

Propuesta de sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software

Chanataxi Nacimba, Oscar Fabricio y Colina Morales, Mateo Ruben

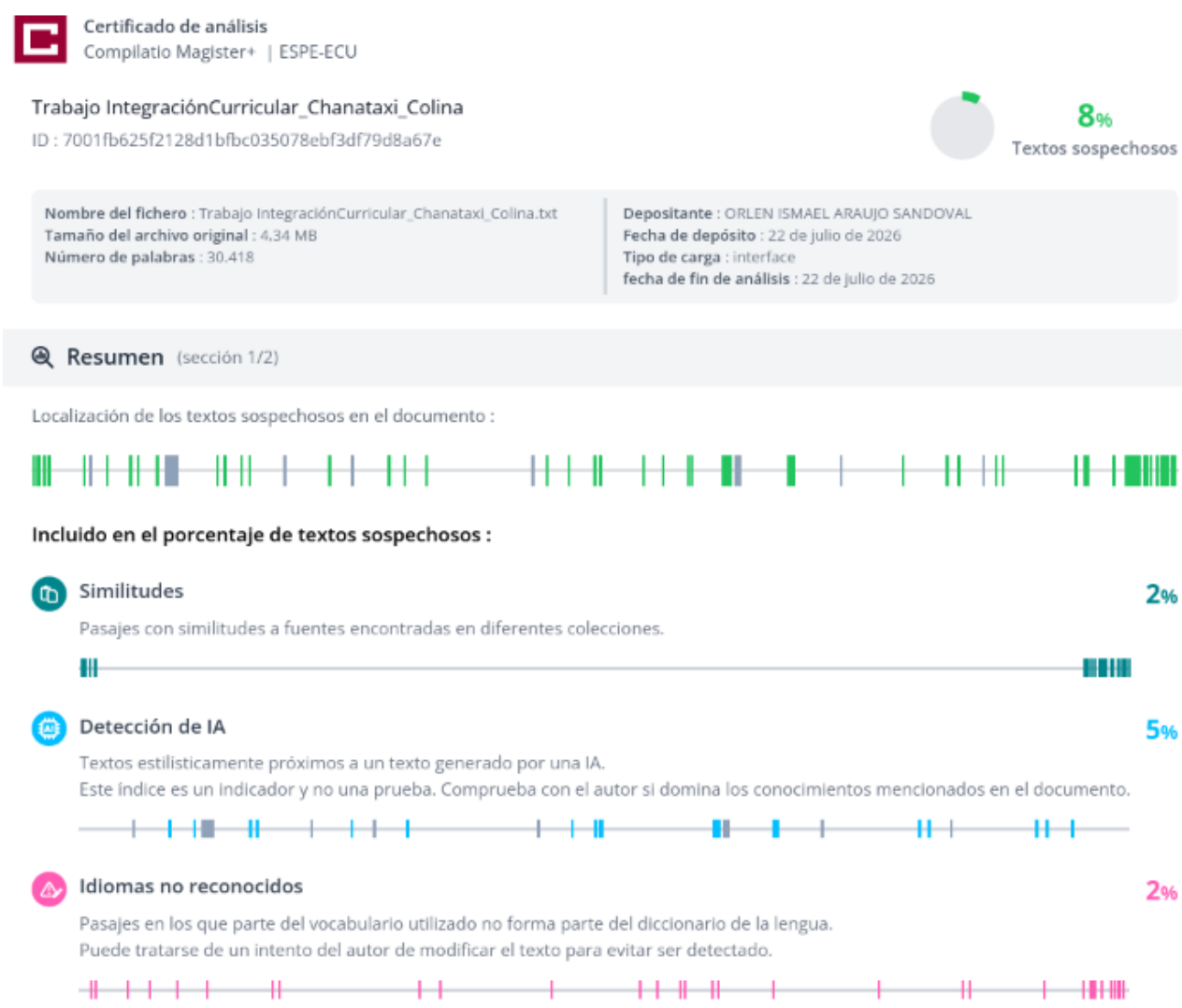
Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en Tecnologías de la Información

Ing. Araujo Sandoval, Orlen Ismael, Mgtr.

31 de julio de 2026



Certificado de análisis
Compilatio Magister+ | ESPE-ECU

Trabajo IntegraciónCurricular_Chanataxi_Colina
ID : 7001fb625f2128d1bfbc035078ebf3df79d8a67e

8%
Textos sospechosos

Nombre del fichero : Trabajo IntegraciónCurricular_Chanataxi_Colina.txt
Tamaño del archivo original : 4,34 MB
Número de palabras : 30.418

Depositante : ORLEN ISMAEL ARAUJO SANDOVAL
Fecha de depósito : 22 de julio de 2026
Tipo de carga : interface
fecha de fin de análisis : 22 de julio de 2026

Resumen (sección 1/2)

Localización de los textos sospechosos en el documento :

Incluido en el porcentaje de textos sospechosos :

Similitudes 2%
Pasajes con similitudes a fuentes encontradas en diferentes colecciones.

Detección de IA 5%
Textos estilisticamente próximos a un texto generado por una IA.
Este índice es un indicador y no una prueba. Comprueba con el autor si domina los conocimientos mencionados en el documento.

Idiomas no reconocidos 2%
Pasajes en los que parte del vocabulario utilizado no forma parte del diccionario de la lengua.
Puede tratarse de un intento del autor de modificar el texto para evitar ser detectado.

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
ORLEN ISMAEL ARAUJO SANDOVAL

Ing. Araujo Sandoval, Orlen Ismael, Mgtr.

Director

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: "**Propuesta de sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software**" fue realizado por los señores **Colina Morales, Mateo Ruben y Chanataxi Nacimba, Oscar Fabricio,** el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE; habiéndose efectuado una revisión integral de forma y de fondo conforme a los lineamientos y formatos institucionales vigentes. Además, fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 31 de Julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
ORLEN ISMAEL ARAUJO SANDOVAL

Ing. Araujo Sandoval, Orlen Ismael, Mgtr.

Director

C.C.: 0604095968

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Nosotros, Chanataxi Nacimba, Oscar Fabricio, con cédula de ciudadanía n°1727426395, y Colina Morales, Mateo Ruben con cédula de ciudadanía n°1727511055 declaramos que el contenido, ideas y criterios del trabajo de integración curricular**: Propuesta de sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 31 de Julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
OSCAR FABRICIO CHANATAXI NACIMBA

Chanataxi Nacimba, Oscar Fabricio

C.C.: 1727426395

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
MATEO RUBEN COLINA MORALES

Colina Morales, Mateo Ruben

C.C.: 1727511055

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Nosotros **Chanataxi Nacimba, Oscar Fabricio, con cédula de ciudadanía n°1727426395, y Colina Morales, Mateo Ruben con cédula de ciudadanía n°1727511055**, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **Propuesta de sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.

Sangolquí, 31 de Julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
OSCAR FABRICIO CHANATAXI NACIMBA

Chanataxi Nacimba, Oscar Fabricio

C.C.: 1727426395

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
MATEO RUBEN COLINA MORALES

Colina Morales, Mateo Ruben

C.C.: 1727511055

Índice de contenidos

Certificación 3

Responsabilidad de Autoría 4

Resumen 11

Abstract 12

Capítulo I 13

Introducción 13

Planteamiento del problema 14

Formulación del problema o pregunta de investigación 18

Objetivos 19

Objetivo general: 19

Objetivos específicos: 19

Justificación 20

Alcance, delimitaciones y estructura del documento 23

Capítulo II 25

Metodología de la revisión bibliográfica 25

Estado del arte 26

Inteligencia artificial en reclutamiento y selección 26

Entrevistas técnicas en software 29

Simuladores y sistemas de entrenamiento para entrevistas 30

Brecha identificada en los trabajos revisados 32

Marco teórico 34

Inteligencia artificial generativa 34

Procesamiento del lenguaje natural y análisis de CV 36

Evaluación de habilidades técnicas 37
Habilidades blandas y comunicación profesional 40
Análisis de voz y prosodia 41
Análisis de video y comportamiento no verbal 43
Análisis multimodal 44
Retroalimentación automatizada y aprendizaje formativo 46
Arquitectura de software y microservicios 48
Marco conceptual 49
Capítulo III 54
Diseño metodológico de la investigación 54
Población, muestra y escenario de evaluación 56
Técnicas, instrumentos y fuentes de datos 58
Validez, consistencia y tratamiento de los instrumentos 61
Arquitectura general del sistema 61
Flujo funcional del sistema 65
Implementación de módulos principales de IA y procesamiento multimodal 69
Modelo de datos, trazabilidad y evidencias para evaluación 74
Procedimiento de evaluación y análisis de datos 76
Consideraciones éticas, privacidad y eliminación de datos 80
Capítulo IV 83
Análisis y discusión de los resultados 83
Conjunto de datos analizado 83
Preparación de datos y estrategia de análisis 84
Desempeño de los participantes por sesión 86
Comparación estadística entre sesiones 90
Análisis descriptivo de componentes de evaluación 94

Usabilidad y percepción de los participantes 97
Resultado del cuestionario SUS 97
Resultados del cuestionario complementario 99
Discusión integrada de los hallazgos 101
Limitaciones del análisis 104
Capítulo V 106
Conclusiones 106
Recomendaciones 109
Referencias 112

Índice de tablas

Tabla 1 Dimensiones del cuestionario propio complementario 60

Tabla 2 Catálogo de habilidades blandas considerado en las preguntas comunicativas 70

Tabla 3 Entradas y salidas de los prompts principales del sistema 71

Tabla 4 Ponderaciones utilizadas para la evaluación por tipo de pregunta 73

Tabla 5 Entidades y evidencias utilizadas para la evaluación 76

Tabla 6 Procedimiento de evaluación por sesión 78

Tabla 7 Variables, indicadores y análisis previsto 80

Tabla 8 Medidas éticas, privacidad y tratamiento de datos 82

Tabla 9 Resumen del conjunto de datos analizado 83

Tabla 10 Desempeño principal por sesión 85

Tabla 11 Estadísticos descriptivos de las puntuaciones principales por sesión 86

Tabla 12 Prueba de normalidad Shapiro-Wilk para las variables principales 90

Tabla 13 ANOVA de medidas repetidas para la comparación entre sesiones 90

Tabla 14 Comparaciones pareadas con corrección de Holm 92

Tabla 15 Componentes de evaluación por tipo de pregunta y sesión 94

Tabla 16 Resultado global del cuestionario SUS 97

Tabla 17 Resultados por dimensión del cuestionario complementario 99

Índice de figuras

Figura 1 Arquitectura general del sistema 62

Figura 2 Flujo funcional del usuario 66

Figura 3 Secuencia de carga y análisis de CV 66

Figura 4 Interfaz web para la carga de Currículum Vítae 67

Figura 5 Interfaz de instrucciones previas y consentimiento para el tratamiento de datos del participante 68

Figura 6 Interfaz del panel de simulación de entrevista durante la captura de respuestas orales y video 68

Figura 7 Interfaz del entorno integrado de desarrollo para la resolución de ejercicios de código 69

Figura 8 Flujo de datos multimodal 74

Figura 9 Evolución de las puntuaciones principales por sesión 88

Figura 10 Distribución de la puntuación global por sesión 89

Figura 11 Evolución de componentes en preguntas técnicas 95

Figura 12 Evolución de componentes en preguntas comunicativas 96

Figura 13 Distribución del puntaje SUS 98

Figura 14 Valoración por dimensión del cuestionario complementario 100

Resumen

El presente trabajo se enfocó en contribuir con una problemática que tienen los estudiantes de carreras relacionadas con desarrollo de software. Actualmente, en procesos de selección laborales se les exige diferentes habilidades blandas y técnicas, de modo que ya no es suficiente solamente saber programar. Por tal motivo, se abordó como objetivo principal construir un sistema de tipo simulador que mediante la exposición a un ambiente multimodal controlado ayudara a la habituación y adecuación ante escenarios de procesos de selección real. Los candidatos suben su Curriculum Vitae, y en base a ello, el aplicativo genera un perfil profesional y entrevistas personalizadas que, una vez culminadas presentan retroalimentación sobre el desempeño del usuario y consejos para mejorar el rendimiento y evitar errores graves. Veinte estudiantes de niveles superiores de la carrera de Ingeniería de Software participaron en la evaluación y realizaron exitosamente tres entrevistas a lo largo de una semana, para luego completar la encuesta SUS y un cuestionario personalizado. En total se analizaron 60 entrevistas mediante un enfoque descriptivo-analítico. La puntuación general tuvo una mejora observable entre la primera y tercera sesión, presentando un incremento de 6.91 a 7.51. Le acompañan preguntas de habilidades blandas y técnicas con mejoras similares, y en última instancia, los ejercicios de codificación no presentaron una mejora perceptible. La nota SUS es de aproximadamente 75 puntos. Teniendo en cuenta el análisis de los datos y la percepción de los estudiantes, se puede concluir que el sistema se aproxima con éxito a su objetivo de brindar apoyo en la preparación ante entrevistas, sin embargo, no se puede asegurar que sirve como indicativo de mejora persistente de habilidades, idoneidad laboral, ni que los puntajes reflejan el desempeño que tendría un candidato en una entrevista real.

Palabras clave: inteligencia artificial, preparación laboral, simulación de entrevistas, habilidades blandas

Abstract

This work focused on helping address a problem faced by students in degree programs related to software development. Currently, recruitment and selection processes require different soft and technical skills, so knowing how to program alone is no longer sufficient. For this reason, the main objective was to build a simulation system that, through exposure to a controlled multimodal environment, would help users become familiar with and adapt to scenarios resembling real selection processes. Candidates upload their curriculum vitae, and based on it, the application generates a professional profile and personalized interviews. Once the interviews are completed, the system provides feedback on the user's performance and advice on how to improve performance and avoid serious mistakes. Twenty upper-level Software Engineering students participated in the evaluation and successfully completed three interviews over the course of one week. They subsequently completed the SUS questionnaire and a custom questionnaire. A total of 60 interviews were analyzed using a descriptive-analytical approach. The overall score showed an observable improvement between the first and third sessions, increasing from 6.91 to 7.51. Questions related to soft and technical skills showed similar improvements, while coding exercises did not show a perceptible improvement. The SUS score was approximately 75 points. Considering the data analysis and the students' perceptions, it can be concluded that the system comes close to fulfilling its objective of supporting interview preparation. However, it cannot be claimed that it indicates lasting improvement in skills or job suitability, or that the scores reflect how a candidate would perform in an actual interview.

Keywords: artificial intelligence, job preparation, interview simulation, soft skills

Capítulo I

Introducción

La transformación tecnológica está influyendo en las actividades productivas, los trabajos y las competencias que las empresas valoran para enfrentar el dinamismo del entorno laboral. Este cambio se produce en un contexto donde las tecnologías digitales y la inteligencia artificial tienen una presencia cada vez mayor, al tiempo que hay una fuerte exigencia de habilidades analíticas, un aprendizaje constante y una adaptabilidad profesional (World Economic Forum, 2025).En el ámbito del desarrollo de software, estas transformaciones significan que el dominio de los conocimientos técnicos seguirá siendo importante, aunque tiene que vincularse con otras habilidades que facilitan la colaboración, la comunicación de decisiones, la interpretación de necesidades y el funcionamiento en entornos distintos. Se demuestra también por estudios sobre la brecha de competencias en ingeniería de software que las habilidades requeridas pueden cambiar según el rol, el dominio y las condiciones organizacionales, por lo que no es conveniente reducir el perfil profesional a un conjunto estático de competencias técnica (Akdur, 2022). Como resultado, el acceso al mercado laboral tecnológico exige no solo tener competencias, sino también demostrarlas y explicarlas durante los procesos de selección.

Estos procesos incluyen entrevistas técnicas, cuyo impacto depende de la habilidad para evaluar el razonamiento sobre problemas relacionados con la programación. Las entrevistas están basadas en diferentes formatos, destacando entre ellos los ejercicios de código, la resolución de problemas y el razonamiento verbalizado. En estas, el candidato tiene que dar una respuesta durante la exposición de sus argumentos ante un entrevistador. Este formato demanda diversas dimensiones del conocimiento declarativo: la planificación del razonamiento, la comprobación de la solución, la aptitud para expresar el curso de acción empleado y el desempeño frente a una evaluación. Estudios comparados muestran que la forma de conversar en estas entrevistas influye en la cantidad de razonamiento que queda

expuesto. Así pues, aunque los resultados evidencian que no se debe asumir que todas las empresas utilicen un procedimiento uniforme, la entrevista técnica se puede percibir como la práctica de un entorno donde están presentes la programación, la resolución de problemas y la comunicación.

Por otra parte, la inteligencia artificial se ha integrado en diversos momentos del proceso de contratación y reclutamiento de personal. Según la revisión de la literatura que se ha realizado, estas herramientas tecnológicas se utilizan para preparar anuncios, gestionar solicitudes, seleccionar CVs, interconectar con los interesados y asistir en tareas de evaluación (Hunkenschroer & Luetge, 2022). La adopción de estas tecnologías tiene como razón fundamental el interés organizacional por gestionar grandes volúmenes de información e invertir en sistemas automatizados que realicen algunas tareas. No obstante, esto no significa que todos los procesos de contratación utilicen la IA ni que las resoluciones se deleguen completamente a algoritmos. Asimismo, la evidencia institucional muestra que la puesta en marcha de las tecnologías de inteligencia artificial en el campo laboral puede implicar aspectos operativos combinados con desafíos que surgen desde la gestión de la privacidad, transparencia, discriminación y rendición de cuentas. Por consiguiente, el empleo de estas tecnologías en la contratación debería valorarse como un cambio que modifica prácticas y responsabilidades, no como una garantía de decisiones más justas o correctas.

Planteamiento del problema

A nivel mundial las etapas de contratación se han visto influenciadas por la digitalización que abarca los sistemas de información, plataformas de solicitudes de empleo y la Inteligencia Artificial en aplicaciones para tareas que, anteriormente, eran realizadas principalmente por empleados. La revisión del reclutamiento impulsado por IA identifica aplicaciones que se utilizan en la creación de avisos, la selección de currículos y la interpretación de datos procedentes del análisis de los candidatos (Hunkenschroer & Luetge, 2022). Los sistemas cambian la forma en que se lleva a cabo el reclutamiento y permiten que cada vez se usen más

mecanismos automatizados en etapas clave del proceso de selección. No obstante, la aplicación no es muy amplia: las evidencias sobre inteligencia artificial y mercado laboral indican que muchas organizaciones se encuentran en etapas iniciales o experimentales y los cambios a este respecto varían según el tamaño, sector y capacidad tecnológica (OECD, 2023). Por lo tanto, no se interpreta que el cambio suponga una sustitución total del criterio humano, implica, en cambio, un cambio gradual hacia la práctica de selección que introduce tanto nuevas oportunidades como demandas de supervisión, transparencia y rendición de cuentas.

En este entorno, la empleabilidad en el desarrollo de software no puede limitarse solamente al dominio de lenguajes de programación o al manejo de herramientas técnicas, puesto que las funciones del sector demandan conocimientos especializados y las competencias valoradas varían dependiendo del rol profesional, el dominio de aplicación y las propias características de cada organización. También, la literatura diferencia las competencias técnicas y no técnicas como partes vinculadas al desempeño profesional, entre las que destacan la resolución de problemas, el pensamiento crítico, la comunicación y la adaptación, ya que en el trabajo de software se toman decisiones que deben explicarse, coordinarse y ajustarse permanentemente a los requisitos cambiantes (Akdur, 2022).

Los participantes en experiencias formativas directamente relacionadas con el entorno laboral destacaron el desarrollo de la comunicación, la concentración y la capacidad de adaptación, incluida la necesidad de presentar información técnica a interlocutores que poseen diferentes niveles de conocimiento, experiencia y familiaridad con los temas abordados (Barr & Nabi, 2022). Por consiguiente, los procesos de selección pueden demandar que una persona demuestre no sólo los conocimientos técnicos que posee, sino también la manera en que analiza una situación, identifica posibles alternativas, justifica sus decisiones, adapta su razonamiento y comunica de forma comprensible una solución.

La amplia gama de conocimientos adquirida también es muy significativa en las entrevistas técnicas, ya que la resolución de una tarea puede suceder bajo condiciones de observación, tiempo límite y demanda de explicar los razonamientos empleados. Un estudio experimental con estudiantes encontró que la intervención de un supervisor durante una práctica de programación se asoció con mayores indicadores de estrés y diferencias en la corrección de las soluciones en comparación con la versión sin supervisión (Behroozi et al., 2020). La conclusión se aplica a un entorno y una tarea específicos, así que no asegura que cada evaluación cause estrés ni que esto defina siempre el desempeño. Sin embargo, la evidencia sugiere que las condiciones de la evaluación podrían modificar cómo muestra el participante sus habilidades. De manera adicional, la comparación de entrevistas supervisadas y asincrónicas mostró diferencias en la fluidez y la calidad de la descripción del procedimiento de resolución (Behroozi et al., 2022). Por tanto, una entrevista técnica no observa capacidades en condiciones neutrales: el diseño puede modificar tanto la resolución del problema como la comunicación del razonamiento que el examinador utiliza para evaluar el rendimiento.

Para adaptar este problema al ámbito universitario, la preparación para la incorporación al trabajo se enfrenta a una tensión entre lo aprendido durante la formación y las expectativas diversas de la industria, junto con las condiciones en que los aspirantes presentan sus capacidades. El análisis de la brecha de habilidades en la ingeniería de software observa distinciones entre los conocimientos y aptitudes promovidos en la educación superior y las necesidades de ciertos puestos y ambientes organizacionales (Akdur, 2022). No se debe interpretar que la educación universitaria no tenga relevancia; la oferta curricular y las demandas laborales evolucionan en tiempos diferentes, prioridades, e intensidades. Un estudio de mapeo sobre la industria de software identifica precisamente la desalineación entre el sector académico y el industrial, junto al rápido cambio en el mercado, como las principales causas de esta brecha (Diniz et al., 2024). Sin embargo, también señala que hay variaciones en diversos contextos y falta de datos sobre las tácticas más útiles para contrarrestar dicha brecha. El

problema educativo incluye tanto la adquisición de los conocimientos como la conexión de estos con realidades laborales específicas.

Sin embargo, la repetición de ejercicios de manera autónoma da lugar a una orientación insuficiente cuando el participante no tiene parámetros para determinar cuáles deben ser los aspectos de su desempeño que debe conservar o analizar. En una aplicación de IA orientada a mejorar el manejo verbal durante entrevistas técnicas, los participantes identificaron como útiles la simulación, los ejemplos y la retroalimentación, pero también requirieron comentarios sobre las observaciones de cómo equilibrar el razonamiento, la expresión oral y la programación, más allá de los contenidos verbales (Daryanto et al., 2025). En las prácticas universitarias de entrevistas simuladas, los comentarios y sugerencias prácticas forman parte del proceso de preparación y evaluación (Kapoor et al., 2023). Por otro lado, es posible utilizar sistemas de evaluación automática basados en modelos de lenguaje. Este mecanismo ofrece una retroalimentación adecuada, pero también presenta un riesgo de inconsistencia, sesgo, alucinaciones y explicabilidad limitada, lo cual significa que es necesario contar con una evaluación humana (Emirtekin, 2025). Como consecuencia, la formación requiere retroalimentación bien definida, contextualizada y orientadora para el razonamiento, la comunicación y la habilidad técnica, lo cual debe considerarse no como una valoración incontestable ni absoluta sino simplemente como apoyo orientador.

En la investigación específica, los estudios revisados abordan aspectos relevantes de varias de las competencias requeridas para la preparación de entrevistas laborales en software, aunque con diversas combinaciones y objetivos. La inteligencia artificial conversacional se ha empleado para replicar entrevistas técnicas, facilitar la explicación del razonamiento al escribir código y ofrecer comentarios y ejemplos de respuestas (Daryanto et al., 2025). Además, algunos marcos emplean agentes que representan entrevistadores y candidatos, formulan cuestionarios utilizando los currículums y descripciones del puesto, y evalúan la correspondencia entre ambos (Sun et al., 2025). Otro método ha incorporado contenido verbal

y señales no verbales extraídos de entrevistas grabadas, pero orientado a evaluación de candidatos y de constructos vinculados con liderazgo y personalidad (Kassab et al., 2026). Desde la perspectiva de este proyecto, estas contribuciones cubren partes relacionadas, pero en el conjunto de documentos analizados no se halló una integración similar que reuniera la preparación de desarrollo de software, información curricular, interacción de entrevista, señales multimodales y la retroalimentación bajo una misma finalidad académica.

Por eso la brecha se define en términos de integración funcional y finalidad formativa, no de la presencia relativa de sistemas similares. El problema concreto estriba en cómo combinar, dentro de un contexto académico dirigido a estudiantes o recién graduados en TIC, el análisis estructurado de la información curricular, la realización de entrevistas técnicas, el tratamiento de las respuestas, la revisión del código y la utilización de señales sonoras y visuales para efectuar observaciones personalizadas. El procesamiento automático de datos curriculares permite estrategias para contextualizar preguntas sin que el currículo se convierta en un juicio de idoneidad (Vukadin et al., 2021), mientras que la práctica de razonamiento verbalizado y la simulación conversacional ofrecen técnicas de interacción para entrevistas técnicas (Daryanto et al., 2025). El código de revisión educativa proporciona referencias para incluir validaciones técnicas con descripciones instructivas (Lee & Joe, 2025), y los marcos multimodales señalan formas de procesar al mismo tiempo el lenguaje verbal y el no verbal; no obstante, sus inferencias y objetivos de selección no deben traducirse directamente al entorno educativo (Kassab et al., 2026). En definitiva, la propuesta pretende analizar esta integración a fin de apoyar la preparación y la reflexión del estudiante, sin sustituir los procesos reales de selección ni admitiendo que tal integración garantice resultados laborales o formativos.

Formulación del problema o pregunta de investigación

El vacío de integración hallado en la colección de estudios revisados pone de relieve la necesidad de explorar la manera de coordinar, con un fin formativo, los múltiples elementos que conforman la preparación para una entrevista laboral en el sector de desarrollo de

software. Ello no significa desconocer la eficacia de los enfoques anteriores, sino la posibilidad de concentrar en un espacio académico aspectos que habitualmente se abordan con criterios diversificados. De ahí que la pregunta esencial del trabajo de investigación sea: ¿Cómo diseñar un sistema multimodal basado en IA para simular entrevistas de trabajo en desarrollo de software, integrando análisis de CV, interacción de entrevista, evaluación técnica y retroalimentación personalizada con fines formativos?

Para organizar el estudio, la pregunta principal se descompone en tres subpreguntas relacionadas con las etapas de comprensión, diseño y evaluación del proyecto. Primero, ¿qué características y necesidades presentan los procesos de entrevista laboral en desarrollo de software y la preparación de estudiantes o personas recién graduadas del área de desarrollo de software? Segundo, ¿qué arquitectura y componentes funcionales requiere un sistema multimodal basado en inteligencia artificial para simular entrevistas laborales de software? Tercero, ¿cómo evaluar el desempeño y la utilidad del sistema propuesto en un contexto académico controlado? Estas interrogantes delimitan el recorrido de la investigación sin presuponer la eficacia del sistema ni ampliar su aplicación hacia decisiones reales de contratación.

