



**SISTEMA MULTI-AGENTE PARA LA PRESELECCIÓN DE CANDIDATOS EN
VACANTES PÚBLICAS DE EMPLEO UTILIZANDO INTELIGENCIA
ARTIFICIAL GENERATIVA**

Cristian Mauricio Blandón Londoño

Tesis de Maestría

Asesores

Claudia Patricia Álvarez Barrera

Juan David Martínez Vargas

UNIVERSIDAD EAFIT
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA
MEDELLÍN
2025

TABLA DE CONTENIDO

INTRODUCCIÓN	10
PLANTEAMIENTO DEL PROBLEMA	12
JUSTIFICACIÓN	16
OBJETIVO GENERAL	18
OBJETIVOS ESPECÍFICOS.....	18
MARCO CONCEPTUAL.....	19
1. Sistema Multi-agente (MAS).....	19
2. Matching semántico de candidatos y vacantes.....	20
3. Procesamiento de Lenguaje Natural (NLP) en RRHH	21
4. Modelos de Lenguaje de Gran Tamaño (LLM)	23
5. Evaluación de sistemas de IA en Recursos Humanos.....	25
6. Web Scraping en sistema de recursos humanos.....	26
DISEÑO METODOLÓGICO.....	28
1. Entendimiento del Negocio.....	28
2. Entendimiento de los Datos	29
2.1 Estructura de los descriptores de empleo	29
2.2 Estructura de un currículum.....	29
3 Modelado del Sistema.....	30
3.1. Preprocesamiento y vectorización.....	30
3.2. Representación Estructurada de la Información Extraída.....	31
3.4. Agente de Extracción de Datos	33
3.5. Agente de Extracción de Datos de Currículos (CV Parser)	35
3.6. Agente Evaluador.....	36
4. Herramientas Implementadas.....	36
5. EVALUACIÓN.....	38
5.1. Marco General de Evaluación.....	38
a) Fase de Calibración Técnica	38
5.2. Evaluación Automatizada (LLM as a Judge).....	39

5.3. Evaluación por Criterio Humano (Human as a Judge)	40
RESULTADOS.....	42
1. Resultados por Dimensión de Evaluación (LLM as a judge)	43
1.1. Precisión en la clasificación semántica	43
1.2. Distribución de Afinidades Asignadas.....	50
1.3. Métricas de evaluación por área.....	51
1.4. Consistencia Operativa.....	53
1.5. Consideraciones Éticas y Sesgo en el Proceso de Evaluación	55
1.5.1. Evidencia de la salida del sistema	55
2. Evaluación por criterio humano	58
2.1. Coincidencia Sistema-Humano	58
2.2. Evaluación Cualitativa (Escala 1-5).....	58
2.2.1. Comentarios cualitativos adicionales:.....	59
3. Interpretabilidad	59
3.1. Evaluación manual de las sugerencias hechas por el sistema.	59
3.2. Evidencia de la salida del sistema.....	60
CONCLUSIONES	63
TRABAJO FUTURO.....	65
REFERENCIAS.....	67
Apéndice A.....	72
Protocolo de etiquetado de relevancia para la evaluación de candidatos.....	72
1. Criterios de Relevancia para el Etiquetado.	73
2. Instrucciones para la Puntuación General de Relevancia del Candidato	75
3. Escala de puntuación detallada por criterio.	75
4. Ejemplo Práctico	76
5. Construcción del Ground-Truth	77

Lista de Figuras

FIGURA 1. PREPROCESAMIENTO Y VECTORIZACIÓN.....	31
FIGURA 2. REPRESENTACIÓN ESTRUCTURADA DE LA INFORMACIÓN EXTRAÍDA	32
FIGURA 3. ARQUITECTURA MULTI-AGENTE Y FUNCIONES ESPECÍFICAS	33
FIGURA 4. AGENTE DE EXTRACCIÓN DE DATOS	34
FIGURA 5. FORMATO DE LA INFORMACIÓN EXTRAÍDA	34
FIGURA 6. AGENTE DE EXTRACCIÓN DE DATOS DE CURRÍCULOS (CV PARSER)	35
FIGURA 7. FORMATO DE LA INFORMACIÓN EXTRAÍDA (CVS PARSER).....	36
FIGURA 8. EJEMPLO CURRÍCULO EVALUADO	45
FIGURA 9. EJEMPLO CURRÍCULO EVALUADO	46
FIGURA 10. DESCRIPCIÓN DE VACANTE	47
FIGURA 11. INTERPRETACIÓN REALIZADA POR EL SISTEMA (AFINIDAD ALTA)	48
FIGURA 12. EVALUACIÓN REALIZADA POR EL SISTEMA (AFINIDAD ALTA)	48
FIGURA 13. INTERPRETACIÓN REALIZADA POR EL SISTEMA (AFINIDAD MEDIA)	49
FIGURA 14. EVALUACIÓN REALIZADA POR EL SISTEMA (AFINIDAD MEDIA)	49
FIGURA 15. INTERPRETACIÓN REALIZADA POR EL SISTEMA (AFINIDAD BAJA)	50
FIGURA 16. EVALUACIÓN REALIZADA POR EL SISTEMA (AFINIDAD BAJA).....	50
FIGURA 17. TIEMPO DE EVALUACIÓN POR CANDIDATO	53
FIGURA 18. INTERPRETACIÓN DEL SISTEMA – CANDIDATO MASCULINO.....	56
FIGURA 19. INTERPRETACIÓN DEL SISTEMA – CANDIDATA FEMENINA	57
FIGURA 20. INTERPRETACIÓN DEL SISTEMA (SECTOR FINANCIERO).....	61
FIGURA 21. INTERPRETACIÓN DEL SISTEMA (SECTOR TECNOLOGÍA)	62

