e iOS (27.68%) [2], y la necesidad de optimizar recursos de desarrollo. El concepto fundamental *"write once, run everywhere"* constituye el núcleo de esta filosofía.

Históricamente, esta tecnología evolucionó desde enfoques basados en WebView (PhoneGap/Cordova) hacia frameworks de compilación a código nativo, hasta las soluciones actuales que equilibran rendimiento nativo con eficiencia de desarrollo. El rendimiento de nativo es de especial interés para el proyecto blind wiki, ya que eso proporciona una mayor compatibilidad con los lectores de pantalla (TalkBack en Android y VoiceOver en iOS).

2.1.1 Frameworks Considerados

Flutter (Google) Framework que utiliza Dart como lenguaje y un motor de renderizado propio (Skia). Su arquitectura se compone de tres capas: el framework Dart, el motor C/C++ con Skia, y los wrappers específicos de plataforma. Entre sus ventajas, Flutter ofrece un rendimiento similar al nativo con 60 FPS en animaciones complejas, una interfaz de usuario consistente y personalizable en todas las plataformas, y widgets adaptables tanto de Material Design como de Cupertino. Además, cuenta con sistema de árbol semántico con descripciones para cada widget, perfectamente integrado con las APIs nativas de accesibilidad (TalkBack/VoiceOver).

React Native (Meta) Utiliza JavaScript y React para crear aplicaciones móviles mediante una arquitectura puente que conecta JavaScript con componentes nativos de la plataforma. Sus ventajas incluyen el aprovechamiento del ecosistema JavaScript/npm, la utilización de componentes UI nativos reales, y la facilidad de transición desde el desarrollo web. Además, React Native ofrece integración directa con los componentes nativos de accesibilidad, heredando automáticamente las mejoras del sistema operativo. No obstante, el puente JavaScript-Nativo puede limitar el rendimiento en comparación con Flutter.

2.1.2 Análisis Comparativo

En la siguiente tabla 1 se muestra un resumen de las características de los frameworks considerados para el desarrollo de la aplicación blind wiki.

Para el desarrollo de la aplicación blind wiki, se ha optado por React Native, por su mayor madurez, y por que presenta una curva de aprendizaje más suave para el desarrollador. Las limitaciones en cuanto a al rendimiento de React Native en comparación con Flutter no son críticas para el proyecto,

Table 1: Comparación de Frameworks Multiplataforma

Característica	Flutter	React Native
Lenguaje	Dart	JavaScript
Arquitectura UI	Motor propio	Componentes nativos
Rendimiento	Excelente	Bueno
Hot Reload	Sí	Sí
Curva Aprendizaje	Moderada	Baja (dev JS)
Madurez	Media (2017)	Alta (2015)
Accesibilidad	Árbol semántico	Componentes nativos

ya que los principales cuellos de botella son las conexiones con el servidor y no el rendimiento interno de la aplicación.

2.2 Modelos Clave en Traducción de Voz

2.2.1 El Panorama Actual de la Traducción Automatizada

La traducción automática ha experimentado una revolución radical con el advenimiento de modelos neuronales avanzados. Esta evolución ha transformado un campo tradicionalmente dominado por sistemas basados en reglas y estadísticas hacia arquitecturas de aprendizaje profundo cada vez más sofisticadas, capaces de capturar sutilezas lingüísticas previamente inaccesibles para sistemas automatizados.

El paradigma actual se caracteriza por varios avances fundamentales:

- **Arquitecturas Transformer:** Han reemplazado a las redes neuronales recurrentes (RNN) como base de los sistemas más avanzados, permitiendo el procesamiento paralelo y la captura de dependencias a larga distancia en el texto y audio.
- **Modelos Multimodales:** La integración de diferentes modalidades (texto, voz, imagen) ha permitido sistemas que comprenden el lenguaje de manera más holística, similares a la comprensión humana.
- **Modelos de Extremo a Extremo:** Los sistemas modernos tienden a minimizar los pasos intermedios, traduciendo directamente entre modalidades (por ejemplo, Speech-to-Speech) sin necesidad de convertir explícitamente a representaciones intermedias.
- **Preprocesamiento y Representaciones Universales:** El uso de representaciones universales del lenguaje (*embeddings*) ha permitido que los modelos generalicen mejor entre idiomas, incluso aquellos con pocos recursos de entrenamiento.

2.2.2 Análisis Comparativo de Modelos

En la Tabla 2 se presenta un análisis comparativo de los principales modelos de traducción actuales.