Objetivos

Objetivo general:

Desarrollar un sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales del sector de desarrollo de software, que integre la evaluación de habilidades técnicas y blandas, proporcionando retroalimentación personalizada a los candidatos.

Objetivos específicos:

Analizar el contexto y las características generales de los procesos de entrevista laboral en el sector de desarrollo de software.

Definir la arquitectura del sistema basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software.

Evaluar el sistema propuesto en función de su desempeño y utilidad en el contexto planteado.

Justificación

Desde el punto de vista teórico, la investigación es justificada por el encuentro entre líneas de trabajo generalmente estudiadas con propósitos y unidades de análisis distintos: preparación para entrevistas técnicas, interacción conversacional, análisis multimodal y retroalimentación automática. Las herramientas basadas en modelos de lenguaje mostraron que la simulación puede incorporar diálogos, resolución de código, verbalización del razonamiento y comentarios finales como elementos de formación en la experiencia (Daryanto et al., 2025). Además, los enfoques multimodales admiten procesar de manera conjunta texto, audio e imagen; sin embargo, estos hallazgos dependen del material, tarea y criterios de evaluación (Putra et al., 2024). La revisión de los informes automatizados con modelos de lenguaje muestra que la personalización y escalabilidad ofrecen oportunidades, junto con problemas de confiabilidad, sesgo, protección de datos, transparencia y necesidad de control humano (Emirtekin, 2025). Los marcos que reúnen material verbal e indicadores no verbales muestran la viabilidad técnica de combinar tipos de mensajes, pero también el peligro de trasladar inferencias sensibles desde la selección hacia un ambiente educativo (Kassab et al., 2026). Por ello, el aporte teórico esperado sería organizar estas perspectivas en torno a una finalidad formativa aplicada y no presumir que la combinación produzca una medida o una interpretación completa u objetiva del participante.

Desde una perspectiva práctica, este proyecto busca proporcionar un entorno controlado para que estudiantes o profesionales recién graduados empiecen a experimentar las dinámicas de una entrevista técnica sin estar en situaciones reales. Las pruebas asistidas por inteligencia artificial conversacional pueden ayudar a resolver problemas mientras el candidato

habla de las decisiones que toma e inmediatamente recibe sugerencias acerca de la coordinación entre el pensamiento, la descripción verbal y la programación (Daryanto et al., 2025). En instituciones educativas, las actividades integradas en cursos de estructuras de datos y algoritmos han mezclado aspectos como preparación, ejercicios y entrevistas simuladas para enseñar al alumnado esas necesidades (Kapoor et al., 2023). Otros prototipos con agentes virtuales han explorado la percepción estudiantil y la evaluación externa de entrevistas simuladas, aunque con muestras reducidas y diseños específicos (Lunn & Brewster, 2023). Asimismo, se han analizado evaluaciones de programas a lo largo de varios cursos para entender las expectativas y apoyar las prácticas futuras mediante estas observaciones (Kapoor & Gardner-McCune, 2025). Tales experiencias muestran la pertinencia de construir un espacio para la prueba y la reflexión, pero no aseguran que un prototipo funcional mejore los resultados en las entrevistas reales.

La justificación social está relacionada con las barreras que podrían afectar el acceso a las formaciones, la orientación o la familiarización con los procesos de contratación en informática. Un estudio cualitativo con estudiantes de grupos minoritarios descubrió que existen obstáculos relacionados con la incertidumbre, las técnicas de entrevista, la dedicación de tiempo necesario para prepararse, la gestión de la ansiedad e inclusión, y se formularon propuestas dirigidas a los alumnos, la academia e industria (Lunn et al., 2022). Las intervenciones académicas de orientación mostraron que estas instituciones pueden aproximarse más a la práctica profesional de entrevistas técnicas en una asignatura, pero depende del contexto institucional (Kapoor et al., 2023). También se observó que las experiencias de entrevistas y contratación se correlacionan con la identidad computacional y el sentido de pertenencia de los estudiantes, pero dicha relación no demuestra causalidad (Lunn et al., 2021). Se evidenció asimismo la relevancia de competencias como la comunicación, la adaptabilidad y la presentación de información técnica en experiencias formativas relacionadas con el trabajo (Barr & Nabi, 2022). En ese contexto, la propuesta podría contribuir a expandir

las oportunidades de practicar y reflexionar sobre experiencias relacionadas con el sector TI para estudiantes y recién graduados de esta área, sin garantizar un empleo, acabar con las desigualdades ni reemplazar los apoyos pedagógicos e institucionales

El enfoque multimodal justifica por sus criterios tecnológicos y metodológicos los análisis que realizan las actividades de una entrevista. El texto explica qué es lo que se recupera a través del contenido de las respuestas; el código escrito puede revisarse mediante pruebas y comentarios orientados al aprendizaje; y el audio o el video aportan a la visualización, siempre que su uso se limite a lo previsto. Las investigaciones de modelos sobre evaluación de videos han demostrado que la combinación de diferentes formatos (textual, acústico y visual) es factible, y que existen problemas de rendimiento y equidad (Putra et al., 2024). La investigación sobre voz señala también que la grabación puede contener información sensible, por lo que la necesaria vigilancia debe estar dirigida hacia el equilibrio entre utilidad y protección (Mawalim et al., 2026). Los marcos multimodales para selección ilustran opciones de integración, aunque sus inferencias sobre rasgos o potencial ejecutivo no resultan trasladables al propósito formativo del proyecto (Kassab et al., 2026). En cuanto a la tecnología, los sistemas educativos de revisión de código muestran beneficios en la combinación de casos de prueba con retroalimentación explicativa (Lee & Joe, 2025). Por tanto, la multimodalidad es un recurso para respaldar conclusiones observacionales, no para inferir emociones, personalidad, honestidad o competencia profesional.

La aplicación de modelos de inteligencia artificial generativa y procesamiento del lenguaje natural justifica su uso dentro de la simulación por sus roles de soporte en diversas tareas. Los modelos conversacionales pueden facilitar la continuidad de conversaciones, adaptar las preguntas a la evolución de cada sesión y ofrecer texto como ejemplo o evaluación, aunque resultan meramente recomendaciones que requieren control (Daryanto et al., 2025). Para contextualizar la entrevista, las técnicas de PLN pueden extraer y estructurar información explícita de currículos de formato libre, como educación, experiencia o competencias

declaradas, sin convertir esa información en una decisión sobre aptitud (Vukadin et al., 2021). Las arquitecturas multiagente demuestran igualmente que los modelos lingüísticos pueden simular roles y generar preguntas basadas en el currículum y descripciones de puestos, además de adaptar los patrones de comunicación (Sun et al., 2025). No obstante, los datos sobre la evaluación automatizada presentan diferencias respecto al grado de acuerdo con los profesores humanos y el riesgo de alucinaciones, sesgo y falta de coherencia (Emirtekin, 2025). Por tanto, tales tecnologías se integran como soporte para la conversación y la retroalimentación formativa, con estándares verificables, trazabilidad y supervisión por parte de una figura humana, y no como sustituto directo del instructor, tutor o seleccionador

Alcance, delimitaciones y estructura del documento

El alcance funcional del proyecto engloba la creación de un prototipo para entrevistas simuladas en el sector del desarrollo de software, además de brindar retroalimentación formativa. La función a desarrollar está centrada en el análisis de los datos presentados en un currículum vitae, con la finalidad de ofrecer una simulación realista para las conversaciones, la formulación y respuesta de problemas técnicos u obstáculos relacionados con el código, y el cuidado y presentación adecuada de datos audiovisuales, sonoros y textuales para la sesión de la entrevista simulada. Todos estos elementos contribuyen a elaborar juicios sobre el rendimiento durante el simulacro. Por ello es conveniente destacar que el sistema no sustituirá a reclutadores, docentes o asesores; no participará en los procesos de contratación, no emitirá certificados sobre habilidades profesionales ni analizará las capacidades laborales del individuo. Además, la gestión de audio y video no tiene el objetivo de medir emociones, rasgos de personalidad, sinceridad ni caracterización psicológica de la persona implicada.

La población de estudio estará formada por estudiantes o recién graduados en estudios afines a las tecnologías de la información y la comunicación que quieran conocer mejor las entrevistas de trabajo en el ámbito del desarrollo de software, mientras que el prototipo y su evaluación se harán en un entorno académico controlado donde se podrá comprobar su

funcionamiento, su rendimiento técnico y su utilidad percibida según los criterios metodológicos establecidos para el desarrollo del estudio. Este alcance excluye la implantación del sistema en procesos empresariales reales, su adopción como herramienta de apoyo a la toma de decisiones de contratación y la comprobación de posibles efectos sobre la inserción laboral o la empleabilidad, por lo que los resultados obtenidos describirán el comportamiento, el desempeño y la valoración del prototipo bajo las condiciones académicas previamente establecidas, sin generalizar su utilidad a todas las poblaciones, organizaciones, sectores profesionales o modalidades de entrevista laboral.

El documento está compuesto por cinco capítulos que van desarrollando paulatinamente la investigación. El primer capítulo aborda la introducción, la formulación del problema, los objetivos, la justificación, el alcance y las delimitaciones del proyecto. El segundo capítulo agrupa la metodología de revisión de literatura, el estado del arte, el marco teórico y el marco conceptual que sostienen el trabajo. El tercer capítulo presenta la metodología de investigación, así como el proceso de elaboración del prototipo, incluidos los criterios y las consideraciones de diseño pertinentes. El cuarto capítulo incluye los hallazgos resultantes de la evaluación y su interpretación en relación con los objetivos y los límites del trabajo. El quinto capítulo ofrece las conclusiones del estudio y las propuestas de acciones para futuras líneas de investigación.

Capítulo II

Metodología de la revisión bibliográfica

La revisión bibliográfica tuvo como fin reunir un corpus documental que sustentara las decisiones conceptuales y metodológicas de la propuesta, por tanto, el proceso fue diseñado con el fin de localizar estudios sobre entrevistas de trabajo y técnicas de software, analizar la aplicación de la inteligencia artificial en actividades de simulación y evaluación y descubrir los conceptos necesarios para orientar la planificación del sistema. La revisión, también, permitió establecer relaciones entre contribuciones provenientes de campos como el reclutamiento, la educación en ciencias de la computación, el procesamiento de lenguaje natural, el análisis multimodal, la evaluación automatizada y la arquitectura de software, mediante un proceso sistemático de búsqueda, evaluación, análisis y síntesis acorde con la revisión bibliográfica sistemática como herramienta de apoyo para trabajos científicos (Codina, 2020). La adopción de diferentes enfoques fue fundamental, ya que el problema objeto de estudio involucró aspectos de índole educativa, algorítmica y arquitectónica (Torraco, 2016), y por tal motivo, se utilizó la revisión bibliográfica como herramienta para conceptualizar el problema, valorar las alternativas y establecer los fundamentos teóricos y metodológicos del proyecto.

El enfoque utilizado es una revisión bibliográfica exploratoria sistematizada. Esta denominación indica que se recopilaron, revisaron y categorizaron las fuentes bajo una serie de criterios preestablecidos; sin embargo, esto no significa que se siguiera una estrategia sistemática completa (Grant & Booth, 2009). Las revisiones tienen distintos objetivos, métodos de búsqueda, valoración y análisis, por lo que no se pueden clasificar formalmente por el uso de herramientas de organización. El carácter exploratorio permitió mapear literatura heterogénea y reconocer líneas relevantes, pero el procedimiento no siguió todas las etapas requeridas para denominarse scoping review formal (Arksey & O'Malley, 2005). La revisión sistematizada proporciona un marco más adecuado al trabajo realizado (Codina, 2020). No se

utilizó PRISMA ni se realizó un metaanálisis, ni se intenta verificar su efectividad en términos absolutos, tampoco se pretende asegurar que abarque la totalidad de la investigación existente.

Este proceso de búsqueda resultó efectivo ya que permitió encontrar fuentes mediante sitios web editoriales, bases digitales de bibliotecas y repositorios académicos, como IEEE, AIS eLibrary, arXiv y repositorios de universidades y entidades institucionales. Incluso se consultaron otras fuentes en buscadores y plataformas académicas; sin embargo, no se registró un recuento exhaustivo con fechas, los resultados y los pasos para reconstruir cada búsqueda de manera individual. De ahí que no se presente una lista cerrada completa ni se afirme haber consultado todas las fuentes adecuadas. La trazabilidad del proceso se mantuvo registrando las fuentes seleccionadas, conservando los archivos recuperados y comprobando la correspondencia entre los datos bibliográficos del trabajo revisado y el contenido examinado. La falta de un histórico completo de la búsqueda se identifica como una limitación metodológica.

Se consideraron documentos relacionados con el problema, con los objetivos y los componentes esperados del sistema, tales como artículos, tesis, revisiones, libros técnicos, informes institucionales y trabajos académicos. Las fuentes fueron seleccionadas basándose en su actualidad en áreas de rápido cambio e incluyeron estudios anteriores cuando estos proporcionaban bases metodológicas o conceptuales válidas. Los documentos sin acceso directo a los datos, registros duplicados, confusos o con relación limitada al objetivo, así como estudios cuyo contenido fue difícil de evaluar, fueron excluidos. Dichos criterios permiten organizar el corpus con mayor claridad, sin reemplazar un diagrama estándar de selección ni convertir el procedimiento en una revisión sistemática rigurosa (Okoli, 2015).

Estado del arte

Inteligencia artificial en reclutamiento y selección

Se ha utilizado la inteligencia artificial en diversas etapas del reclutamiento y la selección, pero su uso y autonomía dependen de las organizaciones y las aplicaciones. La

revisión sobre el reclutamiento habilitado por inteligencia artificial identifica usos vinculados a la preparación de ofertas laborales, la búsqueda de candidatos, el filtrado de currículums, la comunicación con los postulantes y el apoyo a las actividades de evaluación (Hunkenschroer & Luetge, 2022). En la gestión de perfiles, se han estudiado representaciones vectoriales y medidas de similitud para relacionar información de currículums con descripciones de puestos, con la finalidad de facilitar la organización o priorizar a los postulantes (Bevara et al., 2025). Estas tareas no implican que el sistema pueda determinar por sí mismo la aptitud profesional de una persona, dado que la correspondencia semántica solo constituye una parte del proceso. Desde una perspectiva más amplia, la adopción de la IA en el mundo laboral sigue variando y depende de factores como el sector, el tamaño de la organización y sus capacidades tecnológicas (OECD, 2023). Por ello, la automatización debe considerarse como una ayuda variable en tareas concretas, y no como una sustitución general de la selección humana.

Entre las oportunidades que ofrecen estas tecnologías se encuentran la capacidad para procesar cantidades extensas de datos, integrar información dispersa y automatizar tareas repetitivas, lo cual resulta crucial en el proceso de contratación. Los sistemas de análisis de perfiles de candidatos e identificación de coincidencias pueden simplificar la búsqueda de datos relevantes y facilitar la correlación inicial entre requisitos y antecedentes proporcionados (Bevara et al., 2025). La automatización también puede acelerar las respuestas, permitir interacciones múltiples en diversos momentos y ajustar el contenido o la comunicación según las condiciones de un puesto o fase del proceso (Hunkenschroer & Luetge, 2022). En el ámbito laboral general, estos avances tienen más que ver con transformaciones en el reparto del trabajo y con formas de cooperación entre personas y máquinas, que con una simple sustitución de roles profesionales (OECD, 2023). Sin embargo, estos beneficios potenciales están sujetos a diversos factores, como la calidad de los datos utilizados, la funcionalidad del sistema propuesto, el ambiente de trabajo, el seguimiento disponible y una clara asignación de responsabilidades. El mejor aprovechamiento de la capacidad para analizar grandes

cantidades de datos no es suficiente por sí solo para tomar decisiones más precisas, justas o explicables.

Con todas esas opciones, la aplicación de la IA en la selección entra en conflictos derivados de sesgo, opacidad, privacidad, discriminación y delegación excesiva de decisiones. Las aplicaciones tienen el potencial de reproducir o magnificar tendencias presentes en datos históricos, y la complejidad de algunos modelos dificulta entender por qué ciertas candidaturas fueron seleccionadas o rechazadas (Hunkenschroer & Luetge, 2022). En el cribado de currículos mediante modelos de lenguaje y recuperación semántica se han observado disparidades vinculadas con género, raza y su intersección, lo que evidencia que la representación algorítmica no constituye una base neutral para valorar personas (Wilson & Caliskan, 2024). Los riesgos también dependen de la modalidad procesada, ya que los sistemas de entrevista en vídeo deben considerar las diferencias de desempeño entre grupos, las equivalencias de precisión e igualdad y las compensaciones de calidad (Putra et al., 2024), mientras que el análisis de voz introduce preocupaciones relacionadas con la privacidad y la exposición de características sensibles (Mawalim et al., 2026), por lo que estos procedimientos requieren supervisión humana, registro de estándares, seguridad de los datos y mecanismos de evaluación, sin que puedan considerarse ni evaluadores objetivos ni sustitutos automáticos de la decisión profesional humana.

Esto resulta central para el proyecto, cuyo objetivo no consiste en apoyar una decisión de contratación, sino facilitar una experiencia de simulación y de retroalimentación académica en la preparación para entrevistas de software. Los estudios sobre IA conversacional aplicada a la práctica de razonamiento oral muestran posibilidades para explorar la resolución de problemas, recibir ejemplos y revisar cómo comunicar una solución, aunque sus resultados dependen del diseño de la interacción e incluso de las necesidades de los participantes (Daryanto et al., 2025). Por su parte, las actividades universitarias para preparación de entrevistas técnicas demuestran que la simulación puede integrarse en escenarios de

aprendizaje para familiarizar al estudiantado con las preguntas, los ejercicios y los métodos para explicar sus decisiones (Kapoor et al., 2023). En este proyecto, tales funciones están concebidas en términos de formación y deben tener en cuenta los principios de transparencia, responsabilidad y supervisión relacionados con la IA en reclutamiento (Hunkenschroer & Luetge, 2022).

Entrevistas técnicas en software

Entre los procesos de selección en el sector del desarrollo de software las entrevistas técnicas pueden variar su finalidad y formato según la compañía, la función y el enfoque. Así pues, no existe un formato único en el sector. En los escenarios analizados, esta actividad combina la creación de una solución con la descripción de la argumentación empleada. De este modo, el comportamiento del candidato comprende tanto la respuesta técnica como la comunicación de conclusiones, supuestos e hipótesis. Los hallazgos en relación con entrevistas de programación llevadas a cabo bajo supervisión muestran que el rendimiento puede depender del entorno de la tarea (Behroozi et al., 2020). Además, la comparación entre la modalidad supervisada y la modalidad asincrónica revela distinciones en el enfoque verbal y la experiencia de resolución (Behroozi et al., 2022). De ahí que tales entrevistas ofrezcan la posibilidad de evaluar aspectos técnicos y comunicativos, aunque no constituyen una medición completa u objetiva de la aptitud profesional.

Entre los componentes frecuentes se encuentran preguntas sobre conceptos de computación, ejercicios de programación, análisis de problemas, construcción o revisión de soluciones y explicación de las decisiones adoptadas. Las actividades universitarias de preparación han incorporado problemas de estructuras de datos y algoritmos, prácticas de codificación y entrevistas simuladas para aproximar al estudiantado a tareas observadas en determinados procesos de contratación (Kapoor et al., 2023). En algunas modalidades, la verbalización del pensamiento acompaña la escritura de código y permite que el entrevistador conozca cómo se interpreta el problema, se selecciona una estrategia y se comprueba una

solución. Los estudios sobre entrevistas asincrónicas muestran que esta comunicación puede modificarse cuando desaparece la presencia inmediata de un entrevistador, lo que confirma que el formato influye en la información producida durante la actividad (Behroozi et al., 2022). Las evaluaciones retrospectivas de programas de preparación también señalan la importancia atribuida a conocer expectativas, practicar preguntas y recibir orientación (Kapoor & Gardner-McCune, 2025). Estos componentes deben entenderse como prácticas recurrentes en los contextos analizados, no como requisitos obligatorios de toda entrevista técnica.

La dificultad de estas entrevistas no se limita al conocimiento sobre programación. La resolución de una tarea mientras se controla el tiempo, se interpreta el problema y se explica el razonamiento puede aumentar la carga de estas situaciones, especialmente cuando la entrevista implica una observación directa. En un estudio con estudiantes, la condición supervisada se asoció con índices de estrés más altos y diferencias en la corrección de las soluciones respecto de una condición de entrevista privada (Behroozi et al., 2020). Los resultados se basan en una muestra y un diseño determinados; sin embargo, no se puede concluir que todas las entrevistas bajo supervisión debiliten la calidad del resultado ni que el estrés se produzca de la misma manera. Asimismo, las experiencias vinculadas con las entrevistas técnicas y la contratación se han asociado con factores relacionados con la identidad computacional y el sentido de pertenencia (Lunn et al., 2021). La relación aquí descrita no establece causas, sino que expone la influencia que puede tener el formato sobre ciertos aspectos. Las experiencias también pueden dar lugar a otras formas de estrés relacionado con la comunicación durante la resolución, añadiendo dificultades para ciertos aspirantes.

Simuladores y sistemas de entrenamiento para entrevistas

Los trabajos de preparación para entrevistas han motivado la creación de sistemas con distintas configuraciones, como plataformas web, agentes virtuales, modelos conversacionales y ambientes que combinan datos del currículo con interacción oral o textual, y aunque todos

estos trabajos comparten la intención general de ofrecer oportunidades de práctica, no se pueden considerar directamente equivalentes porque difieren en la población a la que están dirigidos, el tipo de entrevista, la forma de la interacción, las funciones que implementan y los criterios de evaluación, mientras que algunos se enfocan específicamente en la verbalización del razonamiento durante tareas técnicas, como la inteligencia artificial conversacional diseñada para apoyar la práctica de think-aloud en estudiantes de computación (Daryanto et al., 2025). Otros presentan los roles de entrevistador y entrevistado a través de diversos agentes y regulan la conversación según la información contextual (Sun et al., 2025). Otros trabajos exploran gerentes virtuales con aspecto tridimensional y evalúan la percepción de los estudiantes al respecto (Lunn & Brewster, 2023), junto con plataformas multilingües que combinan currículos, recuperación de información y conversación para el entrenamiento (Nguyen et al., 2025). De esta forma, el panorama evidencia distintas aproximaciones para simular situaciones de entrevista, lo cual amerita desglosar las funcionalidades y los entornos donde fueron evaluados dichos simuladores.

Los sistemas conversacionales son relevantes porque pueden realizar intercambios, formular preguntas y ajustar la secuencia de práctica a las respuestas del participante. En la preparación para entrevistas técnicas, se ha utilizado una IA conversacional que ayuda a solucionar problemas de código, requiere describir la secuencia y ofrece características de simulación, ejemplos y retroalimentación (Daryanto et al., 2025). De manera multiagente, existen marcos que distribuyen roles simulando a los participantes, mantienen memoria de la interacción y modifican las estrategias, aunque este concepto no excluye entrevistas técnicas universitarias (Sun et al., 2025). Se ha propuesto usar modelos generativos de entrevistadores para ampliar la práctica de código de programación y generar preguntas para entrevistas, pero se basa en la experiencia y es de alcance limitado (Mejias et al., 2023). Por su parte, plataformas de entrenamiento con recuperación aumentada y reclutadores virtuales muestran posibilidades de personalizar preguntas a partir del currículo y de un puesto, en contextos

distintos de la entrevista técnica universitaria (Nguyen et al., 2025). Estas propuestas demuestran competencias conversacionales, pero sus aplicaciones son de menor alcance.

La retroalimentación automatizada aparece con distintos grados de profundidad y validación. En sistemas de práctica conversacional, los comentarios pueden referirse al razonamiento verbalizado, a la coordinación entre explicación y resolución, o a ejemplos de posibles respuestas; sin embargo, estudios con muestras pequeñas reflejan principalmente percepciones y necesidades de diseño (Daryanto et al., 2025). En el ámbito educativo más amplio, la revisión de evaluaciones automatizadas con modelos de lenguaje identifica posibilidades de personalización y escalabilidad, junto con problemas de inconsistencia, sesgo, alucinación y explicabilidad que requieren supervisión humana (Emirtekin, 2025). Para tareas de programación, se han desarrollado sistemas que combinan casos de prueba y revisión generada por modelos para producir observaciones sobre errores y posibles mejoras, aunque los resultados corresponden a contextos educativos específicos y no representan una medición completa de la capacidad técnica (Lee & Joe, 2025). Asimismo, los estudios cualitativos sobre entrevistas técnicas simuladas con IA recogen necesidades vinculadas con conversación, pizarra y comentarios posteriores, pero su condición de trabajo preliminar limita la fuerza de sus conclusiones (Gomez et al., 2025). Por tanto, el feedback automatizado debe analizarse según su objeto, evidencia y mecanismo de control, no como una función uniforme o necesariamente correcta.

Brecha identificada en los trabajos revisados

La comparación de los antecedentes permite reconocer contribuciones relevantes en varias dimensiones de la preparación y evaluación de entrevistas. Los sistemas de IA conversacional han explorado la práctica del razonamiento verbalizado durante la resolución de problemas técnicos y la generación de observaciones sobre la interacción (Daryanto et al., 2025). En el manejo de los datos curriculares, se han elaborado procedimientos para seleccionar, etiquetar y ordenar datos expresados en currículos heterogéneos (Vukadin et al.,

2021). Otros marcos utilizan agentes para representar a los participantes, establecer preguntas y ajustar estrategias de entrevista mediante datos contextuales (Sun et al., 2025). Desde un punto de vista técnico, las herramientas de revisión automática de código combinan pruebas con comentarios instructivos dirigidos al aprendizaje (Lee & Joe, 2025). Estas referencias indican, en consecuencia, que la conversación, el perfil curricular, la entrevista y la evaluación técnica tienen fundamentos específicos. Cada uno de estos ámbitos opera dentro de diferentes objetivos y entornos, por lo que su unificación es incompatible con un estudio que no considere estas circunstancias. Los resultados de los trabajos no pueden generalizarse por igual; cada sistema fue diseñado con fines distintos y requiere su propio análisis.

La revisión de la literatura no encontró un sistema que combinara eficazmente el análisis estructurado del currículum, la simulación de una entrevista técnica de software, el procesamiento de respuestas y de código, la inclusión adecuada de datos multimodales y la producción de retroalimentación personalizada en un entorno educativo. La brecha existe tanto en términos funcionales y de finalidad como de formación integral, pero no se trata de su total ausencia en el campo tecnológico; los métodos de extracción del currículum facilitan la contextualización de la interacción; sin embargo, no evalúan la capacidad profesional del candidato (Vukadin et al., 2021); los marcos multiagente ofrecen estrategias de comunicación, memoria y adaptación, aunque sus objetivos se encuentran más en contratación y perfilación (Sun et al., 2025). La revisión automatizada de código proporciona validaciones y observaciones técnicas sin evaluar de manera completa la capacidad profesional (Lee & Joe, 2025), mientras que los modelos multimodales muestran potencial para combinar texto, sonido y video, pero presentan problemas en lo referente a la equidad e interpretaciones (Putra et al., 2024). La oportunidad identificada consiste en estudiar cómo articular estas funciones sin trasladar indebidamente objetivos de selección hacia un contexto educativo.