Lista de Tablas

TABLA 1. DIMENSIONES A EVALUAR.....	38
-------------------------------------	----

TABLA 2. DISTRIBUCIÓN DE AFINIDADES ASIGNADAS	51
TABLA 3. MÉTRICAS DE EVALUACIÓN POR ÁREA	52
TABLA 4. MATRIZ DE EVALUACIÓN DEL SISTEMA FRENTE AL GROUND-TRUTH	52
TABLA 5. COMPARACIÓN DEL TIEMPO PROMEDIO DE EVALUACIÓN POR PERFIL	54
TABLA 6. CONSISTENCIA OPERATIVA	54
TABLA 7. EVALUACIÓN CRITERIO HUMANO. ANÁLISIS DE CONCORDANCIA.....	58
TABLA 8. CRITERIO EVALUACIÓN	58
TABLA 9. INTERPRETABILIDAD	60

RESUMEN

Históricamente, los procesos de selección se han realizado de forma manual. En ellos, los candidatos atraviesan por una serie de filtros que varían de acuerdo con los requerimientos de la vacante. Al no ser procesos estándar, se puede prolongar su duración y, por consiguiente, incrementar la tasa de vacantes no cubiertas, afectando negativamente la competitividad organizacional.

Encontrar el candidato ideal para una vacante de trabajo es una tarea que requiere tiempo y recursos. En la actualidad, éste es el desafío al que se enfrentan diariamente las organizaciones del sector de recursos humanos donde cada día invertido en la búsqueda de talento adecuado para sus vacantes se traduce en costos, por lo que los retrasos en estas tareas afectan directamente la consecución de objetivos estratégicos organizacionales.

A pesar del uso creciente de tecnologías para seguimiento de candidatos como los Sistema de Seguimiento de Candidatos (*Applicant Tracking System or ATS*), dichas herramientas, tienden a tener limitaciones semánticas y no se adaptan bien a contextos locales, especialmente en países como Colombia, donde gran parte de la población se encuentra en la informalidad, lo cual dificulta no solo una evaluación objetiva de la idoneidad de los candidatos, sino también la introducción de sesgos evaluativos.

Estudios recientes han comenzado a abordar la integración de arquitecturas multi-agente con modelos de lenguaje de gran tamaño para procesos de pre-selección automatizados. Este proyecto propone la implementación de un sistema multi-agente para la evaluación de candidatos, que en combinación con técnicas de procesamiento de lenguaje natural (NLP) y modelos de lenguaje de gran tamaño (LLM), permita analizar la información de los postulantes y ofertas laborales que ayude a los profesionales de recursos humanos a identificar la idoneidad de un candidato a una vacante específica.

El sistema por implementar estará diseñado para optimizar los procesos de evaluación en contratación y se estima que, mediante esta implementación, se pueda disminuir el tiempo promedio evaluación y selección.

Esta implementación, pretende también disminuir las operaciones manuales y mitigar los sesgos en la evaluación, contribuyendo al desarrollo sostenible del capital humano. Se proyecta que el sistema incremente la eficiencia en dichos procesos y facilite una mayor alineación entre las competencias requeridas por las empresas y las habilidades disponibles en el mercado, beneficiando tanto a empleadores como a postulantes. Indirectamente se ayudará al cierre de la brecha entre la oferta y la demanda de habilidades en el mercado laboral colombiano, particularmente en un contexto marcado por altos índices de informalidad, al establecer criterios de evaluación imparciales basados exclusivamente en las competencias que indiquen los postulantes.

ABSTRACT

Historically, recruitment processes have been carried out manually. In such processes, candidates go through a series of filters that vary depending on the specific requirements of each vacancy. As these procedures are not standardized, their duration can be extended, leading to an increase in unfilled positions and, consequently, negatively impacting organizational competitiveness.

Identifying the ideal candidate for a job vacancy is a task that demands both time and resources. Today, this represents a significant challenge for organizations within the Human Resources sector, where each day spent searching for the right talent translates into operational costs. As a result, delays in recruitment activities directly affect the achievement of strategic organizational goals.