Table 2: Comparación de Modelos de Traducción

Característica	SeamlessM4T	Whisper	Google T
Arquitectura	Transformer multimodal	Transformer	Transfo
Modalidades	Voz-voz, voz-texto, texto-texto, texto-voz	Voz-texto	Texto-texto,
Idiomas (ASR/S2S/T2T)	96/100→35/95↔95	97/Ninguno/Limitado	~100/~60
Robustez al Ruido	38% mejor	Referencia	No publ
Fleurs S2TT (BLEU ↑)	20.4 (X→Eng)	16.2 (X→Eng)	No publ
CVSS S2ST (ASR-BLEU ↑)	58.7	N/A	No publ

2.2.3 SeamlessM4T (Meta AI)

SeamlessM4T representa un paradigma revolucionario en la traducción automática al integrar múltiples modalidades y tareas en un único modelo unificado. Desarrollado por Meta AI, este sistema marca un antes y un después en la capacidad de los modelos para manejar traducciones multimodales a escala global.

Arquitectura Innovadora La arquitectura de SeamlessM4T se distingue por su diseño modular pero unificado:

- **Encoder Compartido:** Procesa tanto texto como audio usando una representación intermedia común, permitiendo transferencia de conocimiento entre modalidades.
- **Bridge Multimodal:** Conecta diferentes modalidades mediante un espacio latente que preserva información semántica independientemente del medio de entrada.
- **Decoders Especializados:** Generan salidas específicas para cada tipo de traducción, pero comparten conocimiento a través de parámetros parcialmente compartidos.

Capacidades Lingüísticas SeamlessM4T establece un nuevo estándar en cobertura lingüística:

- Para S2S: maneja traducciones de 100 idiomas hacia inglés y de inglés hacia 35 idiomas.
- Para T2T: soporta traducciones entre 95 idiomas en cualquier dirección.
- Para ASR: reconoce habla en 96 idiomas diferentes.

2.2.4 Whisper (OpenAI)

Whisper, desarrollado por OpenAI, ha establecido un estándar importante en el campo del reconocimiento automático del habla (ASR) multilíngüe.

Arquitectura y Características Whisper se basa en una arquitectura encoder-decoder transformer, optimizada específicamente para ASR:

- **Encoder Convolutivo:** Procesa el audio de entrada transformándolo en representaciones que capturan patrones fonéticos.
- **Transformer Bidireccional:** Analiza estas representaciones en contexto completo.

- **Decoder Autoregresivo:** Genera texto de forma secuencial, aprovechando tanto el audio procesado como el contexto previo.

Capacidades y Limitaciones Whisper destaca por su amplia cobertura lingüística:

- Soporta reconocimiento de voz en 97 idiomas.
- Maneja diferentes acentos, dialectos y variantes lingüísticas.
- Puede identificar automáticamente el idioma hablado.

Sin embargo, carece de capacidad nativa para traducción de voz a voz (S2S), requiriendo sistemas en cascada para traducción completa.

2.2.5 Análisis Comparativo de Robustez

En términos de robustez ante condiciones adversas, Whisper muestra fortalezas y limitaciones:

Presenta buena resistencia general al ruido y variaciones acústicas, gracias a su entrenamiento con datos diversos.

Sin embargo, es menos robusto que SeamlessM4T ante ruido de fondo específico, según evaluaciones en Fleurs.

Su manejo de variaciones del hablante también resulta inferior al de SeamlessM4T en aproximadamente un 49

Estas limitaciones reflejan la naturaleza específica de Whisper como modelo ASR, en contraste con la aproximación multimodal integrada de SeamlessM4T.

2.2.6 Evaluación en Benchmarks Estándar

En evaluaciones cuantitativas, Whisper demuestra rendimiento competitivo pero inferior a modelos más recientes en ciertos aspectos:

En Fleurs S2TT, Whisper-Large-v2 alcanza un BLEU de 16.2 para traducción X→Eng, frente al 20.4 de SeamlessM4T-Large.

Su rendimiento en ASR puro sigue siendo competitivo, especialmente considerando su fecha de lanzamiento anterior.

2.2.7 Ventajas Prácticas y Consideraciones de Implementación

Whisper mantiene ciertas ventajas prácticas que explican su adopción continua:

Su arquitectura relativamente eficiente permite implementaciones en dispositivos con recursos limitados.

Al ser de código abierto, facilita adaptaciones y fine-tuning para dominios específicos.

La madurez de su ecosistema incluye numerosas implementaciones optimizadas por la comunidad.