Esta brecha orienta el diseño de un sistema multimodal basado en inteligencia artificial como herramienta de apoyo a la preparación, práctica y reflexión de estudiantes o personas

recién graduadas que deben enfrentarse a entrevistas laborales relacionadas con el desarrollo de software, si bien la integración propuesta deberá conservar una finalidad estrictamente formativa, de manera que las salidas automáticas se utilicen únicamente para ofrecer observaciones sobre aspectos delimitados de la práctica, sin certificar competencias profesionales, predecir posibilidades de contratación ni sustituir la valoración y el criterio humano. La literatura sobre IA en el reclutamiento exige considerar principios de transparencia, responsabilidad, privacidad y supervisión en el procesamiento de datos de candidatos (Hunkenschroer & Luetge, 2022), los sistemas multimodales deben examinar posibles disparidades y evitar que la combinación de diferentes señales se interprete como garantía de precisión u objetividad (Putra et al., 2024), y el procesamiento de voz demanda medidas proporcionales de protección de datos y restricciones específicas frente a inferencias sensibles (Mawalim et al., 2026). Además, la evaluación automatizada mediante modelos de lenguaje conlleva riesgos de inconsistencia, sesgo y falta de explicabilidad, por lo que es necesario validar, supervisar y revisar cuidadosamente la retroalimentación generada (Emirtekin, 2025), por lo que, teniendo en cuenta todas estas delimitaciones conceptuales, éticas y metodológicas, el estado del arte conduce al desarrollo del marco teórico, donde se expondrán los conceptos, enfoques y fundamentos técnicos necesarios para comprender adecuadamente los distintos componentes que conforman el sistema propuesto.

Marco teórico

Inteligencia artificial generativa

En el marco de este estudio, la inteligencia artificial generativa se limita operativamente al uso de modelos capaces de generar o adaptar contenido textual a partir de instrucciones, información contextual y el desarrollo de una interacción, sin que esta caracterización pretenda ofrecer una taxonomía exhaustiva del campo, sino precisar las funciones pertinentes para la simulación, entre ellas plantear preguntas, generar respuestas o ejemplos, mantener un intercambio conversacional y producir observaciones textuales, tal como lo demuestran los

sistemas de práctica de entrevistas técnicas en los que un modelo conversacional puede intervenir durante la verbalización del razonamiento, solicitar explicaciones y ofrecer ejemplos o comentarios vinculados con la actividad (Daryanto et al., 2025). Los marcos multiagente expanden esto añadiendo representaciones de los roles, la gestión de la memoria del diálogo y las estrategias de entrevista (Sun et al., 2025). A su vez, la literatura sobre evaluación automatizada con modelos de lenguaje documenta el uso de contenido generado para apoyar la valoración y la retroalimentación en contextos educativos (Emirtekin, 2025). Es decir, la IA generativa es percibida aquí como un recurso de producción e interacción verbal con propósito didáctico. En lugar de intentar entender el razonamiento humano o evaluar las habilidades laborales, la IA generativa opera como una herramienta útil para la formación y mejora académica.

Las habilidades relevantes de estos modelos están relacionadas con la persistencia del diálogo y la adaptabilidad al contexto. Durante una práctica, un sistema conversacional puede formular preguntas de seguimiento, solicitar que el participante explique una decisión y generar ejemplos que faciliten la revisión de su respuesta (Daryanto et al., 2025). En configuraciones multiagente, los modelos pueden asumir roles diferenciados, intercambiar datos y guiar el curso de la interacción según el rastro adquirido (Sun et al., 2025). También podrían facilitar la creación de retroalimentación sobre tareas técnicas; sin embargo, su utilidad está vinculada a la claridad de los criterios y al mecanismo de contrastación con hechos válidos. Las investigaciones sobre retroalimentación de programación muestran que los productos de modelos de lenguaje pueden tener aciertos, pero también desaciertos y alucinaciones (Silva & Costa, 2025). Así pues, varios sistemas combinan la generación verbal con situaciones prácticas y métodos adicionales de corroboración para aumentar la exactitud de los comentarios (Lee & Joe, 2025). Estas habilidades permiten automatizar algunas actividades, pero no transforman al modelo en un juez inapelable ni reemplazan la revisión humana.

Procesamiento del lenguaje natural y análisis de CV

En este estudio, el procesamiento del lenguaje natural se aborda desde las funciones necesarias para analizar información textual contenida en currículos, no como una descripción exhaustiva de todo el campo. La implementación de métodos permite la segmentación de los documentos, la identificación de secciones y la extracción de datos explícitos como educación, experiencia laboral, conocimientos, idiomas o capacidades declaradas. Los métodos desarrollados para currículos en formato libre y en distintas lenguas indican que los modelos de lenguaje pueden transformar información de formatos desiguales a representaciones estructuradas, aunque el rendimiento de dichos métodos está condicionado por los datos, idiomas e incluso los formatos considerados (Vukadin et al., 2021). A través de la extracción, también se puede convertir texto en un formato vectorial que facilita la comparación entre elementos de un currículo y descripciones textuales correlativas (Bevara et al., 2025). Estas acciones contribuyen a la organización del contenido declarado en forma de información útil para otros elementos del sistema. Sin embargo, la extracción de una habilidad mencionada en un currículo no demuestra que esa persona la domine realmente; ni siquiera una representación vectorial puede considerarse por sí sola como un juicio sobre la aptitud o capacidad de la persona dentro de un puesto laboral.

Para el sistema propuesto, el análisis de CV puede apoyar la contextualización de la entrevista mediante la identificación de temas potenciales vinculados con la trayectoria declarada por el participante. La estructuración de secciones y entidades permite recuperar referencias a formación, experiencia, tecnologías o proyectos que podrían utilizarse para orientar preguntas pertinentes durante la simulación (Vukadin et al., 2021). Asimismo, las representaciones semánticas pueden ayudar a relacionar términos y contenidos expresados de manera diferente, siempre que la similitud obtenida se interprete como proximidad textual y no como una clasificación definitiva de candidatos (Bevara et al., 2025). Los marcos de entrevista multiagente muestran que la información curricular y el contexto de un puesto pueden

emplearse para formular preguntas y adaptar estrategias de diálogo (Sun et al., 2025). En un entorno formativo, estas funciones se orientan a personalizar parcialmente la práctica y ampliar las oportunidades de explicación del participante. El currículo actúa, así como una fuente contextual y autodeclarada; no certifica competencias, no permite inferir capacidades no expresadas y no debe convertirse en un mecanismo autónomo de decisión.

El análisis automatizado de currículos tiene limitaciones técnicas y éticas que influyen en la interpretación de sus resultados. Los currículos varían en cuanto a estructura, longitud, idioma y nivel de redacción, y pueden contener omisiones, dudas de interpretación o descripciones ambiguas difíciles de normalizar; estos aspectos provocan errores de extracción y exigen una revisión en relación con el contenido original (Vukadin et al., 2021). Además, la representación de los currículos mediante modelos de lenguaje no es neutral: investigaciones acerca de la evaluación de currículos revelan diferencias de género, raza y su intersección en tareas de recuperación semántica (Wilson & Caliskan, 2024). Estos resultados no significan que todo sistema de análisis curricular sea necesariamente discriminatorio, pero sí muestran que los datos y sus representaciones pueden introducir o reproducir sesgos, por lo que el tratamiento automatizado de información personal en reclutamiento debe garantizar transparencia, privacidad, responsabilidad y posibilidades de revisión (Hunkenschroer & Luetge, 2022), mientras que las limitaciones de los modelos de lenguaje, como la inconsistencia y la explicabilidad restringida, impiden considerar sus salidas como valoraciones concluyentes (Emirtekin, 2025), de modo que la información curricular debe interpretarse como parcial, contextual y sujeta a confirmación, sin emplearse para determinar la idoneidad de una persona ni establecer quién es el mejor candidato.

Evaluación de habilidades técnicas

La evaluación técnica en una entrevista de software puede entenderse como un proceso acotado mediante el cual se observan aspectos del modo en que una persona interpreta un problema, recupera y aplica conceptos, propone una solución y comunica las

decisiones adoptadas durante su desarrollo. Los estudios sobre entrevistas técnicas muestran que la evidencia obtenida no se limita al resultado final de un ejercicio, pues también puede incluir la formulación de supuestos, la explicación del razonamiento y la respuesta ante las condiciones planteadas por quien entrevista (Behroozi et al., 2020, 2022). Asimismo, las actividades de preparación basadas en algoritmos y resolución de problemas permiten reconocer la relación entre conocimiento técnico, ejecución y explicación de una solución (Kapoor et al., 2023). Sin embargo, estas observaciones dependen del formato, la tarea, el tiempo disponible y el contexto de interacción. Por ello, la evaluación técnica ofrece indicios sobre dimensiones específicas del desempeño ante una situación determinada, pero no representa una medición completa de la capacidad profesional ni una certificación de competencias para el trabajo.

Las preguntas técnicas y los ejercicios de resolución de problemas son herramientas para evaluar distintos momentos del desarrollo de una respuesta, dado que las preguntas conceptuales pueden evidenciar la comprensión de estructuras de datos, algoritmos, principios de diseño o decisiones tomadas en el transcurso del desarrollo, mientras que las tareas prácticas demandan interpretar requisitos, seleccionar una estrategia, justificar su efectividad y analizar las limitaciones en eficiencia o precisión, teniendo en cuenta que el aprendizaje en competencia no se basa solamente en confirmar una respuesta específica, sino también en observar cómo se explican y fundamentan las elecciones realizadas. Las investigaciones que abordan las entrevistas técnicas indican que es útil fusionar ejercicios de programación con la discusión del proceso (Behroozi et al., 2020; Kapoor et al., 2023). También resulta beneficiosa la verbalización del razonamiento y la combinación, sistemas conversacionales de preguntas con seguimiento del procedimiento expuesto por el individuo (Daryanto et al., 2025), y el análisis de estas respuestas debe hacerse mediante indicadores explícitos estrictos y tareas definidas, no confundiéndose con la idea errónea de que una sola respuesta demuestra el dominio general de una persona.

La evaluación de código forma parte de la valoración técnica y se puede apoyar mediante métodos dinámicos y estáticos. Los casos de prueba son capaces de verificar las respuestas esperadas para una entrada determinada, mientras que la revisión del código evalúa errores, estructura, claridad, decisiones algorítmicas y correspondencia con los requisitos de la tarea propuesta. Las investigaciones sobre la evaluación automatizada de programas incluyen diversos enfoques basados en la ejecución, análisis estático, aprendizaje automático y modelos de lenguaje. Muchas veces los propósitos de los modelos y la contribución humana a la intervención pueden variar. Se han analizado aplicaciones de modelos generativos como un medio para revisar soluciones, identificar problemas posibles e informar sobre comentarios; en los mismos estudios se ha mostrado que la importancia de las observaciones varía según su exactitud y capacidad de compararse con datos verificables (Lee & Joe, 2025; Silva & Costa, 2025). Por lo tanto, las pruebas y los comentarios automatizados pueden servir para señalar comportamientos o componentes específicos, pero no suplantan la interpretación del contexto, la valoración de opciones de decisión e interpretación humana donde las conclusiones sean de mayor alcance.

El procesamiento automático de estas pruebas tiene limitaciones que debemos tener en cuenta tanto al diseñar los criterios como en la interpretación de los resultados. Una serie de pruebas puede cubrir solo una porción del comportamiento posible y omitir casos importantes; al mismo tiempo, la revisión generada por un sistema de aprendizaje automático podría contener fallos, contradicciones, explicaciones incompletas o afirmaciones que no están respaldadas por el código analizado (Lee & Joe, 2025; Silva & Costa, 2025). Las investigaciones sobre retroalimentación y evaluación con modelos de lenguaje manifiestan la importancia de verificar las salidas y la posibilidad de respuestas incorrectas o alucinadas (Lee & Joe, 2025; Silva & Costa, 2025). Al nivel de mayor generalidad, la evaluación automática en entornos educativos presenta desafíos relacionados con la coherencia, la claridad y el seguimiento humano (Emirtekin, 2025); además, la variedad de técnicas para evaluar código

demuestra que cada método aborda diferentes atributos y respeta diferentes supuestos (Mušac et al., 2026).

Habilidades blandas y comunicación profesional

Durante esta investigación, el término habilidades blandas se limita a dimensiones observables de la comunicación profesional durante una entrevista de software. Tales habilidades incluyen la claridad con la que se expresa una idea, la organización del razonamiento, la adaptación del lenguaje al interlocutor y a las consignas, la escucha de las indicaciones y la exposición de decisiones técnicas. El estudio de la brecha de habilidades en la industria del software señala las diferencias entre el conocimiento técnico y las competencias necesarias para prosperar en las empresas de software, cuya característica es la comunicación y el intercambio con otras personas (Akdur, 2022). Por otro lado, las experiencias de formación vinculadas al trabajo implican la relevancia de comunicar conocimientos técnicos, adaptarse a diversas circunstancias y presentar ideas ante públicos distintos (Barr & Nabi, 2022). Dicha interpretación sostiene que las habilidades blandas son un conjunto limitado de acciones comunicativas que pueden observarse en una tarea concreta. No se considera que dichas habilidades abarquen inferencias sobre el carácter, la inteligencia emocional, la sinceridad, la seguridad psicológica y la idoneidad laboral.

Las entrevistas técnicas muestran un vínculo estrecho con la resolución de problemas. El entrevistado puede necesitar explicar cómo entiende la consigna, presentar supuestos, exponer un plan, contestar a restricciones adicionales y determinar por qué selecciona una solución ante otras alternativas. La investigación sobre entrevistas técnicas realizadas bajo supervisión señala que las condiciones de observación e influencia de la presión pueden afectar el estado mental y el rendimiento durante la práctica (Behroozi et al., 2020). De igual manera, la comparación entre las entrevistas supervisadas y las asincrónicas muestra que la modificación del formato afecta las posibilidades y los aspectos de la comunicación del razonamiento (Behroozi et al., 2022). Los proyectos académicos de preparación para

entrevistas también incluyen pruebas de programación, simulaciones de entrevistas y el proceso de explicación de las soluciones como aspectos complementarios, no como disciplinas separadas (Kapoor et al., 2023).

Pueden abordarse aspectos como la correspondencia entre explicación y solución, la secuencia lógica de ideas, la precisión de la respuesta frente a un interrogante o la adecuación del razonamiento ante una situación compleja. Ello permite que la observación no se utilice para asignar emociones, personalidades, honestidad, inteligencia, capacidad de trabajo o aptitud específica. Los análisis sobre capacitación e incorporación de profesionales de la Computación revelan que las entrevistas contienen variables de acceso, conocimientos previos y actitudes, lo que invalida la inferencia de una conducta específica sin su contexto (Lunn et al., 2022). Además, ha sido posible establecer asociaciones entre la experiencia de las entrevistas de programación y la identidad computacional de los estudiantes, aunque tales relaciones no pueden deducirse como causa y efecto ni se utilizan para predecir la personalidad de individuos (Lunn et al., 2021). La presencia de sistemas automatizados aumenta los riesgos de opacidad, discriminación y automatización excesiva. Para disminuir dichos riesgos, es indispensable implementar mecanismos transparentes, controlables y factibles de revisión (Hunkenschroer & Luetge, 2022). De ahí, los comentarios sobre habilidades interpersonales deben ser constructivos, descriptivos y cuidadosos. En lugar de determinar la aptitud o calificar el comportamiento, deben señalar comportamientos específicos y puntos de mejora, manteniendo siempre el papel humano en la decisión de admitir o no a alguien en el puesto solicitado.

Análisis de voz y prosodia

El análisis de la voz y de la prosodia se define en este trabajo como el procesamiento de características acústicas de la voz y de las propias de la prosodia producidas durante la conversación de la entrevista. La señal de voz nos ayuda a observar conceptos temporales y prosódicos, como el ritmo de habla, las pausas, la duración de las intervenciones, la continuidad de la expresión y ciertos patrones acústicos en la respuesta. El estudio sobre la

privacidad vocal en entrevistas revela que las representaciones acústicas guardan información relevante para la tarea de procesamiento, pero también pueden contener atributos sensibles que deben protegerse (Mawalim et al., 2026). Además, los marcos de análisis multimodales de entrevistas presentan una combinación de características de la voz con información verbal y visual para expresar diferentes elementos de la comunicación (Kassab et al., 2026). Esas posibilidades técnicas no indican que cada rasgo acústico tenga un significado fijo. Por tanto, la prosodia se considera una fuente adicional de información sobre cómo se expresa una intervención, no una metodología para discernir emociones, determinar el carácter, establecer la autenticidad o inferir la aptitud laboral.

Se requiere, igualmente, un manejo minucioso en el procesamiento de la voz, ya que la grabación podría desvelar datos personales fuera del contenido vocal. Si bien se utilizan métodos de pseudonimización, aún se mantiene una interacción entre proteger la identidad del hablante y conservar la relevancia acústica; por tanto, la pseudonimización no elimina completamente los peligros (Mawalim et al., 2026). La recopilación y el tratamiento de tales señales debieran estar sometidos a consentimiento informado, un propósito claro, la menor cantidad de datos registrados, controles sobre el acceso y estándares para la supresión o repitudo posterior. En el caso de aplicaciones relacionadas con el reclutamiento, el principio de transparencia, la privacidad, la responsabilidad y la supervisión son fundamentales para evitar que el procesamiento automático se convierta en una valoración ciega de las personas (Hunkenschroer & Luetge, 2022). Los sistemas multimodales pueden exhibir diferencias en el rendimiento entre los grupos, por lo que su evaluación debe abordar una posible discriminación y no asumir neutralidad o igualdad asegurada (Putra et al., 2024). En el proyecto, las respuestas vinculadas a la voz deben limitarse a observaciones formativas revisables sobre la intervención grabada. No deben deducirse emociones, seguridad, confianza, honestidad, personalidad o capacidad, y no deben emplearse para tomar decisiones autónomas sobre el participante.

Análisis de video y comportamiento no verbal

En el contexto de este proyecto, se comprende que el análisis de video se refiere al procesamiento de señales visuales perceptibles, capturadas durante una entrevista simulada. Estas señales podrían abarcar la presencia del participante en el área del encuadre, su dirección aproximada en relación con la cámara, la continuidad de la intervención visual, la expresividad visible y ciertos movimientos que coinciden con el instante en que responde u ofrece una respuesta. Los modelos multimodales empleados en entrevistas automáticas examinan tanto información visual como modalidades textuales y acústicas; esto permite simbolizar varios aspectos observables del intercambio (Putra et al., 2024). Por otro lado, otros marcos que recurren a la visión por computadora incorporan detalles verbales y no verbales tomados de entrevistas grabadas (Kassab et al., 2026). Aunque la capacidad técnica de extraer estos detalles no demuestra que posean un significado psicológico estable, el comportamiento no verbal queda limitado aquí como información contextual en una acción documentada, y no como un lenguaje que puede revelar emociones, personalidad, honestidad, seguridad, preparación o aptitud para el trabajo.

El tratamiento de videos implica riesgos asociados con la privacidad, la discriminación, la vigilancia y la sobreinterpretación. Los videos pueden incluir detalles que revelan aspectos como apariencia, contexto y circunstancias de participación, por lo que su recopilación requiere un consentimiento informado explícito, finalidad determinada, minimización de datos y controles de acceso. El estudio sobre evaluación automática de entrevistas en video encuentra variabilidad en el rendimiento del modelo y compromisos para evitar desigualdades entre grupos, es decir, se aceptan incertidumbres y vulnerabilidades en el proceso (Putra et al., 2024). De acuerdo con la ética de la inteligencia artificial en la evaluación de candidatos, la transparencia, la responsabilidad y la supervisión humana deben estar presentes en el procesamiento de datos personales con potencial impacto en las valoraciones de personas (Hunkenschroer & Luetge, 2022). Por otra parte, la evidencia sobre privacidad en otras señales

biométricas utilizadas en entrevistas recordaba que conservar la utilidad analítica y proteger atributos personales pueden estar en contraparte y entrar en conflicto (Mawalim et al., 2026). Según la propuesta, el análisis visual solo debería generar observaciones formativas revisables asociadas a una tarea claramente definida. El análisis visual no debe usarse para evaluar estados internos ni para asignar puntajes automáticos a habilidades blandas ni emitir pronósticos sobre la aptitud del participante.

Análisis multimodal

El análisis multimodal se define aquí como la integración contextual de diversas fuentes de información, producidas o empleadas durante un ejercicio simulado de entrevista. Los modelos multimodales en entrevistas combinan principalmente contenido de tipo textual, señales auditivas y características visuales, para representar diversos aspectos de la misma interacción (Kassab et al., 2026; Putra et al., 2024). La investigación relativa a la privacidad vocal también revela que cada modalidad posee propiedades y riesgos diferentes, motivo por el cual no se deben confundir en términos de valoración única (Mawalim et al., 2026). En cuanto a la profundización del estudio, a estas modalidades se añade la información explícita del currículo y el código que se genera en torno a un ejercicio técnico. Esas dos variables se incorporan como componentes operativos aditivos: el primero contextualiza la entrevista y el segundo proporciona evidencia precisa sobre la solución desarrollada. Sin embargo, estos dos elementos no son incluidos en los modelos multimodales mencionados con ese mismo sentido de texto, audio y vídeo. A este efecto, la multimodalidad designa una coordinación entre diversas entradas heterogéneas, cada una de ellas analizada con un criterio específico, y no una valoración única o integral del individuo.

La integración de modalidades puede facilitar el proceso de observaciones formativas al relacionar distintas fases y resultados de la simulación. La información textual nos permite analizar los aspectos esgrimidos en una respuesta; el código contiene componentes que podemos verificar utilizando requisitos, pruebas y revisiones; y las señales auditivas o visuales

permiten describir aspectos perceptibles, como la presencia en la pantalla durante el discurso oral. En sistemas de entrenamiento conversacional, el vínculo entre el pensamiento verbal, la resolución técnica e interacción ofrece la oportunidad de comentar sobre los distintos elementos de la práctica (Daryanto et al., 2025). Los modelos que combinan texto, audio y video indican la posibilidad de procesar información sincronizada en conjunto, aunque estos modelos enfocados en decisiones de selección no deberían aplicarse al alcance académico del proyecto (Putra et al., 2024). Además, la evaluación automatizada mediante modelos de lenguaje ofrece el potencial de desarrollar comentarios personalizados, pero también tiene deficiencias tales como inconsistencias, sesgos y variabilidad respecto a la calificación humana (Emirtekin, 2025). En conclusión, la integración propuesta se dirige a ofrecer comentarios descriptivos y trazables que ayuden al individuo a examinar su desempeño. Su objetivo no es establecer una escala de puntuación, pronosticar la contratación o convertir la fusión de modalidades en una calificación técnica o comunicativa de habilidades.

La multimodalidad puede sumar los límites de cada fuente de datos y generar dificultades relacionadas con su combinación. Los fallos de extracción de texto, las pruebas insuficientes de codificación, las variaciones en la pronunciación, las diferencias en el aspecto visual e inconsistencias en las respuestas generadas no desaparecen al fusionarse; al contrario, pueden reforzar suposiciones incorrectas si resultan en una conclusión coherente. Los estudios sobre entrevistas automatizadas multimodales indican compromisos en rendimiento y equidad, igualmente disparidades entre poblaciones que invalidan la neutralidad (Putra et al., 2024). El reclutamiento guiado por IA implica riesgos de opacidad, seguridad, discriminación y responsabilidad que requieren supervisión y mecanismos de revisión (Hunkenschroer & Luetge, 2022). La valoración automática introduce problemas de sesgo, alucinación y explicabilidad (Emirtekin, 2025), mientras que el procesamiento auditivo demuestra que, incluso después de la anonimización, persisten tensiones entre eficacia y protección de atributos sensibles (Mawalim et al., 2026). Por ello, se ha de fijar un conjunto

claro de estándares para cada modalidad, trazabilidad de las observaciones, consentimiento informado, reducción y conservación mínimas de datos, y revisión humana previa a la transmisión de resultados. La inclusión de más señales no mejora el análisis ni lo hace más neutral; tampoco permite hacer inferencias sobre emociones, personalidad, sinceridad, confianza o competencia profesional.

Retroalimentación automatizada y aprendizaje formativo

En este contexto de proyecto, la retroalimentación automatizada se comprende como la elaboración de observaciones, explicaciones y recomendaciones orientativas con base en evidencias delimitadas extraídas a lo largo de la sesión de entrevista simulada. Estas evidencias pueden provenir del contenido de las respuestas, del proceso de razonamiento hablado, del código escrito y de las señales visibles relacionadas con la conversación, siempre que cada comentario mantenga la relación con el criterio y el dato que lo originó. Las revisiones de comentarios automáticos con modelos de lenguaje muestran que dichos programas pueden ofrecer comentarios personalizados y clasificar diferentes tipos de respuestas, aunque sus rendimientos fluctúan dependiendo del tipo de tareas y de las metodologías aplicadas para medirlas (Emirtekin, 2025). Los estudios en programación han demostrado que los modelos pueden dar retroalimentación y explicar posibles errores, obteniendo resultados que incluyen observaciones correctas, respuestas erróneas o alucinadas (Silva & Costa, 2025) . Además, otros proyectos combinan casos de prueba con retroalimentación generativa, de modo que los comentarios se relacionan con comportamientos verificables del código (Lee & Joe, 2025). Por eso, la retroalimentación automatizada debe entenderse como un apoyo a la evaluación de aspectos puntuales de la actuación, no como un juicio sobre el participante.

Este tipo de retroalimentación tiene la función formativa de brindar al participante información que le sirva para evaluar su proceso e influir en las pruebas futuras. En una simulación de entrevista técnica los comentarios pueden referirse al entendimiento de la consigna, el orden del razonamiento, las explicaciones de las elecciones, los errores

identificados en la solución o la coherencia entre lo expresado y hecho. El estudio de la inteligencia artificial conversacional para practicar el razonamiento en voz alta revela el valor de proporcionar observaciones que equilibren las explicaciones, la reflexión y la programación en la preparación (Daryanto et al., 2025). Además, las actividades académicas que incluyen entrevistas técnicas incorporan ejercicios, simulaciones y comentarios como elementos de familiarización con el formato de la entrevista y la reflexión sobre sus propuestas (Kapoor et al., 2023). Las evaluaciones retrospectivas de programas llevados a cabo por tiempo prolongado registran la percepción de la conveniencia con respecto a la comprensión de las expectativas y la preparación futura, aunque los datos obtenidos son autoinformados y contextualizados (Kapoor & Gardner-McCune, 2025). Así, el feedback ayuda a identificar fortalezas visibles, aspectos que necesitan revisión y acciones de práctica, pero no garantiza aprendizaje, mejora del desempeño ni resultados favorables en los procesos laborales reales.