Despite the growing adoption of Applicant Tracking Systems (ATS), these tools often face semantic limitations and do not easily adapt to local contexts—especially in countries like Colombia, where a significant portion of the population is employed informally. This reality hinders not only the objective assessment of candidate suitability but also increases the likelihood of evaluative biases.

Recent studies have begun exploring the integration of multi-agent architectures with Large Language Models (LLMs) to automate pre-screening processes. In line with this, the present project proposes the implementation of a multi-agent system for candidate evaluation. By combining Natural Language Processing (NLP) techniques with LLMs, the system aims to analyze applicant and job posting data to support human resources professionals in determining candidate-job fit.

The system will be designed to optimize evaluation procedures in recruitment, with the goal of reducing the average time required for candidate assessment and selection. Furthermore, implementation seeks to minimize manual operations and mitigate bias in the evaluation process, thereby contributing to the sustainable development of human capital.

It is anticipated that this solution will increase the efficiency of recruitment workflows and promote greater alignment between the skills demanded by employers and those offered in the labor market. This, in turn, is expected to benefit both employers and job seekers, and indirectly support efforts to bridge the skills gap in the Colombian labor market—particularly in a context characterized by high levels of informality—by establishing fair and competency-based evaluation criteria.

Keywords: Artificial Intelligence, Human Resources, Talent Acquisition, Automated Evaluation, Language Models, Natural Language Processing, Multi-Agent System, Skills Gap, Colombian Labor Market.

INTRODUCCIÓN

En la era digital actual, el éxito empresarial depende en gran medida de la capacidad de atracción y retención de talento, conseguir a la persona adecuada para un cargo es esencial para obtener una ventaja competitiva, lo que convierte los procesos de selección y contratación en un pilar estratégico dentro de toda organización (Cable & Judge, 1996). Esto resalta la importancia de evaluar cuidadosamente las habilidades de los candidatos durante el proceso de selección y así garantizar un buen ajuste a los requisitos del cargo (Deniz, Noyan, & Ertosun, 2015).

Los avances en la analítica organizacional basada en redes (ONA), la Inteligencia Artificial (IA) y su integración en la gestión del talento humano, se ha desplegado un abanico de posibilidades en la optimización de procesos, entre ellos, los procesos de reclutamiento, mediante la introducción de técnicas de Procesamiento de Lenguaje Natural (NLP) y Modelos de Lenguaje de Gran Tamaño (LLM) ya que han demostrado ser una herramienta eficiente para la automatización de tareas repetitivas (Arenas, 2023).

El proceso de selección de personal tradicional está compuesto por diferentes etapas, desde la publicación de una vacante hasta la selección del candidato, sin embargo, estos procesos no son estándar, dependen de cada compañía y su desarrollo suele complejizarse debido a la necesidad de garantizar un proceso eficiente y eficaz (Bolaños-Cerón, 2020), como consecuencia, no existe un tiempo específico de duración, pero sí un promedio que oscila entre 60 y 65 días hábiles (iDhunt RH, 2023) durante los cuales, puede verse afectada la capacidad de respuesta de las empresas ante las demandas del mercado y puede derivar en vacantes no cubiertas; además, la selección manual introduce sesgos por parte de evaluador, lo que limita la identificación del mejor talento disponible.

La creciente disponibilidad de datos abiertos provenientes de portales de empleo, junto con la capacidad de la IA para analizarlos, plantea una oportunidad para mejorar estos procesos mediante sistemas automatizados de evaluación. Según estudios realizados los

algoritmos de matching reducen en un 40% el tiempo de contratación al analizar competencias clave, sin embargo, la efectividad depende de la calidad de los datos; la anterior afirmación, demuestra que, a pesar de las herramientas existentes, su capacidad para comparar habilidades en contextos específicos empresariales sigue siendo limitada y, en muchos casos, dichas herramientas suelen ser costosas (Nagarhalli et al., 2020).

Este trabajo propone el diseño e implementación de un sistema de matching basado en técnicas de Procesamiento de Lenguaje Natural (NLP), Modelos de Lenguaje de Gran Tamaño (LLM) y el uso de agentes virtuales (chatbots) con el objetivo de automatizar la evaluación de perfiles y facilitar la identificación del candidato ideal para optimizar el proceso de selección.

Se espera que la implementación de esta solución contribuya a mejorar la eficiencia operativa del proceso de selección y contratación, al analizar competencias clave de los candidatos, identificando de forma precisa los perfiles que cumplen con los requisitos definidos en las vacantes, facilitando así una mejor alineación entre la demanda del mercado laboral y las habilidades disponibles; lo cual, podría incidir positivamente en la reducción de la tasa de desempleo.

PLANTEAMIENTO DEL PROBLEMA

Estudios previos han señalado que las credenciales académicas no siempre reflejan las competencias que realmente demanda el sector productivo (Gontero & Albornoz, 2019). En el contexto empresarial actual, la identificación de talentos con alto potencial (HIPOs) representa un desafío crítico en la gestión de recursos humanos, debido a su impacto estratégico en la consecución de objetivos organizacionales.