Estas características hacen que Whisper siga siendo relevante para aplicaciones específicas de ASR, particularmente cuando la traducción no es el objetivo principal o puede manejarse en un paso separado.

2.2.8 Comparación Detallada de Modelos de Traducción

Para comprender completamente las capacidades y limitaciones de los modelos actuales de traducción, es esencial realizar una comparación detallada que considere múltiples dimensiones de rendimiento, funcionalidad y aplicabilidad.

2.2.9 Análisis Multidimensional

La siguiente tabla expandida proporciona una comparación comprehensiva de los principales modelos discutidos, añadiendo métricas relevantes para evaluar su idoneidad en diferentes contextos:

Table 3: Comparación Multidimensional de Modelos de Traducción

Característica	SeamlessM4T	Whisper	Google Translate
Arquitectura	Transformer multimodal	Transformer	Transformer
Modalidades	Voz-voz, voz-texto, texto-texto, texto-voz	Voz-texto	Texto-texto
Idiomas (ASR/S2S/T2T)	96/100→35/95↔95	97/Ninguno/Limitado	~100
Robustez al Ruido (Fleurs WER ↓)	38% mejor que Whisper	Referencia base	N/A
Fleurs S2TT (BLEU ↑)	20.4 (X→Eng)	16.2 (X→Eng)	~15
CVSS S2ST (ASR-BLEU ↑)	58.7	N/A	~50
Flores T2T (chrF++ ↑)	54.3 (Eng→X)	N/A	~50
Latencia (ms)	~300	~500	~1000
Tamaño (GB)	~2.5	~1.5	>10
Código Abierto	Sí	Sí	No
Ejecución Local	Posible (con recursos)	Posible	No
Mitigación de Toxicidad	Reducción del 63%	Limitada	No

2.2.10 Análisis por Casos de Uso

Complementando la tabla comparativa, podemos analizar qué modelos resultan más adecuados para diferentes escenarios:

- **Comunicación Internacional en Tiempo Real:** SeamlessM4T destaca claramente gracias a su capacidad S2S directa y baja latencia relativa.
- **Transcripción Multilíngüe:** Tanto Whisper como SeamlessM4T ofrecen excelente rendimiento, con ventaja para SeamlessM4T en entornos ruidosos.
- **Traducción de Documentos de Alto Volumen:** NLLB-200 y DeepL pueden ser preferibles por su especialización en traducción texto-texto.
- **Entornos con Conectividad Limitada:** Whisper ofrece la mejor relación rendimiento/tamaño para implementaciones locales con recursos limitados.
- **Idiomas de Bajos Recursos:** NLLB-200 presenta la mayor cobertura de idiomas minoritarios para texto, mientras SeamlessM4T lidera en modalidad voz.

Esta comparación revela que no existe un "mejor modelo" universal, sino que la elección óptima depende de los requisitos específicos, restricciones técnicas y contexto de aplicación.

2.2.11 Benchmarks: Metodologías y Significancia

Los benchmarks desempeñan un papel crucial en la evaluación y comparación objetiva de modelos de traducción. Comprender sus metodologías específicas y limitaciones resulta esencial para interpretar correctamente sus resultados.

2.2.12 Fleurs: Evaluación Multilingüe Robusta

Fleurs (Few-shot Learning Evaluation of Universal Representations) representa uno de los benchmarks más completos para evaluación de ASR multilingüe y traducción voz-texto:

- Composición del Dataset: Incluye grabaciones de 102 idiomas con múltiples hablantes nativos por idioma, deliberadamente diversificados por acento, edad y género.
- Metodología de Evaluación: Incorpora variaciones controladas de ruido de fondo, reverberación y características del hablante para evaluar robustez.
- Métricas Principales:
 - WER (Word Error Rate) para ASR, donde valores más bajos indican mayor precisión.
 - BLEU (Bilingual Evaluation Understudy) para traducción, donde valores más altos indican mayor calidad.
- Significancia: Fleurs evalúa específicamente la capacidad del modelo para funcionar en condiciones del mundo real, más allá de entornos controlados de laboratorio.

2.2.13 CVSS: Enfoque en Traducción Voz a Voz

CVSS (Cross-lingual Voice Similarity Search) se especializa en evaluar la calidad de traducción voz a voz (S2ST):

- Metodología: Evalúa tanto la preservación del contenido semántico como la calidad y naturalidad del habla generada.
- ASR-BLEU: Su métrica principal transcribe la salida de voz utilizando un sistema ASR de referencia y compara esta transcripción con una traducción de referencia mediante BLEU.
- Limitaciones Reconocidas: No captura completamente aspectos prosódicos como entonación, ritmo y énfasis emocional, fundamentales para la percepción humana de naturalidad.
- Complementariedad: Debe interpretarse junto con evaluaciones subjetivas humanas para una comprensión completa de la calidad percibida.