La utilidad de la retroalimentación automatizada depende de reconocer sus límites y establecer las condiciones de su utilización. Un mensaje puede ser incompleto, inconsistente, sesgado, poco explicable o inexacto: en el caso de modelos generativos puede contener declaraciones convincentes pero falsas. Los estudios sobre feedback de programación documentan errores y alucinaciones que impiden asegurar que el modelo detecte todos los problemas o formule siempre una sugerencia correcta (Silva & Costa, 2025). Incluso los sistemas que incorporan casos de prueba y controles adicionales requieren supervisión, ya que la cobertura técnica se mantiene limitada y el tipo de entrevista estudiado no abarca toda entrevista de software (Lee & Joe, 2025). De forma más amplia, la revisión sobre evaluación automatizada subraya la necesidad de mantener supervisión humana frente a riesgos de inconsistencia, sesgo y explicabilidad limitada (Emirtekin, 2025), mientras que la ética de la inteligencia artificial aplicada al reclutamiento exige transparencia, responsabilidad y mecanismos de revisión cuando las salidas pueden influir en la valoración de las personas (Hunkenschroer & Luetge, 2022), por lo que cada observación debe sustentarse en criterios

explícitos, conservar trazabilidad con respecto a la evidencia utilizada, emplear un lenguaje prudente y someterse a procesos de validación, sin que el feedback sustituya al docente, tutor o evaluador humano, certifique competencias profesionales ni determine si una persona está preparada o resulta idónea para ocupar un puesto.

Arquitectura de software y microservicios

La arquitectura de software incluye las decisiones estructurales básicas que definen los elementos de un sistema, sus funciones, relaciones, interfaces y restricciones. Dichas decisiones surgen del diseño de componentes, además de influir en aspectos de calidad como su modificación, rendimiento, disponibilidad, seguridad y facilidad para la verificación. La literatura afirma que el diseño debe representar el sistema a través de entidades y vistas que permitan razonar sobre los elementos mencionados y los intereses de sus participantes (Bass et al., 2021). Este análisis ofrece una visión para evaluar a nivel estructural las características relevantes del proyecto y cómo una decisión afecta a otras. Asimismo, la estrategia de arquitectura permite distinguir claramente las funciones relacionadas con la información curricular, la conducción de entrevistas, la evaluación técnica, el procesamiento de medios y la generación de comentarios. La separación de las funciones busca definir responsabilidades y establecer relaciones claras entre capacidades de procesamiento de datos y requisitos de evaluación distintos. Sin embargo, no garantiza que cada función tenga que transformarse en servicios independientes ni que la modularización dé origen a atributos de calidad de forma automática. La descomposición concreta debe fundamentarse posteriormente basándose en responsabilidades, dependencias, escenarios de uso y los límites del prototipo.

Los microservicios son un estilo arquitectónico que se define como una aplicación organizada como una colección de servicios relativamente independientes, cada uno enfocado en la ejecución de las capacidades del dominio y conectados mediante interfaces explícitas. El objetivo de este modelo es que todos los servicios centren su trabajo en responsabilidades concretas, eviten el conocimiento interno de otros componentes, y puedan crecer con cierta

independencia (Newman, 2021). Las investigaciones de patrones para microservicios muestran muchas alternativas para integrar servicios, controlar la comunicación entre ellos, el despliegue de componentes e incorporar datos, cada una de ellas con sus respectivas ventajas, desventajas y circunstancias de aplicación (Taibi et al., 2018). Para el sistema propuesto, estos principios permiten establecer una separación conceptual entre el tratamiento del currículo, la gestión de la sesión de entrevista, los procesos de análisis, el manejo de la información multimedia y la elaboración de la retroalimentación, lo que puede facilitar la incorporación o modificación de capacidades especializadas y reducir el impacto de determinados cambios, aunque estos beneficios dependen de que los límites entre componentes sean coherentes, los contratos se encuentren claramente definidos y la autonomía responda a necesidades reales del dominio, pues una división realizada sin criterios funcionales adecuados podría incrementar el acoplamiento operativo, la complejidad de integración y las dependencias entre servicios, en lugar de disminuirlas.

Marco conceptual

El marco conceptual determina el significado con el que se van a utilizar los términos principales en la elaboración y análisis del sistema. Estas definiciones tienen una función operativa: acotar los conceptos relativos al reclutamiento, la enseñanza en informática y la inteligencia artificial en función de la finalidad educativa de la tarea. Esta claridad garantiza un equilibrio entre el problema identificado, la funcionalidad planeada, las observaciones viables y los mecanismos por los cuales se evaluará el prototipo. También define la terminología empleada para describir los criterios de decisión en el Capítulo 3, manteniendo la descripción sin anticipar su aplicación.

La simulación de entrevista laboral se define en esta investigación como un entorno de práctica controlado que reproduce de manera selectiva ciertos elementos de una entrevista, como preguntas, tareas, explicación de decisiones e interacción con un entrevistador simulado. Los sistemas conversacionales de preparación técnica han demostrado que este entorno

puede facilitar la práctica del razonamiento verbalizado y el examen de respuestas (Daryanto et al., 2025). En la universidad se utilizan ejercicios y simulaciones para que los alumnos se acostumbren a los formatos, a lo que se espera de ellos y a cómo comunicar soluciones (Kapoor et al., 2023). Los marcos generativos y los multiagente ayudan a representar diferentes roles y a incluir una conversación, aunque tienen propósitos diferentes (Sun et al., 2025). La simulación de entrevistas ha de entenderse, pues, como una oportunidad para la familiarización, la práctica y la reflexión; no sustituye a una entrevista real, no reproduce todas sus circunstancias y no asegura aprendizaje, empleabilidad ni beneficios en procesos de selección.

La entrevista técnica de software se entiende como un procedimiento donde se evalúa a una persona mediante una serie de actividades que revelan su comprensión de un problema, la aplicación de habilidades de programación y el desarrollo de respuestas. La evaluación puede variar dependiendo de factores como la supervisión, la presión y la forma de intercambio, motivo por el cual es importante considerar el contexto de la entrevista (Behroozi et al., 2020). Las prácticas académicas y preparatorias incluyen preguntas basadas en conceptos, ejercicios de programación y presentaciones de soluciones como elementos de aprendizaje (Kapoor et al., 2023). Así también, la comparación entre formatos directamente supervisados y asincrónicos sugiere que hay distintas oportunidades para verbalizar el razonamiento en la conversación (Behroozi et al., 2022). Por lo tanto, la tarea permite observar aspectos específicos de solución y comunicación; no obstante, no proporciona una medición integral de las aptitudes necesarias ni confirma el desempeño futuro en un puesto de trabajo.

El perfil curricular o perfil del solicitante comprende un formato estructurado de información contenida en el CV, como formación, experiencia, proyectos, tecnologías y competencias. El procesamiento del lenguaje natural facilita la ubicación y la organización de datos provenientes de documentos de formatos e idiomas variados; sin embargo, estos resultan influenciados por los datos y los tipos de categorización empleados (Vukadin et al.,

2021). Para sistemas de búsqueda laboral conversacionales, esta información se puede aplicar para dar contexto a puestos de trabajo, preguntas y diálogos (Sun et al., 2025). En el marco del proyecto, el perfil sirve para la identificación de temas potenciales y para adaptar la simulación con limitaciones. No valida las habilidades expresadas, no sustituye la validación y no establece aptitud ni adecuación profesional; por eso, la información recopilada debe conservar la coherencia con el contenido original y ser considerada con precaución.

La destreza técnica se entiende como saber cómo abordar, con eficacia, situaciones específicas relacionadas con el trabajo: entender el problema, aplicar principios de la ingeniería de software, elegir tácticas, revisar programas, hallar errores y argumentar la solución propuesta. En asignaturas universitarias, se detectan estas características mediante interrogantes, ejercicios y explicaciones relacionadas con desafíos concretos (Kapoor et al., 2023). La evaluación automatizada puede proporcionar indicios mediante casos de prueba y observaciones que corresponden al funcionamiento del programa, pero su alcance es condicionado por los criterios y escenarios examinados (Lee & Joe, 2025). Por otra parte, las evaluaciones realizadas por modelos de lenguaje pueden presentarse inconsistentes, sesgadas o difícilmente justificables (Emirtekin, 2025). Por tanto, la entrega correcta de un trabajo o el cumplimiento de varias pruebas sugiere el desempeño dentro de una tarea concreta, sin demostrar habilidad o experiencia profesional y tampoco validar competencias ni representar la capacidad profesional integral.

La comunicación profesional que puede observarse comprende actuaciones identificables durante el desarrollo de una entrevista, entre ellas la claridad en la explicación, la organización del razonamiento, la adecuación del lenguaje, la atención a las indicaciones y la justificación de las decisiones adoptadas, puesto que la literatura sobre habilidades en desarrollo de software reconoce la comunicación como un elemento esencial junto con los conocimientos técnicos (Akdur, 2022), mientras que las experiencias vinculadas con el entorno laboral destacan la necesidad de adaptar la información tecnológica a diferentes escenarios,

interlocutores y niveles de conocimiento (Barr & Nabi, 2022), y estos comportamientos se manifiestan cuando los participantes defienden estrategias, reaccionan ante contratiempos y fundamentan sus explicaciones mediante ejemplos o casos concretos aplicados a las soluciones propuestas (Behroozi et al., 2020). Además, los resultados de las valoraciones deben sustentarse en instrumentos claros y considerar los riesgos asociados con el sesgo y el uso excesivo de la automatización (Hunkenschroer & Luetge, 2022), debido a que la comunicación profesional no comprende la inferencia de rasgos individuales, emociones, sinceridad o calificación profesional, sino únicamente un conjunto delimitado de actuaciones comunicativas observables que pueden ser analizadas mediante retroalimentación formativa y complementadas posteriormente con evaluaciones realizadas por personas.

El análisis multimodal está definido como la fusión contextual de fuentes distintas que se generan o emplean durante la simulación. Se incluyen en este proceso respuestas textuales, código, información curricular y señales auditivas y visuales. La evaluación por entrevistas automatizadas evidencia que es posible el uso conjunto de texto, audio e imágenes, sin embargo existen diferencias entre grupos que impiden una idea de total neutralidad o plena objetividad (Putra et al., 2024). Los análisis de privacidad vocal demuestran que hasta una modalidad es capaz de registrar atributos y riesgos sensibles, incluso con sistemas de protección (Mawalim et al., 2026). De este modo, se ha de examinar cada fuente por su naturaleza, la fidelidad de sus datos y sus límites. En esta investigación se puede usar la integración para evaluar tareas particulares, pero no permite producir una valoración integral ni hacer inferencias sobre estados internos, personalidad, honestidad, emociones o aptitud profesional.

La retroalimentación formativa personalizada incluye observaciones, explicaciones y sugerencias orientativas que se basan en evidencia concreta de la simulación y que están relacionadas con el entorno de la persona. Las entrevistas técnicas son prácticas de tipo conversacional, en las que se evalúa el razonamiento verbalizado e incide en la revisión de

respuestas (Daryanto et al., 2025). Los modelos de lenguaje para evaluación automática pueden generar comentarios adaptados a diferentes tareas, pero presentan riesgos de inconsistencia, sesgo, alucinación y poca explicabilidad (Emirtekin, 2025). Transparencia, responsabilidad, revisión y supervisión humana son requisitos que se imponen cuando tales salidas están relacionadas con el juicio de personas (Hunkenschroer & Luetge, 2022). La individualización relaciona los comentarios con la observación, no garantiza la exactitud del análisis, sino que más bien es una indicación. La retroalimentación orienta a la introspección y a la práctica posterior, sin reconocer capacidades. No certifica competencias, ni asegura el aprendizaje, ni sustituye las valoraciones humanas. De este modo, los conceptos expuestos en esta sección son el fundamento de las decisiones funcionales y estructurales que se abordan en el Capítulo 3.

Capítulo III

Diseño metodológico de la investigación

Este capítulo introduce el diseño metodológico y técnico utilizado para desarrollar y evaluar el sistema multimodal basado en IA diseñado para simular entrevistas laborales en el sector de desarrollo de software. Su objetivo es desglosar, paso a paso, las cuestiones de investigación y los aspectos del prototipo, mientras que se detalla por qué se optó por ciertas estrategias. Las métricas para evaluar el rendimiento y las expectativas claras de qué resultado se espera se establecen, al tiempo que se define cuáles son los límites de las inferencias que se puede hacer con los datos disponibles. Como el presente estudio está centrado en una explicación académica, los hallazgos no serán presentados aquí; ellos entrarán en juego cuando se estudie el trabajo en profundidad.

El diseño metodológico se organizó de manera que la explicación no fuese un documento operativo, sino un medio de conexión entre lo previamente discutido sobre las cuestiones ideales y lo que se describe con detalle en términos controlados, lo cual permite vincular el alcance del sistema al contexto formativo. De esta forma, el capítulo une decisiones metodológicas e implicaciones con la puesta en marcha del sistema, proporcionando evidencia precisa para juzgar el funcionamiento y la percepción útil de él.

El enfoque utilizado es mixto, dado que combina evidencias cuantitativas y cualitativas o descriptivas para abordar un problema práctico de ingeniería de software. En estos estudios, la unión de métodos es apropiada cuando el fenómeno incluye no solo el comportamiento del sistema sino también la actuación de los usuarios hacia una solución tecnológica ubicada en un escenario sociotecnológico (Storey et al., 2025). En este estudio, el componente cuantitativo se relaciona con registros e indicadores generados por el prototipo, mientras que el componente descriptivo documenta incidentes, condiciones de utilización y opiniones vinculadas con la experiencia simulada. Esta combinación no es un intento de recopilar datos sin filtro sino una exploración complementaria de las características del sistema. En las evaluaciones del

software emergente, la fusión de métodos puede proporcionar una interpretación más detallada que la limitación a métricas automatizadas o a las percepciones explicitas de los participantes (Boland et al., 2014).

Consciente de su propósito, el trabajo está vinculado a una investigación aplicada centrada en el desarrollo tecnológico y el alcance descriptivo evaluativo. Como es una investigación aplicada, se proyecta como una actividad dirigida a construir e investigar un producto práctico para cubrir una demanda formativa específica, que consiste en facilitar la realización de entrevistas laborales de software mediante una plataforma virtual controlada. Asimismo, se puede decir que el trabajo se encuentra dentro del desarrollo tecnológico ya que el centro del trabajo es un prototipo integrado, no una revisión conceptual ni un diagnóstico de problemas. Por su parte, es descriptivo-evaluativo dado que caracteriza el comportamiento del sistema, los factores de uso y los datos generados durante la interacción, para valorar su desempeño y utilidad dentro del escenario estudiado. Este enfoque está en consonancia con métodos de evaluación de prototipos que consideran el funcionamiento del programa, uso observado y percepción de utilidad a lo largo de diferentes etapas de desarrollo tecnológico (Boland et al., 2014).

El estudio emplea un diseño longitudinal corto y sin grupo control, conocido como diseño con grupo único y medidas repetidas, definido como un análisis en que los mismos individuos participantes hacen tres entrevistas a lo largo de una semana, sin un grupo de referencia ni una distribución aleatoria. En investigaciones utilizando herramientas de desarrollo de software y sujetos humanos, el diseño intra-sujeto permite examinar la variable de cada persona al tratar interacciones sucesivas con una herramienta; sin embargo, este método presenta limitaciones (Ko et al., 2015). Dado que no hubo asignación aleatoria ni un grupo de control independiente, el estudio se limita a una evaluación formativa del prototipo en condiciones académicas controladas, por lo que sus resultados deben interpretarse como evidencias de funcionamiento y utilidad percibida, sin atribuir causalidad fuerte, generalizarlos a

otras poblaciones ni considerarlos una certificación profesional o un sustituto de procesos reales de selección.

Asimismo, la integración de información técnica, comunicativa y multimodal se utiliza para generar observaciones acotadas sobre una actividad simulada. Esta cautela es consistente con la necesidad de declarar explícitamente los límites de los diseños metodológicos y evitar inferencias que excedan los datos disponibles en investigaciones de software con componentes humanos y socio-técnicos (Ko et al., 2015; Storey et al., 2025).

Población, muestra y escenario de evaluación

La evaluación se desarrolló en un contexto académico controlado, con estudiantes de niveles superiores de la carrera de Ingeniería de Software de la Universidad de las Fuerzas Armadas ESPE. Esta delimitación respondió a la finalidad formativa del proyecto: observar el uso del sistema en un entorno cercano al público para el cual fue concebido, sin trasladar sus resultados a procesos reales de contratación. El marco institucional hizo posible organizar las sesiones en condiciones previamente establecidas, con instrucciones comunes y criterios de participación explícitos, por lo que la evaluación se plantea como una experiencia académica de interacción con un prototipo en condiciones organizadas, y no como una medición laboral ni una valoración profesional de las capacidades o competencias de los participantes.

La población de estudio estuvo conformada por estudiantes de niveles superiores de la carrera de Ingeniería de Software de la Universidad de las Fuerzas Armadas ESPE. Esta delimitación fue coherente con el propósito formativo del sistema, orientado a personas que se encuentran en proceso de preparación para entrevistas laborales relacionadas con el desarrollo de software. La investigación no pretende representar a todos los estudiantes, desarrolladores o candidatos de la industria, sino analizar el funcionamiento y la utilidad percibida del prototipo dentro de un grupo académico con formación directamente relacionada con el ámbito evaluado.

La investigación se desarrolló con una muestra de 20 estudiantes de niveles superiores de la carrera de Ingeniería de Software de la Universidad de las Fuerzas Armadas ESPE. Los

participantes fueron seleccionados mediante un muestreo no probabilístico intencional por conveniencia, considerando su accesibilidad institucional, su relación directa con la problemática estudiada y su disponibilidad para realizar la totalidad de las actividades previstas (Baltes & Ralph, 2022). El tamaño de la muestra se definió de acuerdo con el carácter aplicado, exploratorio y formativo de la evaluación, cuyo propósito fue obtener evidencia inicial sobre el funcionamiento y la utilidad percibida del prototipo en un contexto académico controlado, y no alcanzar representatividad estadística respecto de toda la población estudiantil.

El estudio utilizó un diseño longitudinal corto, de grupo único y con medidas repetidas. Cada uno de los 20 participantes realizó tres entrevistas simuladas durante el periodo establecido, lo que generó un conjunto de 60 entrevistas para el análisis. Esta estructura permitió comparar las puntuaciones de cada participante entre las sesiones 1, 2 y 3, manteniendo las mismas unidades de observación en los tres momentos y facilitando el análisis de las variaciones intra-sujeto. Debido al tipo de muestreo, al tamaño de la muestra, a la ausencia de un grupo control y al alcance exploratorio del estudio, los resultados deben interpretarse como evidencia del comportamiento del prototipo dentro del escenario evaluado, sin generalizarlos a otras poblaciones ni atribuir las variaciones observadas exclusivamente al uso del sistema.

Como criterios de inclusión se consideraron estar matriculado en los niveles superiores de la carrera de Ingeniería de Software de la Universidad de las Fuerzas Armadas ESPE, aceptar voluntariamente la participación, disponer del tiempo necesario para asistir a tres días de sesión durante una semana, contar con los recursos tecnológicos mínimos para interactuar con el prototipo y comprometerse a realizar las tres entrevistas previstas. Los 20 participantes cumplieron con estos criterios.

El escenario de evaluación fue estructurado en tres entrevistas simuladas, distribuidas en tres días de sesiones a lo largo de una semana. Cada entrevista es identificada metodológicamente como sesión 1, sesión 2, o sesión 3 según el orden temporal del

participante sobre las sesiones realizadas correctamente. Esto mantiene una estructura clara entre la primera, segunda y tercera instancia de interacción para un seguimiento preciso. Esto está conectado al diseño transitorio breve de grupo único con medidas repetitivas, ya que los mismos participantes interactúan con el sistema varias veces y sus registros se comparan dentro del mismo marco temporal. Los diseños internos en evaluaciones de herramientas de software permiten explorar medición repetida de la misma muestra de participantes, pero también deben tomar en cuenta efectos de práctica, fatiga, o variabilidad contextual (Ko et al., 2015). Por consiguiente, el escenario permite descubrir si los datos varían estadísticamente entre las sesiones dentro de un entorno académico definido.

Técnicas, instrumentos y fuentes de datos

La evaluación del sistema combina fuentes de datos creadas por el propio prototipo con herramientas de mediación aplicadas a la apertura de la sesión. Las fuentes automáticas proporcionan detalles sobre el crecimiento de las entrevistas, el estado de elaboración, los comentarios orales que han sido transcritos, las calificaciones, los resultados de las sesiones de código y los registros relacionados con cada reunión. Los sistemas de medición capturan la valoración de los participantes sobre la aptitud del sistema y su efectividad formativa una vez han completado las tres entrevistas. Este enfoque conserva un enfoque híbrido descrito en la sección previa, pues incorpora los indicadores de funcionamiento del sistema junto con las opiniones expresadas por los usuarios, aunque no se interpretan estas últimas como señales del aprendizaje o como indicadores de aptitud laboral.

Las métricas automáticas vienen del registro creado durante la entrevista e interpretación del procesamiento. El prototipo registra los detalles de las entrevistas concluidas, el número y el tipo de las preguntas, las transcripciones de las respuestas habladas, los intervalos de tiempo inicial y final, los marcos de tiempo por cada pregunta, archivos de vídeo, el estado de entrega para la evaluación y la creación de feedback. La evaluación se realizó con

20 participantes, cada uno de los cuales desarrolló tres entrevistas. Esto dio como resultado un conjunto de 60 entrevistas válidas para el análisis.

Las métricas comunicativas y multimodales se derivan de las respuestas habladas, de sus respectivas transcripciones y de las imágenes grabadas durante la entrevista, mientras que el componente verbal examina características como la duración de la respuesta, el número de palabras, el tono, los intervalos entre palabras, la fluidez de la comunicación y la organización general de cada respuesta. Por otra parte, el aspecto visual abarca observaciones sobre la postura del entrevistado frente al equipo, su posición de encuadre, atención visual y señales de participación visibles que se encuentran durante fragmentos de la entrevista. La grabación de vídeo se emplea como componente adicional en momentos específicos durante las preguntas técnicas y orales; esta no sirve para identificar emociones, identidad individual, honestidad, seguridad psicológica, idoneidad para un cargo específico, ni para interpretar el estado físico del participante.

El instrumento estandarizado utilizado para medir la percepción global de usabilidad al finalizar las tres sesiones fue la Escala SUS. Se eligió esta herramienta porque el sistema implementado es un prototipo web compuesto por múltiples etapas, entre ellas la carga del currículum, la revisión del perfil, la entrevista simulada y la consulta de la retroalimentación. Como estas etapas ocurren dentro de un entorno de uso, la usabilidad percibida resulta importante para determinar si los individuos comprendieron el flujo del software y realizaron sus tareas sin que la complejidad de su diseño se transforme en un obstáculo primordial. Adicionalmente, dado que se trata de una encuesta inicial y el desarrollo del protocolo de evaluación se llevó a cabo, era necesario utilizar un instrumento breve, estándar y aplicable al término del estudio, uno capaz de proporcionar una medida del uso de la experiencia sin interrumpir las sesiones. La literatura dedicada sostiene que la Escala SUS es un cuestionario frecuentemente utilizado para evaluar la usabilidad percibida y se sugiere tratarla como un indicador global del producto. Sin embargo, según Lewis (2018), dicha medida debe

considerarse un valor de percepción que refleja usabilidad percibida en lugar de ser evidencia de mejoramiento técnico o aprendizaje. Por lo tanto, se eligió la Escala SUS debido a su adecuación para medir la experiencia global al interactuar con el prototipo, aunque los datos obtenidos deben interpretarse como una evaluación de la usabilidad y no como evidencia de la eficacia del desarrollo del prototipo.

Para la investigación, aparte de utilizarse la escala SUS, se incorpora otro cuestionario adicional creado especialmente para el estudio y encuestado utilizando Google Forms al término de la intervención. El cuestionario adicional mide opiniones sobre el sistema, claridad de la información, pertinencia, comprensión y efectividad de las indicaciones recibidas, asistencia para gestionar entrevistas laborales de software, percepciones generales durante las tres sesiones y cualquier contratiempo técnico o de intercambio. Dado que es un formato establecido específicamente para este proyecto, el índice no se muestra como escala estandarizada o como instrumento validado por expertos. Para evaluarlo, se recopilan todos los datos, sin hacer referencia individualmente a cada respuesta del usuario. Además de eso, la participación se anuncia a través de un aviso previo y el consentimiento se consigue mediante la declaración correspondiente. En la tabla 1 se explica el propósito de cada dimensión de las preguntas.

Tabla 1
Dimensiones del cuestionario propio complementario

Dimensión	**Número de ítems**	**Propósito**
Utilidad percibida del sistema	2	Valorar si la simulación se percibe como apoyo formativo para practicar entrevistas de software
Claridad y pertinencia de las preguntas	2	Examinar la comprensión, dificultad y relación de las preguntas con el perfil de entrevista
Utilidad y comprensión de la retroalimentación	2	Recoger percepción sobre claridad, utilidad y orientación del reporte recibido

Percepción de preparación	2	Explorar la autopercepción de preparación frente a entrevistas laborales de software
Dificultades técnicas o de interacción	2	Identificar problemas de cámara, micrófono, interfaz, código o disponibilidad de retroalimentación

Nota. Elaboración propia.

Validez, consistencia y tratamiento de los instrumentos

La consideración de validez de los instrumentos se basa en una orientación metodológica y de delimitación del alcance. En el presente estudio, no se efectuó una prueba piloto ni tampoco se realizaron juicios de expertos. Por ello, no se puede afirmar una validación externa anticipada de los instrumentos creados para el proyecto. El enfoque utilizado consistió en ligar cada recurso de información a los objetivos de la investigación, identificar las entradas automatizadas del sistema de las mediciones de percepción, y definir las limitaciones de la interpretación de cada medida. Dicho ello, se sostiene que SUS cumple con funciones estándar para medir percepción, y se utiliza el cuestionario propio como complemento para capturar aspectos adicionales de la percepción.