En este contexto, los métodos tradicionales de selección dependen de la evaluación subjetiva de factores como habilidades de comunicación, trabajo en equipo y autoaprendizaje. Este enfoque, basado en la percepción de expertos, introduce sesgos que pueden afectar la selección de candidatos y pueden limitar la efectividad del proceso (Qin et al., 2023).

En países como Colombia, el desafío es aún mayor debido a los altos índices de informalidad laboral. Según el Departamento Administrativo Nacional de Estadística (DANE, 2023), para enero de 2023, el 58% de la población ocupada se encontraba en condiciones de informalidad, cifra que aún sigue siendo preocupante. De acuerdo con el estudio más reciente del Banco de la República (Otero-Cortés et al., 2025) en febrero del año 2025, más de la mitad de los trabajadores (56% a nivel nacional y 43% en zonas urbanas) aún se emplean en la informalidad.

Lo anterior, no solo restringe el acceso a empleos formales, sino que también aporta información clara de la proporción del talento humano está quedando por fuera de los canales formales de reclutamiento. Este desajuste de habilidades se reconoce como una de las causas principales de los persistentes niveles de desempleo, dificultando las labores de reclutamiento para cubrir las vacantes disponibles y agrava la dificultad de identificar y retener talento cualificado.

Por otro lado, los enfoques actuales de selección de personal presentan limitaciones importantes para abordar esta problemática, en general, las organizaciones han estado utilizando los llamados ATS Applicant Tracking Systems (sistemas de seguimiento de candidatos) u otros filtros automatizados, así como los procesos manuales tradicionales que hoy en día se continúan realizando de manera manual, lo que conlleva a una falta de flexibilidad y contexto en la evaluación.

Estos sistemas, al estar configurados con patrones lingüísticos específicos, pueden omitir información relevante cuando detectan sinónimos en un currículo que no coinciden con sus parámetros predefinidos. Esto puede llevar a que candidatos calificados sean descartados por diferencias menores en la terminología utilizada, lo que evidencia limitaciones en la precisión del proceso de selección debido a una falta de actualización semántica (Worki 360, 2025).

Adicionalmente, los procesos de selección tradicionalmente carecen de objetividad, ya que las decisiones a menudo se ven influenciadas por sesgos del reclutador o por información incompleta sobre el candidato. Esto se debe, en parte, a la falta de validación contextual previa al proceso y a la ausencia de retroalimentación durante el mismo. Por ejemplo, los sistemas ATS frecuentemente incurren en el llamado 'efecto agujero negro', en el que el postulante no recibe ninguna notificación sobre el estado de su postulación (Worki 360, 2025).

Esta situación representa una de las quejas más comunes por parte de los candidatos durante los procesos de selección. No solo genera una percepción de falta de transparencia por parte del evaluador, sino que también dificulta que el candidato pueda alinearse adecuadamente con los requerimientos del cargo.

En consecuencia, los métodos actuales (tanto automatizados básicos como humanos) no logran alinear de forma acertada y precisa las competencias de los aspirantes con las necesidades específicas de las vacantes, especialmente en un entorno laboral dinámico como el colombiano que exige adaptabilidad y conocimiento del contexto local.

Lo anterior refleja con exactitud la postura de (Qin et al., 2023) en la que se hace hincapié en que el 72% de los sistemas actuales no combinan habilidades, experiencia y competencias específicas de forma dinámica. Ante este panorama, al entender estas deficiencias, se plantea la necesidad de un sistema de evaluación de candidatos haciendo uso de herramientas de Inteligencia Artificial Generativa (Gen IA) que aproveche los avances de los modelos en Natural Language Processing (NLP) y aprendizaje automático para mejorar la correspondencia entre candidatos y vacantes.

Por tanto, se propone la implementación de un sistema multi-agente conformado por tres agentes especializados que colaboran de forma coordinada entre sí.

- Un agente de NLP/PLN que analiza las descripciones de ofertas laborales y extrae las competencias y requisitos clave.
- Un agente extractor que procesa hojas de vida en formato PDF y obtiene datos estructurados de la experiencia, formación y habilidades de los postulantes.
- Un agente evaluador que calcula la afinidad o correspondencia entre el perfil del candidato y la vacante.

La integración de técnicas modernas aporta una comprensión semántica profunda de los textos involucrados a diferencia de los filtros basados únicamente en palabras clave, los LLMs pueden interpretar el significado contextual de las habilidades y experiencias descritas en los CV y las ofertas, ofreciendo una evaluación más amplia de la idoneidad de un candidato (Recrew AI, 2025). Con lo anterior, se puede asumir que el sistema propuesto reconocerá correspondencias semánticas aunque los términos utilizados sean similares, superando las limitaciones de los ATS tradicionales.