2.2.14 CoVoST 2: Diversidad de Hablantes

CoVoST 2 se distingue por su enfoque en la diversidad de hablantes:

- Base de Datos: Derivado de Common Voice, contiene grabaciones verificadas por la comunidad, incluyendo mayor diversidad de acentos y dialectos que otros datasets.

- Estructura: Se centra en traducción voz-texto desde inglés a 15 idiomas, con énfasis en la generalización a diferentes hablantes.
- Evaluación: Principalmente mediante BLEU, permite analizar cómo los modelos manejan variaciones en pronunciación, acento y estilo de habla.
- Valor Particular: Su diversidad de hablantes lo hace especialmente valioso para evaluar la equidad y representatividad de los modelos.

2.2.15 Flores: Enfoque en Idiomas de Bajos Recursos

Flores (Facebook Low Resource) destaca por su amplia cobertura de idiomas infrarrepresentados:

- Alcance Lingüístico: Incluye 204 idiomas, abarcando explícitamente lenguas con pocos recursos digitales disponibles.
- Metodología: Proporciona conjuntos de datos paralelos cuidadosamente curados, con textos originalmente escritos en todos los idiomas cubiertos, no simplemente traducidos desde inglés.
- Métrica chrF++: Utiliza F-score a nivel de carácter como principal métrica de evaluación, más robusta que BLEU para idiomas morfológicamente ricos.
- Significancia para Inclusión: Permite evaluar modelos en su capacidad para democratizar la tecnología lingüística más allá de idiomas mayoritarios.

2.2.16 Blaser 2.0: Revolución en Métricas de Evaluación

Blaser 2.0 representa un avance significativo en las metodologías de evaluación:

- Agnosticismo de Modalidad: Puede evaluar traducciones independientemente del formato (texto-texto, voz-texto, voz-voz).
- Alineación con Percepción Humana: Diseñado específicamente para correlacionarse con evaluaciones humanas de calidad.
- Metodología: Utiliza embeddings de modelos multimodales para comparar el significado semántico entre origen y destino, sin depender exclusivamente de solapamiento léxico.
- Ventaja Comparativa: Supera limitaciones fundamentales de métricas tradicionales como BLEU, que penalizan reformulaciones válidas y sinónimos.

2.2.17 Aplicaciones Industriales y Casos de Éxito

La traducción automática avanzada está transformando numerosos sectores industriales y sociales, con aplicaciones que van desde comunicación personal hasta operaciones empresariales globales.

- Comunicaciones Internacionales y Diplomacia: Los modelos avanzados de traducción están revolucionando la comunicación internacional.
- Educación y Difusión del Conocimiento: El sector educativo está aprovechando estas tecnologías para democratizar el acceso al conocimiento.

- Aplicaciones Empresariales y Comercio Global: Las empresas están integrando estos modelos en sus operaciones para facilitar la globalización.
- Turismo y Movilidad Internacional: El sector turístico ha encontrado aplicaciones prácticas inmediatas.
- Atención Sanitaria y Servicios Sociales: Quizás uno de los campos con mayor impacto social.

2.2.18 Desafíos Éticos y Sociales

A pesar de sus beneficios, los sistemas avanzados de traducción automática plantean importantes desafíos éticos y sociales que requieren atención.

- Sesgos Lingüísticos y Representación: Los modelos de traducción pueden perpetuar o amplificar sesgos preexistentes.
- Privacidad y Seguridad: El procesamiento de contenido lingüístico plantea serias consideraciones de privacidad.
- Impacto en Profesionales de la Traducción: La automatización plantea preguntas sobre el futuro de la traducción como profesión.
- Estandarización y Homogeneización Lingüística: La traducción masiva plantea preocupaciones sobre diversidad lingüística.