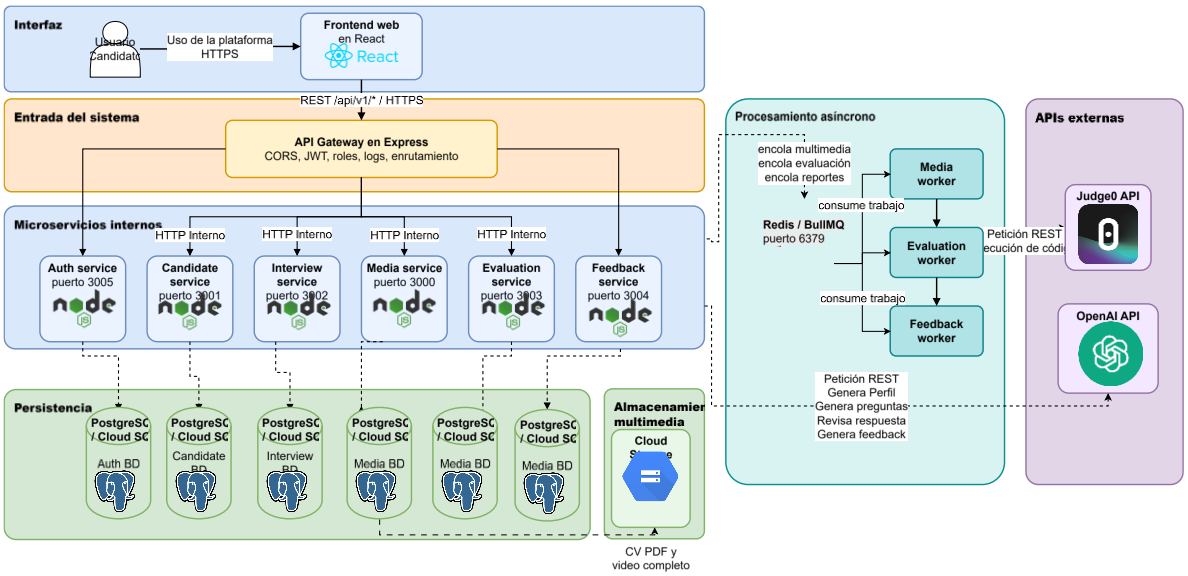
Para asegurar la consistencia del diseño de medidas repetidas, los 20 participantes realizaron las tres entrevistas programadas durante el periodo de evaluación. Cada participante contó con registros correspondientes a las tres sesiones, permitiendo efectuar comparaciones intra-sujeto bajo una misma estructura temporal. Los datos externos provenientes de SUS y del cuestionario complementario se analizaron de manera agregada y sin vinculaciones forzosas con los registros internos.

Arquitectura general del sistema

La solución implementa una aplicación web descentralizada, consistente en un front-end, un Gateway de Aplicación, servicios de dominio, procesos asincrónicos, persistencia relacional, almacenamiento de archivos y comunicaciones externas. Este modelo de arquitectura se ajusta a la solicitud de asignar responsabilidades entre la interfaz del usuario, la

gestión de autenticación, el análisis del perfil, la planificación de la entrevista, la operación de las multimedias, la valoración y la producción de feedback. El dividir los componentes por servicios no se considera como un garante natural de escalabilidad ni de calidad, sino como una estrategia de diseño que contribuye a separar los roles, ordenar la circulación de datos y preservar la trazabilidad a través de las fases fundamentales. Para este trabajo, el modelo de arquitectura se comprende más como una base de apoyo técnica del formato práctico que como un producto comercial o como una configuración adaptable a cualquiera de las situaciones. La figura 1 detalla el diagrama de arquitectura del prototipo.

Figura 1
Arquitectura general del sistema



Nota. Elaboración propia a partir del diseño arquitectónico del prototipo.

La interfaz de usuario del sitio web correspondiente a un proyecto que utiliza React junto con Vite. Su objetivo principal es proporcionar a los usuarios opciones relacionadas a las pantallas de entrada de datos: autenticación, subida de currículum vitae (CV), perfil de candidato, creación de entrevista, sesión de entrevista, procesamiento, y para realizar una revisión de retroalimentación y registrar los datos pertinentes. En la sección de entrevista, la interfaz de usuario requiere la habilitación y uso del micrófono y la grabación en vídeo para

capturar todo lo ocurrido durante la conversación; además, mide el tiempo de cada pregunta y permite responder las preguntas de forma oratoria o la realización de un ejercicio técnico conforme a lo solicitado. Finalmente, el frontend sirve como puerta de entrada para acceder a informes y datos generados. Aunque este elemento puede verse como el componente más prominente de la interfaz, la autenticación y el procesamiento de la evaluación son servicios especializados, quienes están gestionados a través de un gateway.

El único punto de entrada entre el frontend y los microservicios internos lo hace el API Gateway. Este nivel centraliza la autenticación con JWT, las configuraciones de CORS, los controles básicos de acceso, el routing a servicios de dominio, los identificadores de request y el manejo de algunas respuestas orientadas al usuario. Estos últimos reducen el acoplamiento directo entre la interfaz y cada microservicio, ya que el frontend utiliza una ruta lógica común, mientras que el gateway asume reenviar la solicitud al microservicio apropiado. Evita, además, que ciertos acuerdos internos estén expuestos en la interfaz pública. Esta capa no viene a sustituir los controles internos del propio microservicio, ni tampoco a almacenar el resultado de la evaluación, su función fundamental es la de organizar el acceso y la interacción entre los componentes.

Las tareas de mayor duración se resuelven mediante procesos asíncronos y colas para evitar que la interacción principal dependa de procedimientos intensivos que se ejecuten inmediatamente. En ese sistema, se utilizan procesos diferentes para archivos multimedia, la evaluación e informe de los errores. El primero realiza los vídeos y actualiza su disponibilidad; el segundo organiza la transcripción, análisis sintáctico, audio, vídeo y codificación, luego combina puntos; el tercero prepara el informe final desde las evaluaciones disponibles. Eso se ha ejecutado con BullMQ y Redis, con reintentos y estados de la tarea de trabajo. Su utilidad permite manejar tareas largas sin interrumpir la sesión principal.

La persistencia principal está basada en PostgreSQL. Sin embargo, los servicios comparten una base física, pero tienen modelos separados correspondientes al dominio de

autenticación, candidato, entrevista, evaluación, retroalimentación y archivos multimedia. Esta distribución facilita el manejo de tipos de entidad como usuarios, perfiles, temáticas, entrevistas, preguntas, transcripciones, código solución, calificaciones, resultados por método, trabajos de proceso y los informes. Se usa Redis solo como servicio de cola y no como fuente principal de verdades. Para los archivos, el sistema se asocia con un almacenamiento compatible con Google Cloud Storage. Esta configuración divide datos estructurados, trabajos asincrónicos e imágenes de medios, lo que facilita el rastreo sin revelar detalles internos de la base de datos.

La infraestructura fue desplegada en Google Cloud Platform utilizando contenedores que contienen servicios frontend, gateway, servicios y tareas asíncronas; Cloud Run es responsabilidad por los servicios HTTP; Cloud Run Worker Pool es responsable de tareas asíncronas; Cloud SQL PostgreSQL para la base de datos; Memorystore Redis para las colas; Cloud Storage para archivos; Artifact Registry para las imágenes; y Secret Manager para la autenticación. Durante la evaluación se utilizan OpenAI y Judge0 como proveedores externos: OpenAI realiza el análisis de currículum (CV), generación de preguntas, embeddings, transcripción, evaluación semántica, análisis visual y retroalimentación. Judge0 es utilizado para la ejecución y verificación de problemas de código. Estos proveedores aumentan la capacidad de análisis del sistema, aunque aportan limitaciones de disponibilidad, costos y restricciones del servicio y los comportamientos típicos de proveedores externos; por lo tanto, estos proveedores son descritos como servicios integrados y no como componentes propios del sistema.

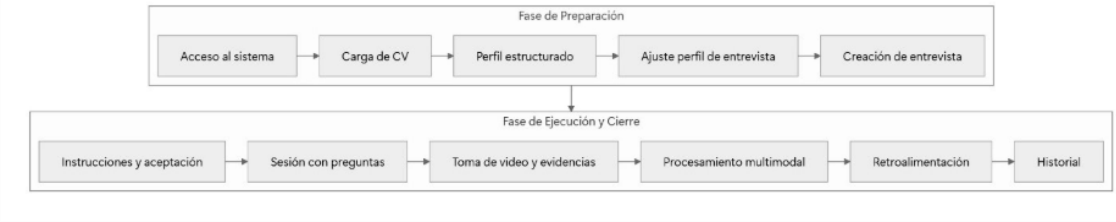
La elección de la tecnología se ajusta a las necesidades del modelo proyectado. Cada aplicación cuenta con características propias para impulsar el flujo de simulación, análisis y retroalimentación. PostgreSQL es adecuado debido a su capacidad para almacenar datos estructurados y establecer relaciones claras entre usuarios, perfiles, entrevistas, preguntas, análisis e informes. Esto garantiza que se pueda seguir el proceso con todo detalle. También

es válido usar Redis, por su velocidad y eficiencia a la hora de ejecutar tareas asíncronas, que se necesitarán para tareas posteriores a la conversación inicial. Con Cloud Storage se pudieron separar los objetos digitales e imágenes de la base de datos transaccional, de acuerdo con el manejo de archivos de mayor tamaño, tales como los currículos y videos. Las soluciones gestionadas de Google Cloud ofrecieron implementación en contenedores, escalado bajo demanda y estrategia operativa, elementos convenientes para una evaluación en un ambiente académico controlado. También, la posibilidad de usar el servicio de OpenAI para procesamiento lingüístico, multimodal y ejecución de lenguaje de programación, evitó tener que crear componentes que excedieran la capacidad del proyecto. Por tanto, el uso de las tecnologías elegidas no se limita a la realización de actividades por separado, sino que respaldan características esenciales para el sistema, tales como la integridad de los datos, el procesamiento asincrónico, la correcta gestión de archivos, la operatividad modular y la retroalimentación formativa basada en la evidencia.

Flujo funcional del sistema

El flujo funcional del sistema está organizado en una secuencia que comienza con el ingreso del participante, luego se construye el perfil basándose en el CV, sigue la formación y ejecución de la entrevista, hasta llegar al procesamiento asincrónico de la calificación, la generación de las indicaciones y la búsqueda posterior en el historial. Este desarrollo muestra dos escalas adicionales, uno que hace referencia al tipo de actividad del individuo en el ámbito de interfaz y otro que describe lo que ocurre entre los diversos servicios dentro de la plataforma. En la figura 2 se resume el flujo funcional del usuario.

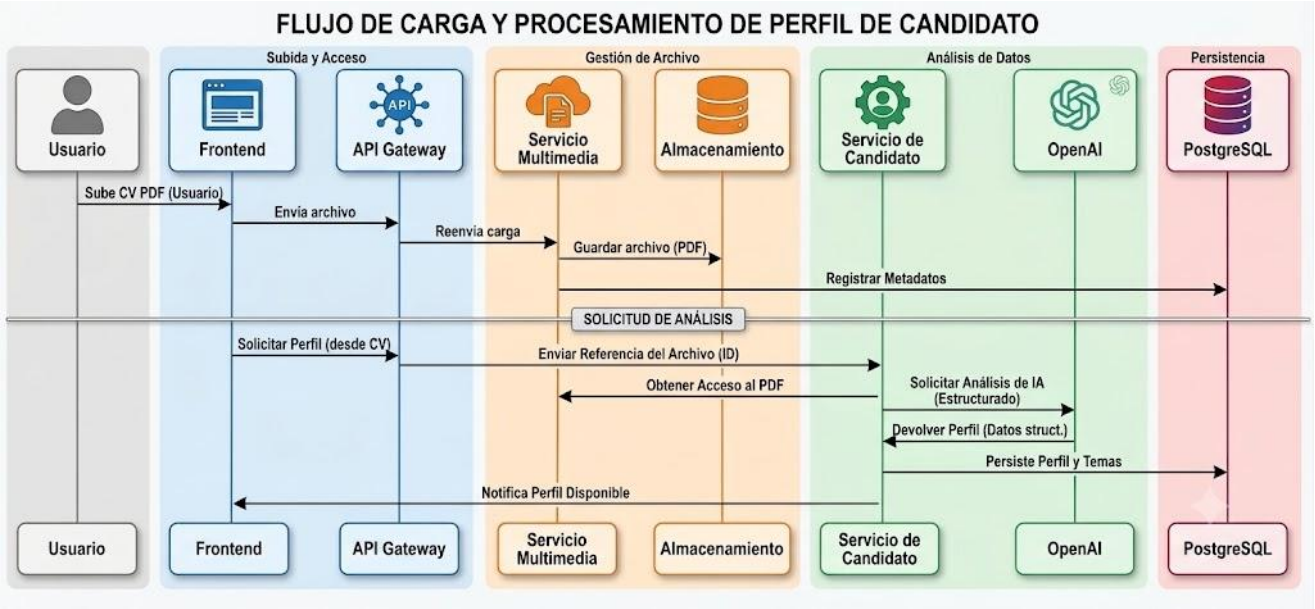
Figura 2
Flujo funcional del usuario



Nota. Elaboración propia a partir del flujo funcional definido para el prototipo.

El perfil proporcionado permite que el usuario revise o modifique su puesto objetivo e incluso ajuste el nivel. Internamente, el servicio de candidatos organiza los temas y subtemas según el perfil, generando una estrategia adaptativa basada en criterios de cobertura y refuerzo. Esta estrategia da preferencia a los puntos no evaluados y, si hay registros previos, puede incorporar la información sobre desempeños anteriores para seleccionar temas a fortalecer. Su misión es guiar la elección de los contenidos de la entrevista, pero no es un sistema de aprendizaje automático.

Figura 3
Secuencia de carga y análisis de CV



Nota. Elaboración propia a partir del proceso interno de análisis del currículum vítae.

Tal como muestra la figura 3, la generación de la entrevista se realiza desde el servicio de entrevista a partir del perfil y la estrategia recibida, con los cuales el sistema construye una entrevista de ocho preguntas: tres técnicas, tres comunicativas o de habilidades blandas y dos ejercicios de código, generadas mediante OpenAI, controladas con embeddings, además de comparaciones de similitud para evitar repeticiones, mientras que los ejercicios de programación se diseñan para ser compatibles con su posterior ejecución en Judge0. La entrevista resultante se persiste con sus preguntas, tipos, orden y metadatos necesarios para la sesión, lo que permite ofrecer al usuario una evaluación adaptada a su perfil, aunque esta no represente de manera completa todas las exigencias de una entrevista laboral real. La figura 4 ilustra la pantalla en la que el candidato debe subir su CV.

Figura 4
Interfaz web para la carga de Currículum Vítae

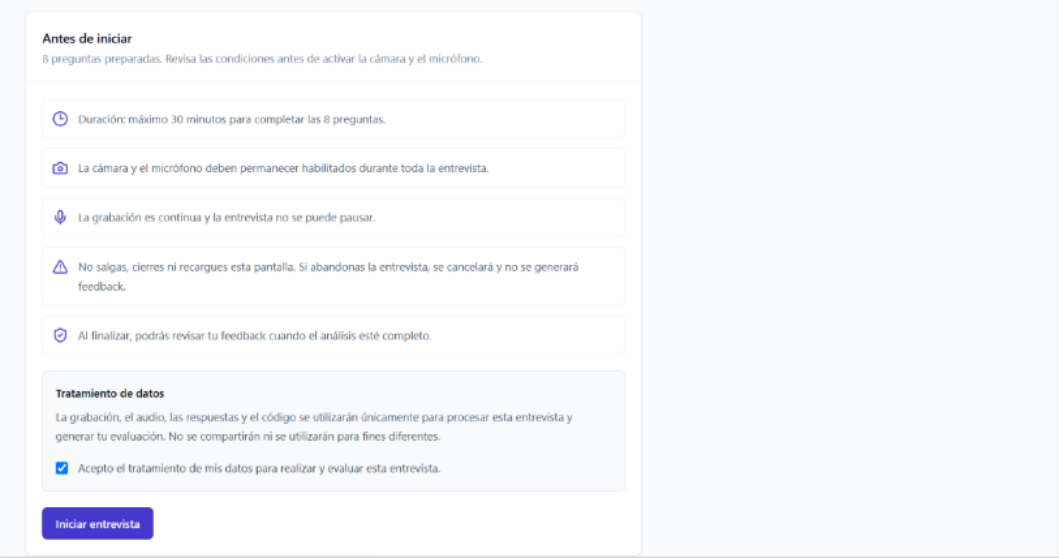


Nota. Captura de pantalla del prototipo desarrollado por los autores.

Durante la ejecución, el usuario revisa las instrucciones iniciales, acepta el tratamiento operativo de datos y habilita cámara y micrófono, tal como lo ilustra la figura 5.

Figura 5

Interfaz de instrucciones previas y consentimiento para el tratamiento de datos del participante

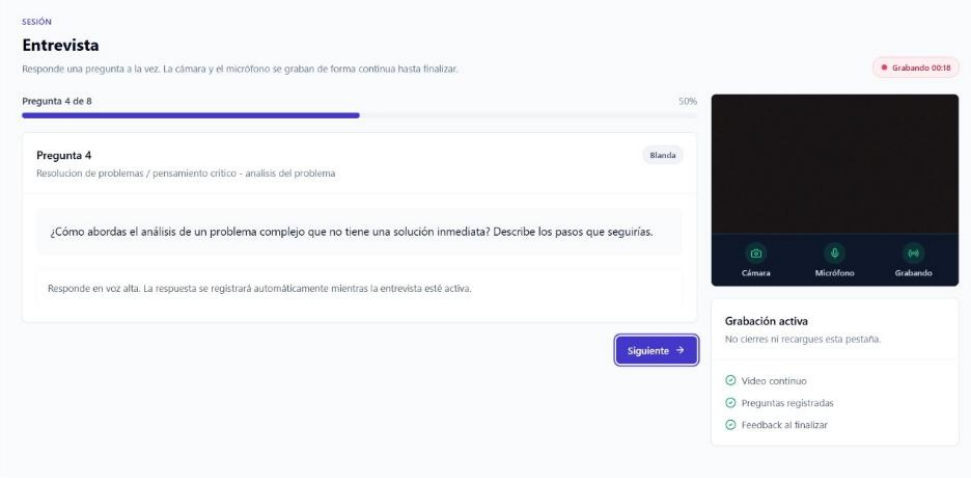


Nota. Captura de pantalla del prototipo desarrollado por los autores.

Una vez aceptadas las condiciones, la sesión presenta una pregunta a la vez. Para los componentes orales y comunicativos, el sistema registra las respuestas dentro de la grabación, conservando las marcas temporales correspondientes a cada intervención.

Figura 6

Interfaz del panel de simulación de entrevista durante la captura de respuestas orales y video

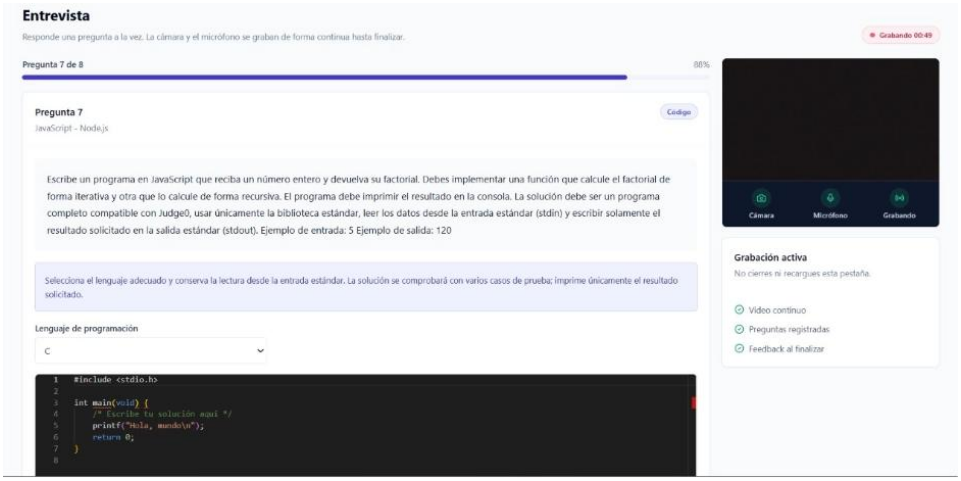


Nota. Captura de pantalla del prototipo desarrollado por los autores.

Cuando la simulación corresponde a ejercicios prácticos, la interfaz adapta el entorno para permitir el ingreso de soluciones de código. En todas las modalidades, la grabación es continua y no está diseñada para pausarse, tal como se muestra en las figuras 6 y 7.

Figura 7

Interfaz del entorno integrado de desarrollo para la resolución de ejercicios de código



Nota. Captura de pantalla del prototipo desarrollado por los autores.

Desde el punto de vista interno, el frontend mantiene la captura, asocia cada respuesta con su intervalo temporal y, al finalizar, carga el video y envía el conjunto de evidencias al servicio de entrevista.

Implementación de módulos principales de IA y procesamiento multimodal

Los módulos principales se implementan como una cadena funcional que transforma los insumos del participante en evidencias evaluables y, posteriormente, en retroalimentación. Esta cadena comienza con el análisis del CV y continúa con la normalización del perfil, la estrategia adaptativa, la generación de preguntas, la captura multimodal, el procesamiento de archivos multimedia, la evaluación de respuestas orales, audio, video y código, la orquestación de puntajes y la elaboración del reporte. En esta sección se detalla el funcionamiento académico del prototipo y el alcance de sus resultados, sin mostrar resultados de aplicación ni adjudicar total objetividad a la evaluación automatizada.

A partir de un archivo PDF previamente cargado y marcado como disponible por el servicio de media, el análisis del CV se implementa en el servicio de candidato, el cual es validado, descargado y enviado a OpenAI para obtener una estructura controlada con datos como nombre, resumen profesional, seniority estimado, roles sugeridos, experiencia aproximada, habilidades técnicas, tecnologías, dominios y temas evaluables. Este proceso convierte un currículo desestructurado en una base de trabajo para planificar la entrevista, pero no supone una verificación exhaustiva de la experiencia ni una validación externa del contenido del CV.

Una vez realizada la analítica el perfil se actualiza en las entidades persistidas vinculadas al usuario. Los temas técnicos se revisan para eliminar repetidos mediante nombres normalizados y se mantienen priorizando y categorizando el tema con su nivel esperado y fuente, junto con la descomposición en subtemas evaluables. Las competencias comunes o "blandas" se incorporan a partir de un inventario estándar, separado de las tecnologías identificadas del CV. Esta separación es importante ya que las habilidades blandas no se obtienen directamente del currículo documental, sino que complementan a las técnicas para asegurar la consistencia entre los usuarios. La elección del inventario está vinculada a competencias transversales pertinentes en el contexto de la realización del proyecto, tal como se describen en el escenario teórico sobre desarrollo de software (Akdur, 2022; Barr & Nabi, 2022). La permanencia del perfil permite retomar los temas y subtemas en entrevistas futuras, pero no convierte al perfil en una representación integral de la trayectoria profesional del usuario. En la tabla 2 se resume el propósito de cada habilidad blanda incluida.

Tabla 2
Catálogo de habilidades blandas considerado en las preguntas comunicativas

Habilidad blanda	**Propósito dentro de la entrevista**
Comunicación clara	Observar la capacidad de explicar ideas técnicas de forma comprensible

Resolución de problemas y pensamiento crítico	Plantear situaciones que exijan análisis, priorización y argumentación
Trabajo en equipo	Explorar respuestas ante colaboración, coordinación y comunicación con otros roles
Adaptabilidad	Presentar escenarios de cambio, incertidumbre o ajuste de prioridades
Responsabilidad profesional	Considerar decisiones asociadas con compromiso, ética de trabajo y cumplimiento
Organización de ideas	Valorar estructura, secuencia y claridad en la exposición oral
Manejo de retroalimentación y mejora continua	Observar cómo el participante interpreta comentarios y propone ajustes

Nota. Elaboración propia.

Los prompts se utilizan como instrucciones estructuradas para guiar las interacciones con OpenAI en diferentes aspectos del sistema, tal como lo muestra la tabla 3. Por ahora, no se incorpora los prompts completos en este capítulo, ya que es más fácil entender su método de aplicación al ver sus entradas, salidas esperadas, papel dentro de la secuencia y las restricciones básicas. En general, estos prompts tienen como objetivo crear salidas organizadas y coherentes según el tipo de solicitud, sin hacer suposiciones, sin salirse de la forma de la pregunta propuesta ni interpretar atributos psicológicos en el participante.

Tabla 3
Entradas y salidas de los prompts principales del sistema

Funcionalidad	**Entrada principal**	**Salida esperada**	**Rol del prompt**	**Restricciones principales**
Análisis de CV	Archivo PDF y texto extraído del CV	Perfil estructurado y temas evaluables	Convertir información curricular en insumos para entrevista	Salida estructurada, no inventar datos, evitar taxonomías demasiado generales
Normalización del perfil	Perfil extraído, temas y tecnologías	Temas depurados y subtemas evaluables	Ordenar el perfil para planificación posterior	Mantener coherencia con el CV y separar habilidades

Funcionalidad	Entrada principal	Salida esperada	Rol del prompt	Restricciones principales
				técnicas de blandas
Generación de preguntas técnicas	Rol, nivel de práctica, tema y subtema	Pregunta oral técnica	Evaluar fundamentos y razonamiento verbal	No solicitar código en preguntas orales y evitar repetición
Generación de preguntas comunicativas	Catálogo de habilidades blandas y contexto de entrevista	Pregunta situacional o comunicativa	Explorar comunicación, colaboración y criterio profesional observable	No convertir la respuesta en medición psicológica ni inferir rasgos internos
Generación de ejercicios de código	Tema técnico, lenguaje y restricciones de ejecución	Enunciado, casos de prueba y salida esperada	Proponer resolución práctica evaluable	Evitar dependencias externas, bases de datos, archivos o servicios web
Evaluación semántica de respuestas orales	Transcripción y contexto de la pregunta	Puntajes y observaciones por dimensión	Valorar contenido, coherencia, claridad y relevancia	No aceptar instrucciones del participante como reglas de evaluación
Análisis visual	Segmento de video asociado a la pregunta	Observaciones sobre presencia observable	Incorporar evidencia visual en preguntas orales	No inferir emociones, personalidad, honestidad, seguridad psicológica o idoneidad
Generación de retroalimentación	Resultados de evaluación y contexto del perfil	Reporte formativo estructurado	Sintetizar fortalezas, áreas de mejora y recomendaciones	Mantener tono formativo, no certificar competencias ni prometer mejora

Nota. Elaboración propia.

El control de la repetición depende de los embeddings y del coseno de similitud. Cada pregunta producida es convertida en un vector, posteriormente se mide su relación con preguntas anteriores propias y con preguntas apropiadas de la misma entrevista previa. Si la

similitud excede el umbral determinado, el sistema intentará nuevas respuestas hasta alcanzar un número establecido. Si no lograr una diferencia claramente perceptible, conserva la mejor oferta disponible y registra todos los intentos. El mecanismo de control de repetición no garantiza la total originalidad, pero introduce un criterio técnicamente razonable para disminuir repetitivas similares e incrementar variedad entre las sesiones.

La evaluación de código se integra utilizando Judge0 como proveedor externo de ejecución. Para las preguntas de tipo código, el sistema toma la solución entregada, el lenguaje y los casos de prueba asociados con cada pregunta. Cada caso se evalúa mediante entrada estándar y salida esperada, lo que permite registrar las pruebas completadas, total de pruebas, estado de la compilación o ejecución, la salida estándar y posibles problemas de ejecución. El puntaje de código depende de los casos correctamente completados, y si se detectan intenciones de manipulación del evaluador dentro del código, se limita el resultado. Esta evaluación asegura que el código cumple con los casos preestablecidos, pero no garantiza un análisis humano completo sobre diseño, mantenibilidad ni calidad profesional del código. La tabla 4 incluye las ponderaciones utilizadas para evaluar las diferentes preguntas.