La información extraída de descripciones y currículos se vectoriza (es decir, se convierte en representaciones numéricas de alta dimensión que capturan su contenido semántico) y se almacena en una base de datos vectorial (ChromaDB), de este modo, el agente evaluador puede realizar comparaciones eficientes entre vectores, midiendo la

similitud semántica entre el perfil del candidato y los requerimientos del puesto en un espacio vectorial común.

Tecnologías de IA generativa recientes han demostrado la capacidad de analizar masivamente currículos y descripciones para hallar candidatos adecuados de forma más objetiva y rápida que los métodos tradicionales (Bravent, 2024), lo que respalda la pertinencia de aplicar NLP y LLMs en este dominio.

Finalmente, es importante aclarar que el alcance del proyecto se limita a un prototipo funcional y no a una prueba de concepto ni a un sistema desplegado en producción para validar la idea. El prototipo funcional propuesto implica la implementación e integración de los componentes clave (los tres agentes y la base de datos vectorial) operando con datos reales o simulados bajo condiciones controladas, esto permite evaluar empíricamente la eficacia del sistema multi-agente en mejorar el emparejamiento candidato-vacante, generando resultados concretos que pueden ser analizados. Dicho prototipo provee un entorno experimental tangible para medir el desempeño del modelo y detectar desafíos prácticos (por ejemplo, calidad de la extracción de datos de CV, relevancia de las recomendaciones de afinidad, tiempos de respuesta en consultas a ChromaDB, etc.).

Asimismo, acotar el proyecto a un prototipo evita las complejidades adicionales de un sistema en producción (como escalabilidad a nivel país, consideraciones de UX masivo, integraciones corporativas o cumplimiento normativo), las cuales no están contemplados en los objetivos de esta fase investigativa. En esta etapa inicial, no se busca desplegar una plataforma comercial completamente operativa, sino demostrar la viabilidad y ventajas del enfoque utilizado, de este modo, el trabajo se alinea con las normas de rigor científico: se explora una solución innovadora al problema identificado, se construye un artefacto funcional para probar sus capacidades en condiciones reales o simuladas, y se obtienen evidencias y lecciones que servirán de base para desarrollos futuros más amplios y eventualmente su transferencia al ámbito productivo.

JUSTIFICACIÓN

La brecha entre la oferta y la demanda de habilidades en el mercado laboral constituye una problemática que impacta tanto a los profesionales como al sector empresarial. Según el Departamento Administrativo Nacional de Estadística (DANE, 2023), gran parte de la población económicamente activa se encuentra en condiciones de informalidad, lo cual dificulta la identificación de talento idóneo para cubrir vacantes estratégicas. Este desajuste limita las oportunidades laborales para la población en general y, al mismo tiempo, representa un obstáculo para la competitividad y sostenibilidad de las organizaciones. En consecuencia, resulta necesario optimizar los procesos de selección mediante el aprovechamiento de tecnologías emergentes, con el fin de favorecer una economía más inclusiva y eficiente.

Los procesos de selección en el área de recursos humanos han sido tradicionalmente manuales, caracterizados por su extensión y por la presencia de sesgos asociados al juicio subjetivo de los evaluadores. Ante esta situación, el presente proyecto plantea la implementación de un sistema multi-agente que reduzca el sesgo evaluativo y aumente la eficacia y confiabilidad de la preselección de candidatos. La propuesta se fundamenta en técnicas de Natural Language Processing (NLP), Large Language Models (LLMs) e inteligencia artificial generativa, orientadas a redefinir los procesos tradicionales y a alinearlos con las necesidades reales del mercado laboral.

El sistema se soportará en tecnologías de pago y de código abierto, y utilizará datos provenientes de fuentes públicas como portales de empleo y currículos de candidatos. El preprocesamiento y la extracción de datos se realizarán mediante web scraping, mientras que los agentes, apoyados en técnicas de Prompt Engineering, transformarán la información no estructurada en datos estructurados y de calidad para los evaluadores. Este enfoque favorece la implementación del sistema sin incurrir en altos costos de infraestructura y, gracias a su diseño multi-agente, garantiza escalabilidad y adaptabilidad a distintos sectores económicos.

La arquitectura distribuida propuesta, potenciada por modelos LLM y capacidades de razonamiento contextual, representa un avance frente a los enfoques tradicionales basados

en Graph Convolutional Networks (GCN), Convolutional Neural Networks (CNN) y Long Short-Term Memory (LSTM). Aunque estos modelos han mostrado eficiencia en la extracción de características para la selección de candidatos, dependen de reglas estáticas y requieren reentrenamientos frecuentes debido a la naturaleza cambiante del mercado laboral. El enfoque planteado incorpora principios de inteligencia adaptativa, mejorando la precisión en la identificación de afinidades laborales y reduciendo los sesgos asociados a decisiones humanas poco informadas.