2.2.19 Recomendaciones para Desarrollo Responsable

Basado en estos desafíos, podemos proponer directrices para el desarrollo e implementación responsable:

- Evaluación de Impacto Lingüístico: Incorporar evaluaciones estructuradas sobre cómo los sistemas afectan diferentes comunidades lingüísticas.
- Participación Comunitaria: Involucrar activamente a hablantes de idiomas minoritarios en el desarrollo y evaluación de sistemas.
- Transparencia Algorítmica: Comunicar claramente a los usuarios las limitaciones y posibles sesgos de los sistemas de traducción.
- Enfoque Complementario: Diseñar sistemas que complementen el juicio humano, especialmente en contextos de alta sensibilidad.
- Inversión en Diversidad Lingüística: Dedicar recursos específicos a mejorar rendimiento en idiomas infrarrepresentados.

Estas recomendaciones buscan maximizar los beneficios de la traducción automática mientras se minimizan sus potenciales impactos negativos.

2.2.20 Convergencias y Tendencias Futuras

Integración de Tecnologías Multiplataforma con Traducción Avanzada La convergencia entre desarrollo multiplataforma y traducción automática avanzada representa una de las tendencias más prometedoras en tecnología, generando sinergias que superan la suma de sus partes individuales.

- Arquitecturas Integradas Emergentes: Estamos presenciando el surgimiento de arquitecturas que integran nativamente traducción en frameworks multiplataforma.
- Casos de Uso Transformadores: La convergencia de estas tecnologías está habilitando casos de uso previamente inviables.

2.2.21 Tendencias Tecnológicas Emergentes

Varias tendencias tecnológicas están moldeando el futuro cercano de estas áreas convergentes:

- Modelos On-Device y Edge Computing: El procesamiento local está ganando predominancia.
- Personalización y Adaptación Contextual: La personalización está reemplazando los modelos monolíticos.
- Multimodalidad Ampliada: La integración de múltiples modalidades continúa expandiéndose.

2.2.22 El Horizonte de Investigación

Las fronteras actuales de investigación prometen avances transformadores en los próximos años:

- Modelos Fundacionales Multilíngües: Los modelos foundation específicamente entrenados para comprensión multilíngüe están emergiendo.
- Interfaces Cerebro-Computadora para Traducción: Aunque en etapas iniciales, la investigación en BCIs para traducción muestra potencial.
- Sistemas Auto-supervisados Multilíngües: La auto-supervisión está transformando el entrenamiento lingüístico.

2.2.23 Implicaciones Socioeconómicas y Culturales

El impacto de estas tecnologías convergentes se extenderá más allá de consideraciones puramente técnicas:

- Democratización del Acceso Global: La combinación de desarrollo multiplataforma accesible y traducción avanzada está democratizando el acceso a mercados globales.
- Transformación Laboral y Educativa: El paisaje laboral y educativo está experimentando transformaciones significativas.

2.2.24 Consideraciones de Equidad y Acceso

La distribución equitativa de estos beneficios tecnológicos plantea desafíos importantes:

- Brecha Digital Lingüística: Existe el riesgo de que idiomas y dialectos no incluidos en estos sistemas queden aún más marginalizados digitalmente.
- Acceso a Hardware Adecuado: Los dispositivos capaces de ejecutar estos modelos avanzados no están uniformemente distribuidos globalmente.
- Alfabetización Digital Multilingüe: Se requieren nuevas aproximaciones a la alfabetización digital que consideren contextos multilingües.

Abordar estos desafíos será crucial para asegurar que estas tecnologías actúen como fuerzas democratizadoras y no amplifiquen desigualdades existentes.

3. Specification and design of the solution

3.1 Overview

Disseny de blind wiki 1.0 vs disseny de blind wiki 2.0.

Comentar breument el perquè escullo React Native per fer la app.

Detall sobre el model de traducció el·legit.

4. Development of the work

Esquema Gantt de la meua planificació

Aquesta secció representa que ha de ser un diari del meu progrés???

Detalls de implementació.

5. Experimentation and evaluation of the work

Mesures del rendiment algorísmiques del sistema de traducció podrien ser interessants. Mesures pràctiques del sistema en el entorn de producció podrien ser interessant. Potser hi ha certs paràmetres que es poden ajustar en funció del servidor del que disposem.

6. Sustainability analysis and ethical implications

Anàlisi del cost de manteniment.

Parlar de privacitat? Acció benèfica?

Potser és més interessant enfocar aquesta secció al tema de la accessibilitat digital.

7. Conclusion

Conclusió.

References

- [1] Ye Jia, Michelle Tadmor Ramanovich, Quan Wang, and Heiga Zen. CVSS corpus and massively multilingual speech-to-speech translation. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, pages 6691–6703, 2022.
- [2] StatCounter. Mobile operating system market share worldwide - march 2025, 2025. Recuperado de StatCounter Global Stats.