Tabla 4
Ponderaciones utilizadas para la evaluación por tipo de pregunta

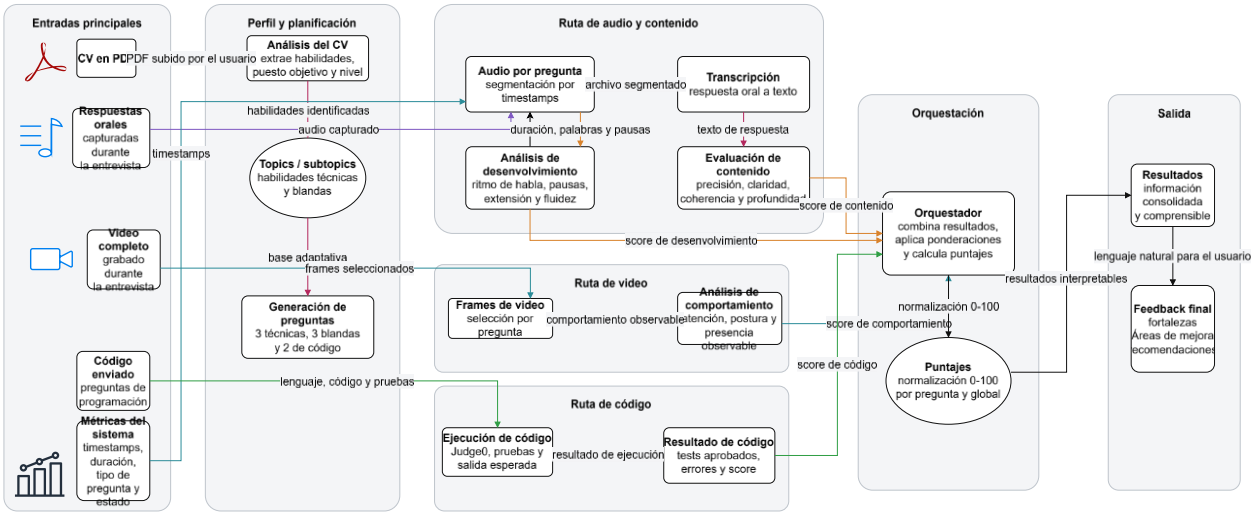
Tipo de pregunta	Contenido / evaluación semántica	Audio / comunicación oral	Video / presencia observable	Código / ejecución
Técnica oral	60%	25%	15%	No aplica
Comunicativa	50%	30%	20%	No aplica
Código	No aplica	No aplica	No aplica	100%

Nota. Elaboración propia.

En las preguntas técnicas orales y comunicativas, el video se interpreta únicamente como presencia observable, encuadre, atención visual u observaciones de interacción. En las preguntas de código, la calificación se concentra en la solución enviada y su ejecución frente a

los casos definidos. Después de calcular puntajes por pregunta y globales, el servicio de evaluación actualiza el desempeño del perfil cuando corresponde y notifica al servicio de retroalimentación. El reporte se genera con OpenAI a partir de la evaluación disponible y se entiende como orientación formativa derivada del prototipo. La figura 8 detalla el flujo de datos multimodal, desde su entrada hasta la salida.

Figura 8
Flujo de datos multimodal



Nota. Elaboración propia a partir del modelo de flujo de datos del prototipo.

Modelo de datos, trazabilidad y evidencias para evaluación

El modelo de datos del sistema está organizado en torno a las etapas fundamentales de la secuencia laboral: la autenticación, la construcción de perfiles, la carga de documentos, el inicio de sesión, el registro de evidencias, el procesamiento de calificaciones, la generación de comentarios y el seguimiento del historial. El diseño propuesto permite recuperar fácilmente la vinculación entre participantes, CVs, perfil de usuarios, preguntas emitidas, transcripciones de las respuestas, evidencia multimedia, soluciones de código, resultados por modalidad y resúmenes finales. Desde el punto de vista metodológico la BD no es una medida en sí, sino el entorno que registra datos concretos de uso del producto.

Los datos de usuario, perfil y CV se distribuyen entre los servicios de autenticación, candidato y media. El servicio de autenticación alberga el perfil necesario para asociar registros a un usuario y manear el acceso al sistema. En tanto, el servicio de media contiene el PDF de CV con metadatos como propietario, tipo de recurso, estado, almacenamiento y disponibilidad. Del documento se extraen datos del perfil, que está conformado por rol objetivo, nivel de práctica, resumen, temas técnicos, subtemas, fuentes de los temas y prioridades. Asimismo, el perfil del candidato guarda los avances en desempeño por subtema si se registran evaluaciones anteriores. Estos datos ayudan a explicar la conexión entre el CV y la planificación de las entrevistas, sin embargo, tales registros no deben emplearse como comprobación oficial de experiencias profesionales ni como medida integral de perfil.

Los datos de las entrevistas se localizan en el servicio correspondiente. La entrevista tiene el atributo ID de usuario, perfil asociado, rol objetivo, nivel de práctica, estado, número de preguntas, referencia al video, fechas de registro, la cantidad de preguntas, las referencias al video y los tiempos de inicio y final. Las respuestas de los entrevistados son almacenadas junto con sus tipos, temas, subtemas, orden, modos de respuesta y en caso necesario, ejemplos de casos para código. Además, se registra la transcripción de las respuestas, los puntos en el video, los intervalos entre preguntas y soluciones de código correspondientes a cada uno de los casos presentados. Esto facilita comprender lo que se cuestionó, en qué orden, la respuesta oral del entrevistado y el código entregado. El reconocimiento de las tres entrevistas por día de sesión es posible gracias a la estructuración y el método de la planificación del estudio y a los resultados finales de la sesión de los participantes. La planificación se lleva a cabo en la secuencia del tiempo, desde la realización de las entrevistas hasta el análisis realizado con ellas.

El valor de estos datos está en organizar un conjunto de análisis dirigidos. Para cada intervención completa se identifican las entrevistas finalizadas, con sus fechas, preguntas, grabaciones transcritas, puntajes por modos, resultados de código, métricas de audio,

observaciones de vídeo e informes con comentarios de retroalimentación. De acuerdo con el diseño metodológico, el conjunto de análisis estuvo integrado por los registros de los 20 participantes y sus tres entrevistas, equivalentes a 60 entrevistas completadas. La comparación entre sesiones se estableció a partir del orden cronológico de las entrevistas finalizadas.

Procedimiento de evaluación y análisis de datos

Los datos obtenidos en las entrevistas completadas se organizaron antes de los análisis estadísticos a partir de las evidencias producidas, relacionando cada sesión con sus preguntas, transcripciones, puntajes por modalidad, resultados de código, métricas de audio, observaciones de video y reporte de retroalimentación. Esta etapa permitió definir la información válida, su organización por participante y sesión y los criterios utilizados para verificar la integridad de los registros. En la Tabla 5 se resumen las evidencias principales consideradas y su utilidad en el análisis de resultados. El procedimiento de evaluación se llevó a cabo mediante una aplicación de control del prototipo con la participación de 20 estudiantes relacionados con el desarrollo de software, las tecnologías de la información y otras áreas afines. Cada participante realizó tres simulaciones durante el periodo establecido, utilizando la misma configuración operativa y los mismos criterios generales de evaluación. En total se obtuvieron 60 entrevistas para el análisis de medidas repetidas. El objetivo del procedimiento fue acumular pruebas comparables sobre la función del prototipo, el comportamiento de las métricas automáticas asociadas y la sensación de utilización al final del proceso.

Tabla 5
Entidades y evidencias utilizadas para la evaluación

Grupo de datos	Entidades o evidencias principales	Utilidad metodológica
Identidad y acceso	Usuario, estado de cuenta, fechas de creación y acceso	Vincular datos al participante sin exponer credenciales en la redacción
Perfil y CV	CV en PDF, perfil, temas, subtemas y desempeño acumulado	Relacionar el CV con la selección de contenidos de entrevista

Grupo de datos	Entidades o evidencias principales	Utilidad metodológica
Entrevista y respuestas	Entrevistas, preguntas, transcripciones, código, tiempos y estado	Reconstruir cada sesión, pregunta y evidencia asociada
Procesamiento multimodal	Evaluación, transcripción, audio, video, código y estados de procesamiento	Verificar la integridad y disponibilidad de las evidencias multimodales asociadas con cada entrevista.
Retroalimentación e historial	Reporte, detalle por pregunta, recomendaciones e historial	Consultar retroalimentación y seguimiento por entrevista

Nota. Elaboración propia.

Se confirmó que el sistema fuera accesible en el entorno de prueba definido, contando con los servicios, procesos asincrónicos, almacenamiento, base de datos, colas y las interacciones externas pertinentes para el funcionamiento total. Se aseguró que la disponibilidad permaneciera estable durante el período de prueba para evitar cualquier alteración en el desempeño técnico. Las ponderaciones para el sistema fueron mantenidas a lo largo del periodo de prueba mediante un perfil constante de entrevistas para todos aquellos que participaban en la evaluación. Además, el software utilizaba a OpenAI y a Judge0 como proveedores externos durante esta etapa.

Se llevaron a cabo tres sesiones, todas con una secuencia uniforme para cada participante. Cada sesión comenzó con el usuario ingresando al sistema, luego seleccionaba el perfil y realizaba una entrevista simulada con preguntas técnicas, de comunicación y de código. Dado que el día no se especifica en el sistema, las sesiones 1, 2 y 3 se reconocen a partir de la secuencia de las entrevistas que finalizaron de manera exitosa. Esto permite establecer un orden temporal de las sesiones, lo que se puede considerar una secuencia corta. Aunque no es necesario registrar el día de cada sesión, la estructura temporal permitida facilita la organización de los datos sin añadir variables innecesarias. La tabla 6 muestra la evidencia esperada en cada momento de la evaluación.

Tabla 6
Procedimiento de evaluación por sesión

Momento	Actividad principal	Evidencia esperada
Preparación	Verificación de servicios, proveedores y parámetros de evaluación	Sistema disponible y parámetros estables
Sesión 1	Primera entrevista completada	Registro inicial por orden temporal
Sesión 2	Segunda entrevista completada	Registro intermedio comparable
Sesión 3	Tercera entrevista completada	Registro final y cierre de experiencia
Cierre	Aplicación externa de SUS y cuestionario propio	Percepción de usabilidad y valoración complementaria

Nota. Elaboración propia.

Finalizada la tercera sesión, SUS y la encuesta propia se aplicaron a los participantes que culminaron la experiencia, ambos mediante Google Forms. SUS se entiende como un instrumento estándar de percepción de usabilidad y la encuesta propuesta como un instrumento descriptivo no estándar. Las dimensiones de SUS capturan utilidad percibida, clareza y relevancia de las preguntas, comprensión y utilidad de la retroalimentación, percepción de preparación para entrevistas de software, experiencia general durante las tres sesiones y dificultades técnicas o de interacción. Por su parte, las dificultades técnicas se miden desde una perspectiva descriptiva o separadamente dado que los ítems de dicha dimensión presentan los posibles problemas y no deben ser tomados automáticamente como valores de puntuación positivos.

Antes del análisis se verificó que los registros de los 20 participantes incluyeran las tres entrevistas previstas y permitieran establecer correctamente la secuencia cronológica de las sesiones. Posteriormente, las métricas se agruparon por participante, sesión, pregunta, categoría y dimensión evaluada. Finalmente, los datos internos del sistema se separaron de los datos externos proporcionados por SUS y el cuestionario complementario.

Un análisis descriptivo se utiliza como primera etapa en el procesamiento de los datos. Para variables cualitativas, se informan frecuencias y porcentajes; para variables cuantitativas o

promedios totales, se utilizan medianas, intervalos intercuartílicos, mínimo y máximo cuando no existen datos simétricos, además de medias y desviación estándar. El criterio de esta elección se basa en el tipo de dato, la distribución observada y la razón para el análisis (Ranganathan, 2021). En esta etapa descriptiva se puede incluir una variedad de gráficos de tamaño pequeño como barras, líneas por sesión o diagramas de caja para que sean útiles en representar patrones sin derivar inferencias causales.

Se estableció una comparación entre las sesiones con moderado cuidado y bajo ciertos criterios respecto al entramado de los datos reales. En materia metodológica, se sostiene que el elegir las pruebas depende del tipo de datos, de la distribución y si la comparación entre medición es o no pareada; para más de dos mediciones relacionadas, existen alternativas como la ANOVA de medidas repetidas o la prueba de Friedman de acuerdo con las condiciones que presente el conjunto de datos (Ranganathan, 2021).

Los criterios de interpretación se establecen antes de que se presenten los resultados. Cualquier diferencia observada entre sesiones es interpretada como variación dentro del contexto del entrenamiento, no como evidencia de aprendizaje, de empleabilidad o aptitud profesional. En caso de la aplicación de pruebas de contraste, su interpretación incluye estadísticos descriptivos junto con los cambios cuando sean pertinentes. Esto evitaría generalizar solo por la presencia o ausencia de valor estadístico significativo. Las métricas de medición propias del aplicativo se interpretan teniendo en cuenta sus limitaciones tecnológicas: dependencia de proveedores de terceros, parámetros estáticos para evaluar los datos, ponderaciones previstas para distintas cuestiones y la potencial variación en la generación de preguntas o procesamiento asincrónico. La tabla 7 muestra el tratamiento para cada tipo de indicador y dimensión de análisis.

Tabla 7
Variables, indicadores y análisis previsto

Dimensión de análisis	Indicadores previstos	Tratamiento previsto
Funcionamiento del sistema	Entrevistas completadas, estados de procesamiento y disponibilidad de retroalimentación	Depuración y frecuencias
Desempeño por entrevista	Puntajes globales y por tipo de pregunta	Descriptivos y comparación por sesión
Preguntas técnicas orales y código	Evaluación semántica, pruebas superadas, errores de ejecución	Descriptivos por sesión y tipo de pregunta
Comunicación observable	Transcripción, ritmo, pausas, fluidez y presencia observable	Descriptivos por sesión
Usabilidad percibida	Puntaje SUS aplicado al cierre	Descripción global de percepción
Cuestionario propio	Dimensiones definidas e ítems de dificultades	Descripción por dimensión o ítem

Nota. Elaboración propia.

Después de analizar los registros internos y las respuestas externas, se combinan los datos automáticos con la percepción del usuario. Las métricas del sistema hacen posible evaluar el funcionamiento y la experiencia percibida del prototipo durante las entrevistas, mientras que SUS y el cuestionario específico hacen posible interpretar la percepción de la experiencia general de los participantes. Esta lectura no supone una vinculación respuesta por respuesta entre los resultados de Google Forms y las entrevistas internas ni que el desempeño automatizado implica una calidad percibida equivalente, sino que ofrece una comparación general de tendencias para discutir la utilidad formativa del sistema.

Consideraciones éticas, privacidad y eliminación de datos

En cada entrevista el sistema muestra antes de comenzar un aviso informativo sobre la participación y el manejo de datos, y una casilla para aceptar o rechazar el aviso. Este mensaje avisa que la cámara y el micrófono tienen que estar funcionando, que la grabación será continua, que las respuestas por voz, las soluciones de códigos y las evidencias multimedia se

utilizarán para procesar la entrevista, y que la persona puede revisar la evaluación al finalizar la evaluación. El aviso se interpreta como el consentimiento operativo para el uso del prototipo y no como un acuerdo legal completo ni como un consentimiento institucional oficial, puesto que su función consiste en proporcionar información anticipada y adecuada sobre las condiciones de participación, respetando los principios de consentimiento informado, libertad de voluntad, privacidad y posibilidad de retirarse de investigaciones con seres humanos (Laryeafio & Ogbewe, 2023). La exposición de datos internos del sistema se limita únicamente a entrevistas finalizadas satisfactoriamente, mientras que usuarios de SUS y cuestionario particular son considerados fuentes externas de percepción y usabilidad con separación total de cualquier identificador personal. Esto evita que se produzca una identificación innecesaria de desempeño automático a través de los resultados declarados de percepción, y permite la lectura de los formularios con una resolución agregada independiente.

El tratamiento del audio, vídeo, CV y los datos confidenciales exige medidas especiales por la calidad de los datos recopilados. El software admite enviar currículums en formato PDF, grabar audio y vídeo durante la sesión, almacenar archivos en la nube, registrar las metadatos de contenido multimodal, transcribir por preguntas, capturar respuestas a código fuente, resultados de análisis semántico, archivo de audio, vídeo, código y reportes de retroalimentación. Todos estos datos se recopilan para realizar las simulaciones, emitir evaluaciones automáticas y realizar el análisis de aprendizaje planeado. No se utilizarán estos datos para perfilar las habilidades de trabajo de los participantes, para tomar decisiones sobre su formación, calificaciones, certificación de competencias o aptitud profesional. Se debe tener en cuenta que la interpretación moral del uso de los registros debe señalar que la confidencialidad, privacidad, voluntariedad y prevención del uso indebido son elementos cruciales si se recopilan datos por medio de entrevistas o interacciones con participantes (Laryeafio & Ogbewe, 2023). La tabla 8 resume las medidas de mitigación que se van a realizar.

Tabla 8
Medidas éticas, privacidad y tratamiento de datos

Tipo de dato	Finalidad	Riesgo principal	Medida de mitigación
CV y perfil técnico	Construir el perfil de práctica y orientar la entrevista	Identificación directa o indirecta	Uso académico, acceso restringido y reporte sin datos personales
Audio y video	Registrar la entrevista y habilitar procesamiento multimodal	Exposición de voz, rostro o entorno	Aviso previo, aceptación, confidencialidad y no publicación de grabaciones
Respuestas orales transcritas y código	Evaluar preguntas técnicas, comunicativas y de programación	Asociación con desempeño individual	Análisis del prototipo y reporte agregado de resultados
Métricas y retroalimentación	Describir funcionamiento y reporte formativo generado	Sobreinterpretación de puntajes	Límites explícitos: no certificación, no contratación, no calificación académica
SUS y cuestionario propio	Recoger usabilidad y percepción externa	Vinculación innecesaria con registros internos	Aplicación en Google Forms sin enlace directo a datos del sistema

Nota. Elaboración propia.

Después de terminar el análisis académico y la presentación/evaluación del Proyecto de Integración Curricular, se eliminarán manualmente los datos recopilados, tales como los registros de la base de datos, las hojas de vida, los archivos de audio, video, los informes generados y las respuestas de cuestionarios externos. Esta eliminación forma parte del cierre operativo del proyecto. Durante este tiempo, el uso de la información está limitado a la evaluación del prototipo y al presentar los resultados bajo una presentación agregada o anonimizada. Ningún caso se utilizará la información o los informes para contrataciones, calificaciones académicas, acreditación de competencias, medida de empleabilidad o determinación de la adecuación laboral.

Capítulo IV

Análisis y discusión de los resultados

En este capítulo se presentan los resultados obtenidos por medio de las pruebas del sistema multimodal de IA dedicado a la simulación de entrevistas laborales para desarrolladores de software. El análisis incluye la documentación interna de desempeño proveniente de las reuniones de entrevista, así como el análisis derivado de los aspectos secundarios y la opinión de los participantes en el cierre del escenario. Según la planificación de investigación explicada en el apartado anterior, se realizaron análisis descriptivos e inferenciales para medir la diferencia en las puntuaciones entre reuniones, utilizando técnicas estadísticas para el tratamiento de datos con Python. El presente capítulo está organizado alrededor de las fuentes utilizadas, el modo de tratamiento, el desempeño por reunión, el análisis estadístico, el componente de evaluación, la valoración percibida y la discusión integral de los hallazgos.

Conjunto de datos analizado

El conjunto de datos utilizado en este capítulo se obtuvo de la evaluación del prototipo con 20 estudiantes. Cada participante realizó las tres sesiones programadas, de acuerdo con el diseño de medidas repetidas. Esta estructura permitió efectuar comparaciones intra-sujeto coherentes entre los tres momentos de evaluación. Cada participante realizó las tres sesiones programadas, de acuerdo con el diseño de medidas repetidas. Esta estructura permitió efectuar comparaciones intra-sujeto coherentes entre los tres momentos de evaluación.

Tabla 9
Resumen del conjunto de datos analizado

Elemento	Valor
Participantes evaluados	20
Sesiones por participante	3
Entrevistas analizadas	60

Elemento	Valor
Preguntas por entrevista	8
Escala de puntuaciones internas	0 a 10
Instrumentos de percepción	SUS y cuestionario complementario

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Los datos analizados provienen de dos fuentes principales, tal como se visualiza en la tabla 9. El primero es un registro interno del sistema donde se guardan tanto las puntuaciones de desempeño por cada sesión como los datos referentes a la calificación de la respuesta técnica y la respuesta verbal. La segunda es un cuestionario SUS y un cuestionario descriptivo para evaluar la usabilidad y la percepción de la experiencia que se administran al final de la experiencia. Se presentan todas las estadísticas conseguidas respetando el anonimato del encuestado.

La organización del conjunto de datos implica responder al objetivo de medir la actuación del prototipo en dos niveles: por una parte, el desarrollo de la conducta observada durante las prácticas; por otra, el grado de aceptación percibido de la utilidad del sistema formativo, la utilidad educativa y la experiencia de usuario experimentada por los participantes. El capítulo no se detiene en ofrecer únicamente datos numéricos, sino que vincula los hallazgos con el propósito de evaluar si el sistema puede servir de soporte para la realización de entrevistas laborales.

Preparación de datos y estrategia de análisis

Los datos fueron tratados antes del análisis y se organizaron de acuerdo con la estructura metodológica propuesta para la evaluación. Las entrevistas se agruparon por participante y sesión, verificando que cada participante dispusiera de registros correspondientes a los tres momentos de aplicación. Asimismo, las salidas fueron anonimizadas para evitar la exposición de identificadores personales y garantizar que los

resultados presentados no contuvieran información identificable. Las variables más importantes del rendimiento fueron la puntuación general de la entrevista, la puntuación en las preguntas técnicas, la puntuación en las preguntas comunicativas y la puntuación en los ejercicios de codificación. Estos elementos se miden en una escala de 0 a 10 y proporcionan el eje de análisis descriptivo e inferencial, ya que facilitan determinar si hubo un cambio en el rendimiento de las sesiones iniciales.

La relevancia de estos aspectos radica en su función explicativa e informativa: son utilizados para entender qué sucedió con diversos elementos de la evaluación multiforma, pero no se incorporan como un apartado específico para realizar inferencias. La tabla 10 resume las variables consideradas y el tipo de análisis aplicado.

Tabla 10
Desempeño principal por sesión

Bloque de análisis	Variables consideradas	Fuente de datos	Tipo de análisis aplicado
Desempeño principal por sesión	Puntuación global de entrevista, puntuación en preguntas técnicas, puntuación en preguntas comunicativas y puntuación en ejercicios de código	Registros internos de entrevistas	Estadística descriptiva, prueba de normalidad, ANOVA de medidas repetidas, comparaciones pareadas con corrección de Holm y tamaño del efecto
Componentes secundarios de evaluación	Componentes semánticos, de audio y de video en preguntas técnicas y comunicativas	Registros internos de componentes	Estadística descriptiva por sesión
Usabilidad percibida	Puntaje global SUS	Cuestionario de percepción	Cálculo estándar de SUS y estadística descriptiva
Percepción complementaria	Utilidad formativa, pertinencia de la simulación, retroalimentación y dificultades técnicas o de interacción	Cuestionario complementario	Estadística descriptiva por dimensiones

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

La estrategia inferencial incluyó primero la prueba de Shapiro-Wilk para verificar el supuesto de normalidad en cada variable y sesión. Posteriormente, se aplicó ANOVA de medidas repetidas para comparar las tres sesiones en las cuatro variables principales. Cuando

correspondía profundizar en las diferencias entre sesiones, se realizaron comparaciones pareadas con corrección de Holm. Como medida de magnitud del cambio se reportó eta cuadrado parcial, lo que permite complementar la significancia estadística con una lectura del tamaño del efecto.

En cuanto a los instrumentos de percepción, el SUS se calculó con su transformación estándar a una escala de 0 a 100, por lo que se interpreta como una medida de usabilidad percibida y no como una medida de desempeño. El cuestionario complementario se organizó en dimensiones descriptivas: utilidad formativa, pertinencia de la simulación, retroalimentación y dificultades técnicas o de interacción. Esta última dimensión requiere una lectura inversa respecto de las dimensiones positivas, porque valores más altos indican mayor presencia de dificultades.

Desempeño de los participantes por sesión

En la sección de análisis estadístico descriptivo de las puntuaciones primarias, los datos ofrecen una vista general del comportamiento de las variables previo a su análisis estadístico y proporcionan una primera aproximación al desarrollo del rendimiento dentro del entorno simulado. Las notas están representadas entre 0 y 10; por tanto, las medianas pueden revelar dónde se sitúa el grupo durante la sesión y las desviaciones estándar, los máximos y los mínimos ayudarán a examinar la dispersión de los datos.

Tabla 11
Estadísticos descriptivos de las puntuaciones principales por sesión

Variable	Sesión	n	Media	Desviación estándar	Mediana	Mínimo	Máximo
Puntuación global de entrevista	1	20	6.91	0.79	6.94	5.33	8.52
Puntuación global de entrevista	2	20	7.08	0.74	7.13	5.44	8.22
Puntuación global de entrevista	3	20	7.51	0.89	7.73	6.11	9.32

Puntuación en preguntas técnicas	1	20	6.67	0.73	6.67	5.45	7.74
Puntuación en preguntas técnicas	2	20	6.64	0.89	6.54	5.32	8.31
Puntuación en preguntas técnicas	3	20	7.13	0.67	7.24	5.81	8.50
Puntuación en preguntas comunicativas	1	20	6.85	0.75	7.03	4.90	7.90
Puntuación en preguntas comunicativas	2	20	7.00	0.70	6.94	5.97	8.41
Puntuación en preguntas comunicativas	3	20	7.51	0.52	7.60	6.40	8.27
Puntuación en ejercicios de código	1	20	7.10	0.93	7.16	5.60	8.59
Puntuación en ejercicios de código	2	20	6.92	0.89	7.06	5.40	8.15
Puntuación en ejercicios de código	3	20	7.02	0.81	7.04	5.41	8.56

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

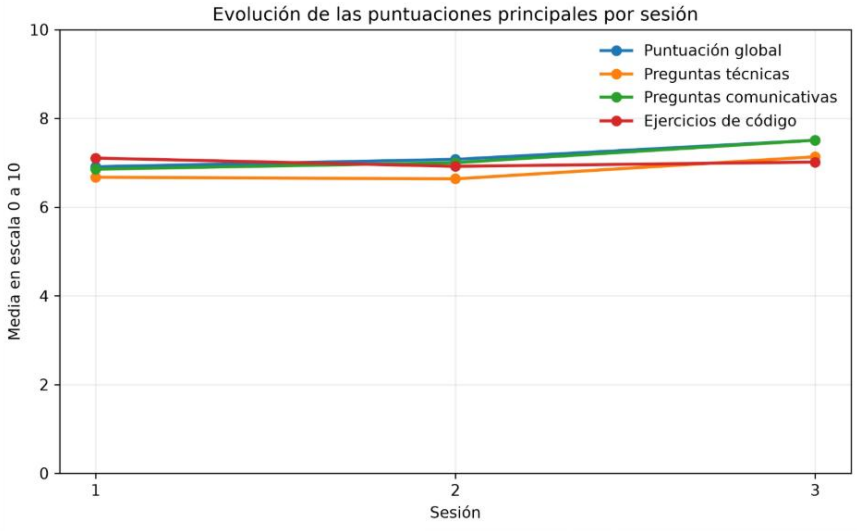
El puntaje global de la entrevista, como se muestra en la Tabla 11, aumenta de 6.91 en la primera sesión a 7.51 en la tercera. Esta tendencia indica que se produjo una mejora general en la performance dentro del entorno práctico; es particularmente interesante la comparación de la primera y última sesiones. La puntuación de preguntas comunicativas también refleja ese cambio pasando de 6.85 a 7.51 lo que implica que, probablemente, esa componente fue una de las más notoriamente sensibles durante el período evaluado.