El proyecto aporta en dos dimensiones principales. En el plano teórico, contribuye al entendimiento del papel de la inteligencia artificial generativa en la gestión del talento, especialmente en contextos emergentes como el colombiano. En el plano práctico, ofrece una solución replicable y escalable que incrementa la eficiencia de los procesos de selección, reduce costos operativos y acorta los tiempos de evaluación, beneficiando tanto a las organizaciones como a los candidatos.

Finalmente, a nivel macroeconómico, el sistema contribuirá al cierre de brechas entre oferta y demanda de habilidades, favoreciendo el desarrollo del capital humano en un entorno económico dinámico y competitivo. De esta manera, la propuesta promueve la equidad, impulsa la competitividad empresarial y aporta al desarrollo sostenible de la fuerza laboral.

OBJETIVO GENERAL

Diseñar e implementar un sistema multi-agente basado en inteligencia artificial, utilizando técnicas de Procesamiento de Lenguaje Natural (NPL) y Modelos de Lenguaje de gran tamaño (LLM) que automatice la preselección de candidatos y los compare con vacantes extraídas de datos abiertos como portales de empleo.

OBJETIVOS ESPECÍFICOS

- Diseñar un agente que permita extraer y estructurar información relevante de los descriptores de vacantes de fuentes abiertas como portales de empleo, mediante el uso de técnicas de web scraping.
- Proponer un agente que permita extraer y estructurar información relevante de los currículos de los candidatos mediante el análisis textual de archivos.
- Implementar un agente virtual que, mediante análisis semántico, permita comparar la información obtenida de los candidatos y la comparare con una vacante de empleo específica, con el fin de identificar el grado de adecuación del postulante a la vacante.
- Determinar el grado de precisión con la cual el sistema realiza las recomendaciones de los candidatos a través de un modelo de evaluación cualitativa del sistema.

MARCO CONCEPTUAL

1. SISTEMA MULTI-AGENTE (MAS)

Un sistema multi-agente (MAS) consiste en múltiples agentes de inteligencia artificial (IA) que trabajan colectivamente para realizar tareas en nombre de un usuario u otro sistema, cada agente de un MAS tiene propiedades individuales, pero todos se comportan en colaboración para conseguir las propiedades globales deseadas. (IBM, 2024).

De acuerdo con la literatura, existen varios tipos de agentes, por un lado, encontramos los agentes reactivos que responden directamente a estímulos del entorno sin mantener un modelo interno complejo y, estos a su vez se caracterizan por ser simples y fáciles de entender, aunque tienden a depender de la arquitectura de agentes previamente establecidos (Carrascosa, 2003)

Por otro lado, encontramos los agentes deliberativos o cognitivos que poseen un modelo simbólico interno del mundo que les permite razonar, planificar acciones e interactuar con otros agentes antes de actuar (Carrascosa, 2003), Estos agentes cognitivos disponen de mayor poder deliberativo para enfrentar entornos complejos, pudiendo incluso incorporar comportamientos reactivos y asumir tareas complejas cuando se requiera lo cual se hace posible mediante una arquitectura adecuada para este tipo de agentes (Universidad Autónoma Metropolitana, Unidad Azcapotzalco, s. f.).

En sistemas multi-agente modernos, es común observar implementaciones de arquitecturas híbridas que combinan el uso de agentes reactivos con agentes deliberativos que permiten aprovechar las ventajas que ambos tipos de agentes ofrecen.

Los sistemas multi-agente pueden resolver y actuar en un amplio espectro de tareas complejas del mundo real, algunos ejemplos de dominios aplicables incluyen desde sistemas de transporte hasta sistemas de defensa (IBM, 2024), por lo que la justificación de su uso en

sistemas de matching de candidatos para el área de RRHH, radica en la descomposición de tareas complejas propias de los procesos de selección en agentes especializados, como bien lo pueden ser un agente para analizar las ofertas laborales, otro para extraer la información de los currículos y otro para realizar el perfilamiento y emparejamiento de los candidatos. Esta sinergia de cooperación distribuida aumenta la escalabilidad y adaptabilidad del sistema al dividir el proceso entre agentes, lo cual, facilita la incorporación de nuevos módulos, teniendo como resultando decisiones más precisas. Estudios recientes incluso proponen el uso de arquitecturas multi-agente para asegurar la equidad y eficacia en reclutamiento.

Un MAS bien diseñado puede fomentar una mejor toma de decisiones, haciéndolas más inclusivas, lo que promueve la reducción de sesgos evaluativos discriminatorios en la toma de decisiones de los procesos de selección, esto a su vez, garantiza el cumplimiento de normatividad legal y mejora la gestión de los procesos de contratación (Fernández & Fernández, 2019).