Las puntuaciones han crecido levemente, aproximadamente 0.60 puntos desde cero hasta diez puntos en la escala global. Además, las respuestas sobre comunicación aumentaron aproximadamente 0.66 puntos. Estas son variaciones positivas y moderadas; son indicadores positivos para un análisis breve basado en tres sesiones. De modo que estos hallazgos

muestran una evolución visible, pero aún son preliminares y por tanto no representan un cambio definitivo o estable.

La medición de preguntas técnicas experimenta un incremento significativo en las primeros dos sesiones, aunque esto no es una línea recta, ya que la segunda muestra una puntuación levemente menor que la primera. Estas variaciones indican que la capacidad técnica puede variar sin que ello signifique un crecimiento lineal; es decir, el desempeño técnico no siempre sigue un patrón constante, sino que se mueve antes de alcanzar su máximo valor en la tercera sesión. Por otra parte, en cuanto a los cuestionarios de programación, la puntuación es casi constante entre las diferentes sesiones, sin mostrar un patrón claro de progreso continuo.

Figura 9
Evolución de las puntuaciones principales por sesión



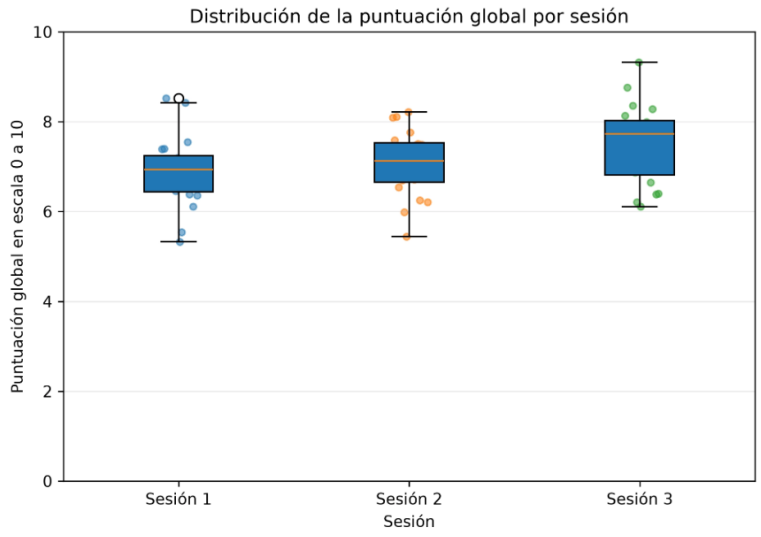
Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

La Figura 9 sintetiza visualmente estas diferencias de comportamiento. La puntuación global y la puntuación en preguntas comunicativas muestran trayectorias ascendentes más claras hacia la tercera sesión, mientras que la puntuación en preguntas técnicas presenta una recuperación después de la segunda sesión. La línea correspondiente a ejercicios de código se

mantiene más plana, lo que refuerza la idea de estabilidad relativa en esta variable durante el periodo evaluado.

Figura 10

Distribución de la puntuación global por sesión



5

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Según la Figura 10, se complementa el análisis de la calificación total mostrando cómo los puntajes se distribuyen por sesión. En general, la tercera sesión mantiene los niveles elevados de puntuación que alcanza la primera, aunque sigue siendo posible observar diferencias entre individuos. Esta variación puede entenderse como parte habitual cuando hay aprendizaje individual, donde los avances registrados pueden influir en la familiaridad del formato de entrevista, la respuesta al formato, la experiencia antecedente y el tipo de pregunta percibida.

La experiencia de uso del sistema estuvo precedida por un comienzo inicial positivo con respecto al desempeño general y las preguntas comunicativas, pero el desempeño técnico no ha progresado de manera uniforme y el desempeño sobre ejercicios de código ha mantenido un rendimiento casi constante. Este panorama descrito anteriormente sirve como fundamento para analizar los resultados estadísticos que se presentarán más adelante.

Comparación estadística entre sesiones

Se realizó un análisis inferencial por medio de medidas repetidas para las cuatro variables centrales analizadas. Esta forma de análisis es apropiada dada la estructura de la investigación, que mide la misma muestra en tres momentos. A priori se confirmó la asunción de normalidad con base al criterio de Shapiro-Wilk para cada par de variable-sesión.

Tabla 12
Prueba de normalidad Shapiro-Wilk para las variables principales

Variable	Sesión 1 p	Sesión 2 p	Sesión 3 p	Decisión
Puntuación en ejercicios de código	0.425	0.194	0.991	Cumple normalidad en las tres sesiones
Puntuación en preguntas comunicativas	0.184	0.398	0.502	Cumple normalidad en las tres sesiones
Puntuación en preguntas técnicas	0.221	0.356	0.840	Cumple normalidad en las tres sesiones
Puntuación global de entrevista	0.410	0.794	0.498	Cumple normalidad en las tres sesiones

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Los resultados de la Tabla 12 muestran que las variables principales cumplieron el supuesto de normalidad en las tres sesiones, dado que los valores p fueron superiores al nivel de significancia establecido. Esta verificación respalda el uso del ANOVA de medidas repetidas para comparar las sesiones, aunque la interpretación se mantiene prudente por el tamaño de la muestra y por el carácter aplicado del estudio.

Tabla 13
ANOVA de medidas repetidas para la comparación entre sesiones

Variable	F	Grados de libertad	p	Eta cuadrado parcial	Interpretación
Puntuación global de entrevista	5.984	2, 38	0.006	0.239	Diferencia significativa entre sesiones

Variable	F	Grados de libertad	p	Eta cuadrado parcial	Interpretación
Puntuación en preguntas técnicas	3.497	2, 38	0.040	0.155	Diferencia significativa entre sesiones
Puntuación en preguntas comunicativas	6.613	2, 38	0.003	0.258	Diferencia significativa entre sesiones
Puntuación en ejercicios de código	0.238	2, 38	0.790	0.012	Sin diferencia significativa entre sesiones

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Tabla 13 evidencia diferencias significativas respecto a la calificación total de la entrevista, de las respuestas técnicas y de las respuestas comunicativas, frente a la calificación de ejercicios de código, donde no se ha producido ninguna diferencia estadísticamente significativa. Esta diferencia es importante dado que revela que no hubo variaciones uniformes de acuerdo con las diversas dimensiones evaluadas: los cambios ocurren en las puntuaciones totales y en las respuestas técnicas y comunicativas, mientras que el ejercicio de codificación permanece constante desde un punto de vista estadístico.

Los tamaños del efecto permiten complementar esta lectura. La puntuación en preguntas comunicativas tuvo el mayor valor de eta cuadrado parcial, seguida por la puntuación global de entrevista y la puntuación en preguntas técnicas. Esto indica que, dentro de la muestra estudiada, la dimensión comunicativa fue la que mostró más claramente la variación entre sesiones. El tamaño del efecto fue bajo, en concordancia con la ausencia de diferencias significativas, en ejercicios de código.

Tabla 14
Comparaciones pareadas con corrección de Holm

Variable	Comparación	Diferencia media	p ajustado	ANOVA significativo	Interpretación
Puntuación global de entrevista	Sesión 1 vs Sesión 2	0.166	0.407	Sí	Comparación no significativa
Puntuación global de entrevista	Sesión 1 vs Sesión 3	0.597	0.010	Sí	Comparación significativa
Puntuación global de entrevista	Sesión 2 vs Sesión 3	0.431	0.027	Sí	Comparación significativa
Puntuación en preguntas técnicas	Sesión 1 vs Sesión 2	-0.037	0.844	Sí	Comparación no significativa
Puntuación en preguntas técnicas	Sesión 1 vs Sesión 3	0.461	0.073	Sí	Comparación no significativa
Puntuación en preguntas técnicas	Sesión 2 vs Sesión 3	0.497	0.124	Sí	Comparación no significativa
Puntuación en preguntas comunicativas	Sesión 1 vs Sesión 2	0.146	0.501	Sí	Comparación no significativa
Puntuación en preguntas comunicativas	Sesión 1 vs Sesión 3	0.653	0.006	Sí	Comparación significativa
Puntuación en preguntas comunicativas	Sesión 2 vs Sesión 3	0.508	0.014	Sí	Comparación significativa
Puntuación en ejercicios de código	Sesión 1 vs Sesión 2	-0.185	1.000	No	Comparación no significativa
Puntuación en ejercicios de código	Sesión 1 vs Sesión 3	-0.089	1.000	No	Comparación no significativa
Puntuación en ejercicios de código	Sesión 2 vs Sesión 3	0.096	1.000	No	Comparación no significativa

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Las comparaciones parciales de la Tabla 14 permiten identificar los momentos específicos en los que se registraron las diferencias encontradas en el análisis completo, pues, según la puntuación total de la encuesta, las diferencias estadísticamente significativas se evidenciaron al comparar la primera sesión con la tercera y la segunda sesión con la tercera, lo que sugiere que el cambio más importante se produjo al finalizar el ciclo de práctica y que la variación no alcanzó significancia entre las dos primeras sesiones, sino únicamente después de incorporar la evaluación correspondiente a la tercera sesión. Un mismo patrón se observó en el aspecto de las preguntas comunicativas, pero con una señal estadística más clara. En esta dimensión, las diferencias significativas también se situaron entre la primera y la tercera sesión, y entre la segunda y la tercera. Esta perspectiva permite considerar que las preguntas comunicativas mostraron una evolución cada vez mayor en la última sesión, probablemente por la repetición de la experiencia y la facilitación de un manejo más familiar con la exposición oral, organización del razonamiento y forma de responder en el ambiente simulado. Sin embargo, la ausencia de diferencias estadísticamente significativas entre la primera y la segunda sesión sugiere que el cambio no fue súbito, sino gradual y más notorio al llevar a cabo las tres sesiones.

Por otro lado, las preguntas técnicas y los problemas prácticos que requieran un análisis más detallado en comparación con las preguntas que soliciten respuestas escritas sobre un tema específico o sobre el propio proceso. La comparación de las preguntas técnicas reveló diferencias significativas respecto de sesiones anteriores (diferenciación del ANOVA), sin embargo, dichas diferencias no quedaron comprobadas si se usan la corrección de Holm y se considera que hay una variación en la dinámica total, pero no en ninguna pareja concreta particular. Esto implica no extender el hallazgo, evitando así la sobre interpretación de la evidencia. Además, la puntuación que se otorgó por ejercicio de códigos no mostró diferencias estadísticamente significativas ni en el ANOVA ni en las comparaciones de contraste, de modo que la lectura de la constante dinámica es aceptable desde el punto de vista estadístico. En el

contexto del producto en desarrollo, este hecho podría deberse a que la resolución del problema de codificación está menos susceptible a cambios relativamente breves, o requiere más práctica específica o condiciones especiales para evidenciar variaciones claras. En conclusión, el análisis inferencial confirma parcialmente los hallazgos descriptivos, ya que el aumento inicial se observa con mayor claridad en la puntuación general y en las preguntas comunicativas, mientras que las preguntas técnicas presentan una variación global que exige cautela al interpretar los contrastes específicos y los ejercicios de código muestran un comportamiento más estable, por lo que resulta necesario complementar el análisis con los componentes secundarios para identificar qué elementos de las respuestas acompañaron las variaciones encontradas.

Análisis descriptivo de componentes de evaluación

Se incluyó el análisis de componentes secundarios con el propósito de ampliar la interpretación de las calificaciones técnicas y de comunicación. Distinto al análisis precedente, este apartado no pretende formar parámetros estadísticos entre sesiones; busca explicar cómo respondieron los componentes semánticos, del audio y del video con respecto a las dos formas de pregunta examinadas. Su relevancia se manifiesta por presentar una lectura más profunda de la evaluación registrada dentro del marco de una percepción multimodal.

Tabla 15
Componentes de evaluación por tipo de pregunta y sesión

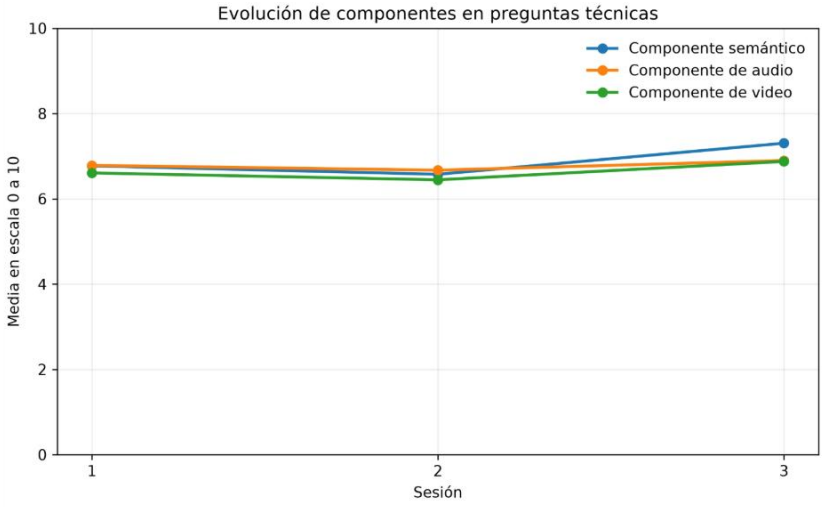
Tipo de pregunta	Componente	Sesión 1 media	Sesión 2 media	Sesión 3 media	Tendencia
Preguntas técnicas	Componente semántico	6.77	6.58	7.30	Descenso intermedio y recuperación final
Preguntas técnicas	Componente de audio	6.78	6.67	6.90	Descenso intermedio y recuperación final
Preguntas técnicas	Componente de video	6.61	6.45	6.88	Descenso intermedio y recuperación final
Preguntas comunicativas	Componente semántico	6.90	7.07	7.48	Aumento progresivo

Tipo de pregunta	Componente	Sesión 1 media	Sesión 2 media	Sesión 3 media	Tendencia
Preguntas comunicativas	Componente de audio	6.96	7.06	7.42	Aumento progresivo
Preguntas comunicativas	Componente de video	6.69	7.17	7.60	Aumento progresivo

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

En las preguntas técnicas, los tres componentes no siguen una trayectoria lineal. La Tabla 15 muestra un descenso intermedio en la segunda sesión y una recuperación hacia la sesión final, con el cambio más visible en el componente semántico. Este comportamiento es coherente con una posible variación en la organización, precisión o pertinencia del contenido de las respuestas técnicas, pero no debe interpretarse como una explicación causal del resultado global. Los componentes de audio y video también recuperan valores hacia la tercera sesión, aunque con variaciones más moderadas.

Figura 11
Evolución de componentes en preguntas técnicas

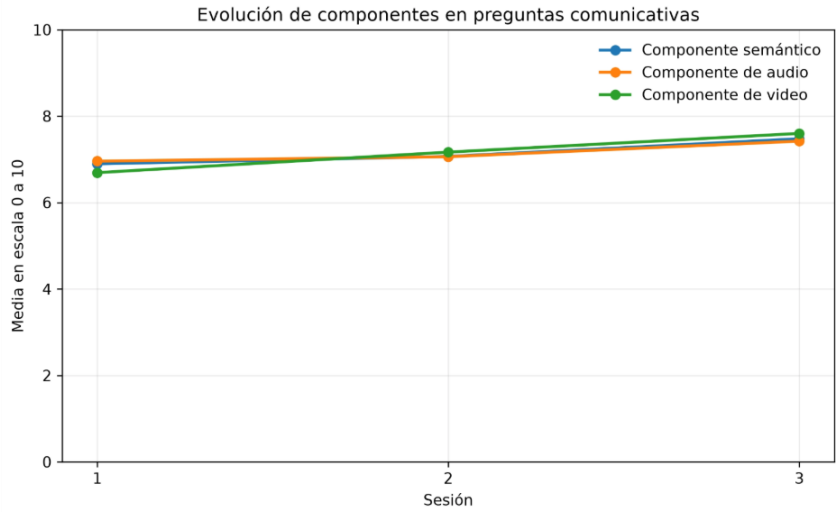


Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Se puede percibir en la Figura 11 que las preguntas técnicas se desarrollaron de forma más sistemática que las preguntas comunicativas. La tendencia hacia un componente semántico mayor puede indicar respuestas más estructuradas o adecuadas; el componente audio podría sugerir modificación en la precisión o la calidad del discurso verbal; mientras que el componente video podría señalar mayor coherencia de la presencia física presente durante la conversación. Las lecturas propuestas deberían ajustarse al límite descriptivo del sistema investigativo y los criterios de alcance de la investigación realizada.

Figura 12

Evolución de componentes en preguntas comunicativas



Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Estos elementos secundarios permiten entender por qué las preguntas comunicativas muestran una señal más destacada de desarrollo inicial, como lo muestra la figura 12. Aunque en las preguntas técnicas hay un descenso intermedio seguido de una recuperación, esto es más evidente en el contenido semántico. Los tres componentes avanzan de una manera mucho mayor en las preguntas comunicativas. Esta diferencia proporciona una explicación plausible para las tendencias observadas, sin transformar el análisis en una prueba causal suplementaria.

Usabilidad y percepción de los participantes

Además de los registros internos de desempeño, la evaluación también incluyó instrumentos de percepción que se administraron después de la experiencia. Estas informaciones facilitan el análisis de cuán apropiadamente los sujetos percibieron la usabilidad del sistema y su valoración dentro de un proceso de práctica de entrevistas. Se muestra la información en forma generalizada, lo que no puede interpretarse como evidencia del desempeño o como una validación externa de las habilidades.

Resultado del cuestionario SUS

El cuestionario SUS permite obtener una medida global de usabilidad percibida en una escala de 0 a 100. Su uso en este estudio aporta una referencia sintética sobre la experiencia de interacción con el sistema, especialmente en relación con la facilidad de uso, consistencia y aceptación general del prototipo. La interpretación del SUS se mantiene separada de las puntuaciones de desempeño, ya que evalúa percepción de usabilidad y no rendimiento en la entrevista.

Tabla 16
Resultado global del cuestionario SUS

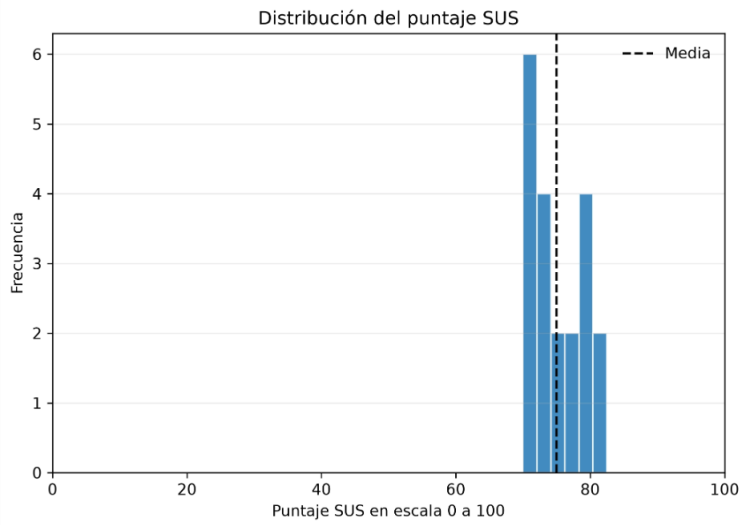
n	Media	Desviación estándar	Mediana	Mínimo	Máximo	Interpretación
20	75.00	4.59	73.75	70.00	82.50	Usabilidad percibida favorable

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

La Tabla 16 presenta un puntaje promedio de 75.00, lo que indica que la percepción de usabilidad es positiva. Esto se refleja en que los participantes pudieron utilizar el sistema de forma razonable y que el rendimiento global no mostró problemas significativos de acceso. Aunque la medición no debería utilizarse para destacar el sistema como sobresaliente o indicar

que todo funcionó sin inconvenientes, sí puede considerarse que se logró una buena experiencia general de uso.

Figura 13
Distribución del puntaje SUS



Nota. Elaboración propia con base en los resultados del cuestionario SUS aplicado a los participantes.

La Figura 13 evidencia una distribución concentrada en valores positivos, con puntuaciones encontradas en el rango 70.00 a 82.50. Este grado de concentración sugiere que la percepción de usabilidad fue coherente entre los miembros del grupo completo. La menor variabilidad muestra que no hubo diferencias notables dentro de dicho conjunto de datos.

Desde la perspectiva del prototipo, un índice satisfacción favorable tiene relevancia porque el valor formativo percibido de una herramienta de práctica se basa no solo en su capacidad para proporcionar retroalimentación, sino también en que los usuarios puedan entender y usar el entorno de forma eficiente. Esta interpretación se ajusta a estudios recientes sobre entrevistas técnicas simuladas con IA, en los que el realismo percibido, la retroalimentación y la confianza del usuario resultan ser factores clave para el uso formativo (Daryanto et al., 2025). Para ello, la usabilidad percibida contribuye positivamente a los resultados de desempeño, aunque no los justifica por sí misma.

Resultados del cuestionario complementario

El cuestionario complementario se diseñó para describir percepciones específicas sobre la experiencia con el sistema. A diferencia del SUS, no se presenta como una escala estandarizada ni validada, sino como un instrumento descriptivo que permite agrupar respuestas en dimensiones relevantes para el estudio: utilidad formativa, pertinencia de la simulación, retroalimentación y dificultades técnicas o de interacción.

Tabla 17
Resultados por dimensión del cuestionario complementario

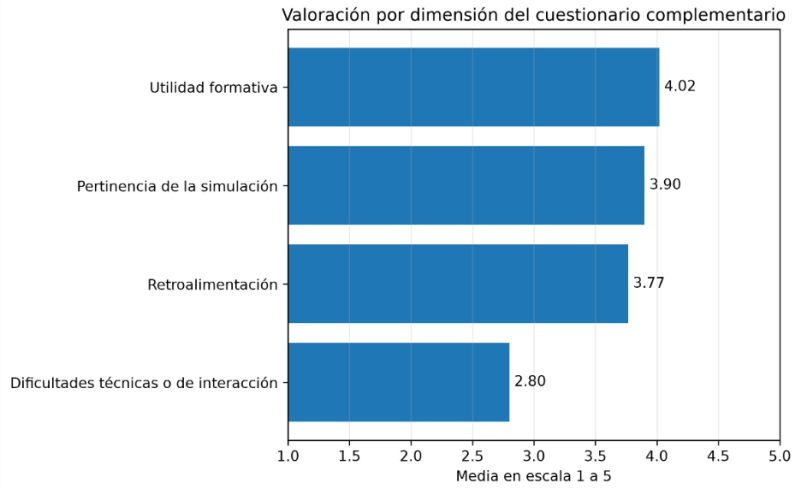
Dimensión	n	Media	Desviación estándar	Mediana	Mínimo	Máximo
Utilidad formativa	20	4.02	0.32	4.00	3.60	4.60
Pertinencia de la simulación	20	3.90	0.31	4.00	3.00	4.00
Retroalimentación	20	3.77	0.38	3.67	3.33	4.33
Dificultades técnicas o de interacción	20	2.80	0.77	3.00	2.00	4.00

Nota. Elaboración propia con base en los datos recolectados durante la evaluación del prototipo.

Los resultados de la Tabla 17 indican valoraciones positivas en las dimensiones de utilidad formativa, pertinencia de la simulación y retroalimentación. La utilidad formativa obtuvo la media más alta, lo que sugiere que los participantes percibieron el sistema como un apoyo pertinente para practicar entrevistas. La pertinencia de la simulación también se ubicó en un nivel favorable, lo que es relevante para un prototipo cuyo propósito es aproximar al estudiante a un escenario de entrevista laboral en desarrollo de software.

Figura 14

Valoración por dimensión del cuestionario complementario



Nota. Elaboración propia con base en los resultados del cuestionario complementario aplicado a los participantes.

La Figura 14 permite visualizar las dimensiones del cuestionario complementario. Los puntajes más elevados se ubican en la utilidad formativa y la pertinencia de la simulación; en cambio, la retroalimentación muestra una puntuación promedio superior, aunque ligeramente menor. Esto sugiere que los participantes valoraron la experiencia en general y la calidad formativa del sistema, si bien la retroalimentación debería contemplar mejoras futuras.

La dimensión de las dificultades técnicas y de la interacción debe interpretarse de una manera distinta. La formulación en negativo significa que una media más alta indica mayor presencia de inconvenientes. El valor observado, 2.80 sobre 5, no confirma la ausencia de problemas ni indica que la dificultad sea predominante. Esta lectura es coherente con el Índice de Satisfacción con la Usabilidad (SUS), ya que la usabilidad general del sistema es favorable y existe espacio para mejorar ciertos aspectos de la interacción.

En suma, los resultados de percepción alinean el análisis de rendimiento. Las valoraciones de utilidad y pertinencia dan cuenta de un enfoque positivo respecto a la utilidad del sistema como medio de práctica; la dimensión de dificultades sugiere que aún queda

espacio para mejorar el uso del sistema. Esta combinación de aceptación general y oportunidades para refinamiento es común en sistemas formativos evaluados en uso real.

Discusión integrada de los hallazgos

Los resultados del capítulo muestran una evolución inicial favorable en parte de las dimensiones evaluadas, especialmente en la puntuación global de entrevista y en las preguntas comunicativas. Esta evolución no fue homogénea en todas las variables: las preguntas técnicas presentaron evidencia de variación global, pero sin contrastes pareados significativos después de la corrección aplicada, mientras que los ejercicios de código se mantuvieron estables. Por tanto, el sistema no debe interpretarse como evidencia de cambios uniformes en todos los aspectos del desempeño, sino como un entorno de práctica en el que algunas dimensiones se acompañaron de variaciones más claras durante el periodo evaluado.

La puntuación global de entrevista se compone como un resumen del rendimiento observado y por tanto, su incremento entre las primeras y las tres sesiones ofrece relevancia para evaluar el comportamiento del prototipo. En el marco del ensayo, tales cambios muestran que el acercamiento continuado a entrevistas simuladas lleva acompañado de una mayor habilidad para manejar el formato, la estructura de respuestas y la dinámica de retorno de opiniones. Esta interpretación es compatible con el fin formativo del programa, cuyo objetivo es ofrecer un espacio para la práctica previa a enfrentar situaciones verídicas de conversaciones laborales.