2. MATCHING SEMÁNTICO DE CANDIDATOS Y VACANTES

El emparejamiento semántico entre perfiles de candidatos y ofertas de trabajo constituye el talón de Aquiles de los procesos de selección automatizados, ya que este proceso implica superar desafíos que van más allá de la simple coincidencia de términos. Consta de dos actividades principales: por un lado, evaluar la afinidad entre las competencias y la experiencia de los candidatos extraídas de su currículum; por otro, analizar los requisitos de una vacante más allá de coincidencias exactas de palabras clave. A diferencia de los métodos tradicionales de matching, que se basan en algoritmos con reglas predefinidas o en la búsqueda de palabras clave, este enfoque busca superar las limitaciones para captar sinónimos o similitudes en contextos dinámicos (Ugale et al., 2025).

En este contexto, los modelos Zero-Shot pueden mejorar la eficiencia del emparejamiento entre candidatos y vacantes al inferir afinidades semánticas sin necesidad de

etiquetado previo de datos (Kurek et al., 2024); estos modelos representan una evolución natural frente a los sistemas tradicionales, ya que permiten detectar relaciones entre perfiles y vacantes sin requerir un entrenamiento supervisado específico para cada tipo de puesto. Este enfoque se aplica en métodos modernos que, apoyados en técnicas de Procesamiento de Lenguaje Natural (NLP), utilizan representaciones semánticas de texto (embeddings) para medir la similitud de significado entre descripciones de empleo y currículos. Para ello, se emplean técnicas de representación vectorial como word embeddings (Word2Vec) y modelos de lenguaje profundo (como BERT), los cuales proyectan la información en un espacio vectorial común; en este espacio, la similitud de coseno entre vectores permite cuantificar el grado de correspondencia semántica, capturando de forma eficiente relaciones contextuales, lo que permite superar las limitaciones de los sistemas basados en coincidencia de términos.

3. PROCESAMIENTO DE LENGUAJE NATURAL (NLP) EN RRHH

El término Procesamiento del Lenguaje Natural (PLN) se utiliza para definir de forma amplia los procesos, herramientas y métodos que manipulan el lenguaje natural, como el habla o el texto normal generados por los seres humanos; en este sentido, las computadoras pueden ser utilizadas para entender y manipular textos en lenguaje natural con un objetivo específico (Luger, E. 2023), y reunir conocimientos sobre cómo los seres humanos entienden y usan el lenguaje, para modelar computacionalmente estos patrones (Chowdhury, G., 2003).

En el contexto de recursos humanos, el Procesamiento de Lenguaje Natural (NLP) aporta las bases fundamentales para analizar textos no estructurados como currículos y descriptores de empleos, dado que se trata de textos escritos de forma libre y sin una estructura específica, es necesario extraer automáticamente la información más relevante.

Una técnica clásica para ello es la representación de documentos mediante TF-IDF (Term Frequency-Inverse Document Frequency), que cuantifica la relevancia de cada término en un documento en relación con un corpus ya establecido. Esto permite identificar

las palabras clave de un currículum o un descriptor de empleo; sin embargo, existen limitaciones, TF-IDF no captura relaciones semánticas. Por ello, los primeros enfoques de NLP utilizaron modelos como Word2Vec, que aprende vectores para cada palabra de modo que palabras contextualmente similares en grandes corpus queden cercanas en el espacio vectorial; así, términos relacionados con una misma competencia tienden a ubicarse próximos entre sí, mejorando la detección de afinidades aun cuando difieran las palabras exactas utilizadas por candidatos y empleadores.

Estudios recientes han destacado los modelos de lenguaje profundo o transformers como BERT (Bidirectional Encoder Representations from Transformers), que permiten obtener representaciones contextuales de oraciones completas, entendiendo cada palabra según en el contexto que aparecen; esta comprensión profunda del contexto habilita comparaciones semánticas mucho más precisas entre currículos y descriptores de empleo.

De hecho, investigaciones han demostrado que BERT y modelos similares superan sustancialmente a técnicas previas (como bag of words con TF-IDF o embeddings promedio) en la tarea de matching de perfiles, logrando mejoras notables en las métricas de exactitud y cobertura (Ugale, 2025).

Por otro lado, otra técnica que se destaca en NLP para el área de RRHH es el Reconocimiento de Entidades Nombradas (NER), empleada para extraer información de currículos, ya que este tiene la capacidad de identificar automáticamente elementos dentro del texto como nombres de personas, títulos de cargos, nombres de empresas, títulos académicos o skills. Este enfoque, integrado en un sistema multi-agente, se podría estructurar de forma tal que, un agente podría dedicarse a la extracción de estos atributos clave como habilidades, años de experiencia, titulaciones y otros rasgos de candidatos, y luego utilizar modelos de aprendizaje profundo para comparar estos atributos estructurados contra los requisitos de la oferta laboral (Ugale, 2025).

Finalmente, este enfoque híbrido podría mejorar la interpretabilidad al explicitar qué criterios del perfil coinciden sin renunciar a las actuales capacidades predictivas de los modelos semánticos.