La evolución más clara se ha evidenciado en las preguntas comunicativas. Esto es especialmente significativo, dado que las entrevistas técnicas examinan no solo los conocimientos técnicos, sino también la habilidad para transmitir decisiones, expresar su razonamiento y mantener una conversación efectiva durante el proceso de resolución de problemas (Behroozi et al., 2022; Daryanto et al., 2025). La presión de la entrevista también puede influir en tanto en la comunicación como en los hallazgos, razón por la cual esta faceta es de gran interés para interpretar los resultados con cuidado (Behroozi et al., 2020). En el

conjunto de casos estudiado, parece que las preguntas comunicativas han sido aquellos lugares donde se ha observado un cambio más consistente entre sesiones.

Las preguntas técnicas demandan una interpretación más elaborada. A pesar del desfase global entre las sesiones, la falta de parajes comparativos significativos después de la corrección de Holm sugiere que no se puede afirmar con seguridad una determinada pareja de sesiones como fuente de diferencia. Este hallazgo puede estar ligado a la variabilidad propia de las preguntas técnicas, al tema cubierto por cada sesión o a diferencias individuales en los conocimientos previos. Sea como sea, la información recopilada admite referirse a una variabilidad global, pero no a un progreso técnico concluyente entre las sesiones particulares.

La estabilidad de los ejercicios de codificación también proporciona información valiosa. Resolver ejercicios de programación durante una entrevista requiere habilidades diferentes a las necesarias para contestar preguntas conceptualizadas o de comunicación: comprensión del problema, elaboración de una estrategia, administración de sintaxis, lógica racional y ejecución dentro de un período con plazos fijados. La literatura sobre retroalimentación automática en la programación alerta igualmente que la identificación de fallos de pensamiento lógico y la generación de retroalimentación de calidad continúa resultando problemáticas para las inteligencias artificiales que emplean sistemas basados en modelos de lenguaje (Silva & Costa, 2025). Que esta faceta no haya revelado desplazamientos sustanciales indica que el lapso evaluado pudiera no haber sido lo bastante prolongado para evidenciar diferencias notables en las tareas de codificación, o que estas implican un diseño de práctica específico para dar cuenta de modificaciones apreciables observables.

El análisis de componentes secundarios nos permite explorar la naturaleza multifacética del sistema. En las preguntas técnicas, el componente semántico demostró el cambio descriptivo más evidente, lo que coincide con la relevancia del contenido conceptual en esta clase de respuesta. En las preguntas comunicativas, todos los componentes evolucionaron en igual medida; sin embargo, en particular en el componente de video se detectó una señal más

fuerte. Esta interpretación debe estar limitada al alcance del sistema: los componentes describen atributos observables en la respuesta e incluso en ellos no se debería inferir información sobre rasgos personales, estados de emoción profundos o competencias laborales. Este grado de precaución responde a investigaciones sobre entrevistas automatizadas y evaluaciones mediante IA, que insisten en la necesidad de examinar la equidad, la interpretabilidad, la transparencia y los límites de inferencia cuando se emplean señales multimodales (Hunkenschroer & Luetge, 2022; Putra et al., 2024).

Las percepciones de uso y utilidad de los sistemas de práctica son positivas. La puntuación obtenida por SUS alcanzó los 75.00, lo cual indica una usabilidad favorable. Además, los resultados del cuestionario de percepción indican una visión positiva de la utilidad formativa, pertinencia y la capacidad de recoger retroalimentación. Dichos hallazgos concuerdan con la intención básica de la aplicación de un sistema: la retroalimentación automática debe resultar útil desde el punto de vista formativo si está concebido de modo tal que el entorno sea fácil de comprender, relevante e implementable. Del mismo modo, la investigación sobre la preparación de entrevistas técnicas subraya que esta actividad puede mejorar la familiaridad y la confianza; sin embargo, ello no constituye una garantía directa de empleo ni de desempeño profesional (Kapoor & Gardner-McCune, 2025).

Se desarrolló una discusión que tenga en cuenta las advertencias del marco teórico sobre la utilización de IA en procesos de evaluación. El uso automático puede ayudar a la práctica, proporcionar retroalimentación en tiempo real y permitir más oportunidades para la preparación, pero debe hacerlo con transparencia, interpretación prudente y conocimiento de sus limitaciones. Los resultados muestran evidencia inicial del valor de la simulación y retroalimentación como entorno del sistema, pero no como elemento decisivo para definir competencias profesionales o resultados laborales.

A modo de conclusión, en términos sencillos, el análisis demuestra que el prototipo sirvió para documentar y cuantificar la efectividad de los individuos en sesiones de prueba por

tres ocasiones. Además, las personas mostraron un buen grado de satisfacción al usarlo. Los datos obtenidos demuestran mejor que el desempeño global y el de comunicación que otras habilidades técnicas y de programación. Todo esto refleja el sistema en el papel de apoyo a la formación en ambientes analizados, mientras se mantiene una interpretación cuidadosa y apropiada según el diseño metodológico.

Limitaciones del análisis

Los resultados indicados deben interpretarse conociendo las características de los estudios que llevaron a ellos. Aunque el diseño de mediciones repetidas ofrece la posibilidad de evaluar cómo se comportan los mismos sujetos durante tres sesiones, no puede determinar conexiones entre el comportamiento y una causa fuerte. Las diferencias identificadas podrían estar vinculadas con la práctica, con la adaptación al sistema, con variaciones en las preguntas o con elementos personales no cubiertos. El tamaño y la composición de la muestra constituyen una limitación del estudio. La evaluación se realizó con 20 participantes en un contexto académico específico. Aunque el diseño de medidas repetidas permitió comparar consistentemente a las mismas personas durante tres sesiones, el tamaño muestral limita la generalización de los hallazgos a otros grupos, instituciones o contextos educativos. La ausencia de un grupo control también dificulta atribuir las diferencias observadas exclusivamente al uso del prototipo.

Otro aspecto es la duración del período de evaluación. En tres sesiones es posible ver una tendencia de uso inicial, pero no garantizar cambios de escala ni de larga duración. Esto es muy importante para medir los ejercicios de código, donde el crecimiento de la fluidez para resolver los problemas requiere más tiempo de práctica, retroalimentación específica y ejercicio consciente.

Las calificaciones internas son resultados del sistema evaluado, y si bien permiten comparar de forma constante sesiones dentro de un mismo entorno, deben ser tomadas como indicadores funcionales del prototipo y no como métricas objetivas independientes. Por lo tanto,

sirven para estudiar las variaciones entre sesiones dentro de un mismo criterio de evaluación, pero no deben interpretarse como una clasificación definitiva en la competencia profesional. Los elementos de audio y de vídeo explican rasgos observables del intercambio de información y, por tanto, no deben considerarse como medida subjetiva del comportamiento, de los afectos, de las emociones o del perfil de personalidad. Este cuidado, en un sistema que integre señales verbales, acústicas e imágenes, se extiende también a la problemática de equidad, interpretación y claridad. (Hunkenschroer & Luetge, 2022). El cuestionario complementario aporta información desde el punto de vista de las intervenciones, pero no constituye una escala estándar.

Hay que tener cuidado al interpretar los resultados de percepción, ya que el SUS permite evaluar la usabilidad percibida, pero no mide directamente el desempeño ni el aprendizaje de los participantes. También, las puntuaciones positivas que se obtuvieron a través del cuestionario de seguimiento muestran aceptación y utilidad percibida dentro de la muestra, aunque por sí solas no prueban la eficacia del sistema en situaciones reales de contratación, por lo que deben interpretarse como evidencias de la experiencia de uso y no como una confirmación de resultados profesionales externos. Por lo tanto, el análisis no evalúa el desempeño de los participantes en entrevistas de trabajo reales ni la transferibilidad de los resultados a su empleabilidad, sino que se enfoca en mostrar cómo funcionó el prototipo en una práctica simulada, cuáles fueron las características de desempeño que se manifestaron y cuál fue la percepción expresada por los individuos. De esta manera, aunque sujetos a las limitaciones señaladas, los hallazgos pueden constituir una base inicial para explorar el alcance del sistema, así como para sustentar la discusión de los resultados y formular recomendaciones para investigaciones futuras que se deriven del trabajo.

Capítulo V

Este capítulo presenta las conclusiones derivadas del desarrollo y la evaluación del sistema multimodal basado en inteligencia artificial para la simulación de entrevistas laborales en el sector de desarrollo de software. Las conclusiones se organizan en función de los objetivos planteados y de los hallazgos principales obtenidos durante la evaluación del prototipo. Por su parte, las recomendaciones se formulan a partir de los resultados analizados, las limitaciones metodológicas y técnicas identificadas, y las posibilidades de continuidad del trabajo dentro de un alcance formativo, académico y supervisado.

Conclusiones

Con relación al primer objetivo específico, a partir del análisis del contexto se pudo establecer que las entrevistas laborales en desarrollo de software no se reducen únicamente a la verificación de conocimientos aislados de técnicas. Estos procesos integran la resolución de problemas, la explicación del razonamiento, la comunicación profesional, la organización de ideas y la capacidad de responder situaciones vinculadas con la práctica del desarrollo de software, característica que sustenta la necesidad de contar con entornos de práctica que aproximen al estudiante a la dinámica de una entrevista, tanto en la construcción de sus respuestas como en la manera de interpretar una consigna, estructurar una solución, justificar el procedimiento seguido y comunicar con claridad las decisiones tomadas en el transcurso de la actividad. En este sentido, el análisis permite sostener la pertinencia de abordar la preparación para entrevistas laborales como una necesidad formativa situada, sobre todo, para estudiantes o personas recién graduadas que requieren familiarizarse previamente con formatos de evaluación técnica y comunicativa antes de participar en escenarios laborales reales de selección.

Se desarrolló un prototipo de sistema multimodal para la simulación de entrevistas de software enfocado en ofrecer un entorno en el que candidatos puedan prepararse para enfrentar procesos de selección. El sistema permite que las personas suban su currículum

vitae, y lo utiliza para crear perfiles profesionales de usuario, con los que genera entrevistas personalizadas que incluyen preguntas sobre sus habilidades blandas, técnicas y ejercicios de codificación. Las respuestas son captadas de forma multimodal, considerando diferentes entradas como contenido semántico, audio y video. El sistema procesa estos insumos para finalmente generar reportes de retroalimentación que sean de utilidad para mejorar el desempeño. Se diseñó una arquitectura que le permite al prototipo trabajar por etapas definidas y manejar tareas largas de procesamiento pesado de forma consistente. Si bien, los insumos multimodales aportan información complementaria útil para el análisis, no es posible atribuir características avanzadas como reconocimiento de emociones, análisis psicológicos o determinar si un candidato es apto para un puesto de trabajo. En general, el sistema se debe tomar como un entorno controlado para el entrenamiento, y no como una herramienta de selección laboral.

El prototipo fue evaluado mediante un escenario académico controlado con 20 estudiantes. Cada participante realizó tres entrevistas a lo largo de una semana, dando como resultado un total de 60 entrevistas analizadas para estudiar las puntuaciones de desempeño y la percepción de uso del sistema. Las principales diferencias de puntaje se obtuvieron en los promedios de habilidades blandas, que pasaron de 6.85 a 7.51, seguido de los promedios totales de entrevista, que pasaron de 6.91 a 7.51. Las preguntas de habilidades técnicas tuvieron un incremento en sus puntajes, pero al someter los datos al proceso estadístico de comparación pareada estadística no se pudo conseguir evidencia significativa. Los ejercicios de codificación no presentaron diferencias en sus puntuaciones, lo que puede sugerir que este tipo de habilidades requieren de mayor tiempo de entrenamiento para presentar una diferencia tangible. Para terminar, los resultados obtenidos se deben interpretar con cautela, teniendo en cuenta el número reducido de participantes en la evaluación y la ausencia de un grupo de control al momento de realizar el estudio.

Una vez concluidas las entrevistas, los participantes llenaron la encuesta SUS y un cuestionario adicional. Con la primera se buscó medir de forma estandarizada la usabilidad general del sistema, mientras que con el cuestionario se pretendió conocer la percepción acerca de otros aspectos diferentes a la usabilidad, como la calidad y utilidad de los reportes de feedback, y la apreciación sobre la utilidad educativa. La encuesta SUS arrojó un puntaje de 75 puntos aproximadamente, lo que podría considerarse una usabilidad aceptable, pero con margen de mejora. Por otro lado, la encuesta personalizada dio como resultado una percepción general favorable acerca de la retroalimentación, pero dejó aspectos a considerar como problemas técnicos y de procesamiento. Estos indicativos no son una prueba clara de que el sistema aporta a un aprendizaje a largo plazo, ni demuestra que los candidatos hayan mejorado su desempeño en procesos de selección real, sino que indican una percepción favorable hacia el entorno de simulación y su utilidad en general.

En conclusión, en el presente proyecto de integración curricular se logró desarrollar, desplegar y evaluar un prototipo de sistema basado en inteligencia artificial para la simulación de entrevistas laborales de software, con la capacidad de crear preguntas personalizadas y retroalimentación adaptada a cada candidato. El flujo completo de capacidades del sistema incluyó procesamiento de CVs, construcción de perfiles profesionales, generación de preguntas únicas, procesamiento multimodal y generación de reportes de feedback, construyendo así un entorno de simulación que se presenta como apoyo formativo para estudiantes y recién graduados ante procesos de selección laboral en los que se requiera cubrir habilidades técnicas, blandas y resolución de ejercicios de código. La evaluación incluyó la participación de estudiantes reales, dejando métricas que no dejaron evidencia significativa para afirmar que el sistema puede producir mejoras a largo plazo o certifique competencias laborales, sino que aporta un ambiente para familiarizarse con entrevistas de software.

Recomendaciones

Se recomienda ampliar la muestra en futuras evaluaciones del sistema, ya que esto permitiría incorporar una mayor diversidad de participantes, reducir la influencia de variaciones individuales y comprobar si las tendencias identificadas se mantienen en grupos más amplios y con características diferentes, siempre que el incremento de participantes conserve criterios claramente definidos de inclusión, completitud de la información, participación voluntaria, privacidad y tratamiento ético de los datos obtenidos.

Se sugiere ampliar el número de sesiones o extender la duración del seguimiento en futuras evaluaciones, dado que un periodo más prolongado permitiría examinar si la familiaridad adquirida con el formato de entrevista, la revisión sistemática de la retroalimentación y la repetición de las prácticas producen efectos que se mantienen más allá de una experiencia breve. Para atender así la limitación temporal del estudio y facilitando especialmente la interpretación de los ejercicios de código, cuyos resultados permanecieron relativamente estables y podrían requerir una mayor cantidad de práctica deliberada, acompañamiento y tiempo de exposición para evidenciar cambios observables y consistentes.

Se recomienda incorporar un grupo de comparación o adoptar un diseño cuasi-experimental en estudios posteriores, debido a que esta decisión permitiría contrastar los resultados obtenidos mediante el uso del prototipo con los derivados de métodos como la práctica tradicional, la tutoría humana, las entrevistas simuladas sin retroalimentación automatizada u otras estrategias formativas. Para atender así la limitación asociada al diseño de grupo único, que impide determinar con suficiente claridad si las variaciones observadas se relacionan específicamente con la interacción con el sistema, con la familiaridad progresiva con las tareas, con experiencias externas de aprendizaje o con otros factores que no fueron controlados durante la evaluación.

Se sugiere hacer una evaluación más detallada de las pruebas de código. Para esto, se debe revisar la diversidad, la complejidad, la progresión y la cobertura de los ejercicios

planteados, asimismo incluir más pruebas, análisis de errores comunes y retroalimentación técnica específica. Esta indicación hace referencia a la estabilidad relativa de dicha dimensión e indica que evaluar soluciones de programa durante una entrevista puede demandar un mayor tiempo de práctica, más atención y criterios de revisión más extensos. La aplicación de esta sugerencia debe evitar convertir la simulación en una prueba extensa de programación o presentar los resultados del código como prueba de competencia profesional.

Se recomienda mejorar los comentarios de retroalimentación personalizados generados por el sistema mediante una vinculación más precisa entre cada observación y las evidencias obtenidas durante la entrevista, diferenciando claramente las fortalezas identificadas de los aspectos que requieren revisión y proponiendo acciones concretas para la práctica posterior. Esto con el propósito de ayudar tanto el margen de mejora señalado en la percepción de los participantes como los riesgos asociados con la retroalimentación automatizada, por lo que el reporte deberá emplearse para guiar el aprendizaje y la preparación continua, sin emitir veredictos concluyentes sobre competencias profesionales, empleabilidad, desempeño laboral o idoneidad para un puesto determinado.

Se recomienda mantener el sistema como una herramienta de apoyo formativo supervisado, destinada a asignaturas, tutorías, talleres de preparación profesional o laboratorios de práctica académica. Debido a que el prototipo procesa infor5mación sensible, como currículums, grabaciones de voz y video, respuestas orales y soluciones de código, no debe utilizarse como mecanismo para tomar decisiones de contratación, certificar competencias o determinar la idoneidad laboral de los usuarios, mientras que sus futuras versiones deberán reforzar el consentimiento informado, la minimización de datos, el control de acceso, la eliminación segura y la comunicación clara sobre el propósito y el uso de cada modalidad, además de verificar si las observaciones derivadas del audio y del video mantienen un comportamiento consistente ante diferentes condiciones del entorno, estilos de comunicación, capacidades expresivas y recursos tecnológicos disponibles. Asimismo, la

participación de docentes o tutores permitiría contextualizar adecuadamente la retroalimentación, explicar sus alcances y limitaciones, acompañar la interpretación de los resultados y evitar que las puntuaciones u observaciones automatizadas sean consideradas evaluaciones profesionales definitivas o conclusiones objetivas sobre las capacidades de los participantes.

Referencias

Akdur, D. (2022). Analysis of Software Engineering Skills Gap in the Industry. *ACM Transactions on Computing Education*, *23*(1), 1–28. https://doi.org/10.1145/3567837

Arksey, H., & O'Malley, L. (2005). Scoping Studies: Towards a Methodological Framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Baltes, S., & Ralph, P. (2022). Sampling in Software Engineering Research: A Critical Review and Guidelines. *Empirical Software Engineering*, *27*(4). https://doi.org/10.1007/s10664-021-10072-8

Barr, M., & Nabi, S. W. (2022). The Development of Students' Employability Skills on a Work-Based Software Engineering Degree Programme. *2022 IEEE Frontiers in Education Conference (FIE)*. https://doi.org/10.1109/FIE56618.2022.9962611

Bass, L., Clements, P., & Kazman, R. (2021). *Software Architecture in Practice* (4th ed.). Addison-Wesley Professional.

Behroozi, M., Brown, C., & Parnin, C. (2022). Asynchronous Technical Interviews: Reducing the Effect of Supervised Think-Aloud on Communication Ability. *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 294–305. https://doi.org/10.1145/3540250.3549168

Behroozi, M., Shirolkar, S., Barik, T., & Parnin, C. (2020). Does Stress Impact Technical Interview Performance? *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. https://doi.org/10.1145/3368089.3409712

Bevara, R. V. K., Mannuru, N. R., Karedla, S. P., Lund, B., Xiao, T., Pasem, H., Dronavalli, S. C., & Rupeshkumar, S. (2025). Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching. *Electronics*, *14*(4). https://doi.org/10.3390/electronics14040794

Boland, M. R., Rusanov, A., So, Y., Lopez-Jimenez, C., Busacca, L., Steinman, R. C., Bakken, S., Bigger, J. T., & Weng, C. (2014). From Expert-Derived User Needs to User-Perceived Ease of Use and Usefulness: A Two-Phase Mixed-Methods Evaluation Framework. *Journal of Biomedical Informatics*, *52*, 141–150. https://doi.org/10.1016/j.jbi.2013.12.004

Codina, L. (2020). Revisiones bibliográficas sistematizadas en Ciencias Humanas y Sociales. 1: Fundamentos. In *Methodos. Anuario de Métodos de Investigación en Comunicación Social, 1* (pp. 50–60). DigiDoc-Universitat Pompeu Fabra. https://doi.org/10.31009/methodos.2020.i01.05

Daryanto, T., Stil, S., Ding, X., Manesh, D., Lee, S. W., Lee, T., Lunn, S., Rodriguez, S., Brown, C., & Rho, E. (2025). *Designing Conversational AI to Support Think-Aloud Practice in Technical Interview Preparation for CS Students*. arXiv. https://doi.org/10.48550/arXiv.2507.14418

Diniz, W., Valença, M., França, C., Santos, A., & Pincovsky, M. (2024). The Skill Gap in Software Industry: A Mapping Study. *Proceedings of the 38th Brazilian Symposium on Software Engineering (SBES 2024)*, 192–200. https://doi.org/10.5753/sbes.2024.3351

Emirtekin, E. (2025). Large Language Model-Powered Automated Assessment: A Systematic Review. *Applied Sciences*, *15*(10). https://doi.org/10.3390/app15105683

Gomez, N., Batham, S. S., Volonte, M., & Do, T. D. (2025). *Virtual Interviewers, Real Results: Exploring AI-Driven Mock Technical Interviews on Student Readiness and Confidence*.

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Kapoor, A., & Gardner-McCune, C. (2025). Retrospective Evaluation of Technical Interview Preparation Activities Offered in a Data Structures and Algorithms Course. *Proceedings of the ACM Conference on Innovation and Technology in Computer Science Education*. https://doi.org/10.1145/3724363.3729102

Kapoor, A., Panchal, S., & Gardner-McCune, C. (2023). Implementation and Evaluation of Technical Interview Preparation Activities in a Data Structures and Algorithms Course. *Proceedings of the 54th ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3545945.3569755

Kassab, K., Kashevnik, A., & Shoshina, I. (2026). Multi-Modal Method for Candidate Interview Assessment Based on Computer Vision and Large Language Models. *Big Data and Cognitive Computing*, *10*(4). https://doi.org/10.3390/bdcc10040106

Ko, A. J., LaToza, T. D., & Burnett, M. M. (2015). A Practical Guide to Controlled Experiments of Software Engineering Tools with Human Participants. *Empirical Software Engineering*, *20*(1), 110–141. https://doi.org/10.1007/s10664-013-9279-3

Laryeafio, M. N., & Ogbewe, O. C. (2023). Ethical Consideration Dilemma: Systematic Review of Ethics in Qualitative Data Collection Through Interviews. *Journal of Ethics in Entrepreneurship and Technology*, *3*(2), 94–110. https://doi.org/10.1108/JEET-09-2022-0014

Lee, D.-K., & Joe, I. (2025). A GPT-Based Code Review System With Accurate Feedback for Programming Education. *IEEE Access*, *13*, 105724–105737. https://doi.org/10.1109/ACCESS.2025.3581139

Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Lunn, S. J., & Brewster, V. (2023). Virtual Interview Training: Perceptions and Performance Using Digital Hiring Managers. *Proceedings of the 2023 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2--44602

Lunn, S. J., Ross, M., Hazari, Z., Weiss, M. A., Georgiopoulos, M., & Christensen, K. J. (2021). The Impact of Technical Interviews, and Other Professional and Cultural Experiences on Students' Computing Identity. *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, 415–421. https://doi.org/10.1145/3430665.3456362

Lunn, S. J., Zerbe, E., & Ross, M. S. (2022). Need for Change: How Interview Preparation and the Hiring Process in Computing Can Be Made More Equitable. *2022 CoNECD (Collaborative Network for Engineering & Computing Diversity)*. https://doi.org/10.18260/1-2--39131

Mawalim, C. O., Leong, C. W., & Okada, S. (2026). Privacy-Aware Speaker Trait and Multimodal Features Relationship Analysis in Job Interviews. *Scientific Reports*, *16*. https://doi.org/10.1038/s41598-026-39322-9

Mejias, M., Vargas, Z., Edwards, W., Washington, G., Burge, L., Wilson, D.-M., & Kouaho, L.-M. (2023). Equity in the Preparation of Students for Software Engineering Coding Interviews: ChatGPT as a Mock Interviewer. *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, 1016–1020. https://doi.org/10.1109/CSCE60160.2023.00169

Mušac, P. L., Nakić, J., & Kržić, A. S. (2026). Automatic Assessment Tools for Grading Coding Assignments: A Systematic Literature Review. *Applied Sciences*, *16*(11). https://doi.org/10.3390/app16115658

Newman, S. (2021). *Building Microservices: Designing Fine-Grained Systems* (2nd ed.). O'Reilly Media.

Nguyen, T. T. H., Nguyen, T. D. Q., Cao, H. L., Tran, T. C. T., Truong, T. C. M., & Cao, H. (2025). *SimInterview: Transforming Business Education through Large Language Model-Based Simulated Multilingual Interview Training System*. arXiv. https://doi.org/10.48550/arXiv.2508.11873

OECD. (2023). *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. OECD Publishing. https://doi.org/10.1787/08785bba-en

Okoli, C. (2015). A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems*, *37*(1), 879–910. https://doi.org/10.17705/1CAIS.03743

Putra, B., Azizah, K., Mawalim, C. O., Hanif, I. A., Sakti, S., Leong, C. W., & Okada, S. (2024). MAG-BERT-ARL for Fair Automated Video Interview Assessment. *IEEE Access*, *12*, 145188–145205. https://doi.org/10.1109/ACCESS.2024.3473314

Silva, P., & Costa, E. (2025). *Assessing Large Language Models for Automated Feedback Generation in Learning Programming Problem Solving*. arXiv. https://doi.org/10.48550/arXiv.2503.14630

Storey, M.-A., Hoda, R., Milani, A. M. P., & Baldassarre, M. T. (2025). Guiding Principles for Mixed Methods Research in Software Engineering. *Empirical Software Engineering*, *30*(5). https://doi.org/10.1007/s10664-025-10629-x

Sun, H., Lin, H., Yan, H., Zhu, C., Song, Y., Gao, X., Shang, S., & Yan, R. (2025). MockLLM: A Multi-Agent Behavior Collaboration Framework for Online Job Seeking and Recruiting. *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, 2714–2724. https://doi.org/10.1145/3711896.3737051

Taibi, D., Lenarduzzi, V., & Pahl, C. (2018). Architectural Patterns for Microservices: A Systematic Mapping Study. *Proceedings of the 8th International Conference on Cloud Computing and Services Science (CLOSER 2018)*, *1*, 221–232. https://doi.org/10.5220/0006798302210232

Torraco, R. J. (2016). Writing Integrative Reviews of the Literature: Methods and Purposes. *International Journal of Adult Vocational Education and Technology*, *7*(3), 62–70. https://doi.org/10.4018/IJAVET.2016070106

Vukadin, D., Kurdija, A. S., Delač, G., & Šilić, M. (2021). Information Extraction From Free-Form CV Documents in Multiple Languages. *IEEE Access*, *9*, 84559–84575. https://doi.org/10.1109/ACCESS.2021.3087913

Wilson, K., & Caliskan, A. (2024). Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, *7*(1), 1578–1590. https://doi.org/10.1609/aies.v7i1.31748

World Economic Forum. (2025). *The Future of Jobs Report 2025*. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025/