4. MODELOS DE LENGUAJE DE GRAN TAMAÑO (LLM)

Los modelos de lenguaje de grandes (LLMs) se posicionan como el avance más reciente en NLP y han estado transformando el análisis semántico, estos modelos son entrenados con grandes volúmenes de datos textuales, lo que los provee de una capacidad abismal para entender y generar texto como lenguaje natural.

Esta capacidad trae consigo ventajas importantes frente a los modelos tradicionales ya que pueden realizar razonamientos complejos e incluso, comprensión contextual profunda de textos. En este sentido, Lo et al. (2025) mencionan usos importantes en el ámbito de RRHH ya que un LLM puede leer la información completa de un descriptor de trabajo o un currículo y extraer la información relevante; uno de los aspectos más importantes a destacar, es su amplio conocimiento que abarca múltiples dominios, esto les permite adaptarse rápidamente a nuevos ámbitos, en este caso, la evaluación de perfiles laborales o descriptores de empleo (zero-shot or few-shot learning), suelen ser muy útiles dado que el mercado laboral es dinámico y constantemente surge terminología especializada.

Ahora bien, Bandara et al. (2025) señalan que, pese a las capacidades transformadoras de los modelos de lenguaje aplicadas a la gestión de talento, su uso conlleva riesgos de sesgo algorítmico y retos éticos ya que estos modelos al ser entrenados con grandes volúmenes de información, aprenden patrones discriminatorios que pueden estar presentes en datos históricos, un ejemplo de ellos sería, (si en los datos de entrenamiento hay menos mujeres en roles de ingeniería, el modelo podría tender a asociar a las mujeres con dichos roles), desventaja que se corrobora por parte de (Gayathri, K., & Jes Bella, Km. (n.d.), 2023) dónde

se hace hincapié en dichos aspectos referente a la transición de la Inteligencia Artificial en los procesos de recursos humanos

El siguiente caso documentado ilustra estos riesgos es el del sistema automatizado de contratación desarrollado por Amazon, el cual fue retirado tras evidenciarse que penalizaba sistemáticamente a candidatas mujeres; según Bandara et al. (2025), este comportamiento sesgado se originó en los datos de entrenamiento, que reflejaban una sobrerrepresentación histórica de hombres en puestos técnicos, lo que llevó al modelo a replicar patrones discriminatorios en lugar de corregirlos.

Este ejemplo evidencia cómo los LLM pueden reproducir prejuicios de género, favoreciendo a candidatos masculinos, y destaca la necesidad de implementar capacidades organizativas para gestionar los sesgos algorítmicos en entornos de recursos humanos (Bandara et al., 2025, p. 3). A partir de ello, las consideraciones éticas en este contexto incluyen entonces garantizar la igualdad de oportunidades (no excluir minorías o ciertos grupos por correlaciones espurias), respetar la privacidad de los datos personales de candidatos, y mantener la transparencia en las decisiones de la IA.

Un problema latente y bastante conocido respecto a los LLM es la opacidad algorítmica ya que los LLM operan como una “caja negra”, por lo que resulta difícil explicar por qué un candidato obtuvo cierta puntuación o recomendación. Esta falta de explicabilidad complejiza la confianza y la adopción responsable de estos modelos en RRHH.

Por esta razón, actualmente se están investigando enfoques para auditar y mitigar sesgos en LLMs, y para dotarlos de mayor interpretabilidad. Algunas propuestas involucran arquitecturas multi-agente donde un agente auditor monitorea las decisiones del LLM para asegurar que sean justificables y neutrales (Universidad Autónoma Metropolitana, Unidad Azcapotzalco, s. f.).

5. EVALUACIÓN DE SISTEMAS DE IA EN RECURSOS HUMANOS

La evaluación de los modelos de Inteligencia Artificial aplicada a RRHH es un punto importante para garantizar su eficacia y confiabilidad ya que estas herramientas influyen en decisiones importantes en los procesos de contratación, en términos cuantitativos se utilizan métricas clásicas de desempeño como precisión, recall y F1-score (Chavez, 2024.).

No obstante, si bien las métricas clásicas son útiles, en el contexto de RRHH no es suficiente tener sólo métricas estadísticas; por esta razón, se hace necesario abordar mecanismos de evaluación propios de la visión otorgada por expertos como lo son la evaluación cualitativa, en la cual los expertos en el área deben revisar muestras de las recomendaciones y descartes hechos por los modelos y verifican si tienen sentido en la práctica.

De hecho, estudios previos han incorporado la opinión de especialistas de RRHH para evaluar la precisión de las recomendaciones generadas por los modelos (Rosenberger et al., 2025), ya que la retroalimentación humana permite identificar falsos positivos/negativos relevantes que las métricas clásicas podrían ocultar. Por ejemplo, una IA podría obtener alta precisión numérica, pero estar sistemáticamente filtrando candidatos de cierto perfil minoritario; en estos casos, solo un experto humano sensible al contexto podría notar ese sesgo. Por tanto, criterios de equidad y transparencia forman parte de la evaluación en RRHH tanto como la precisión técnica.

En consecuencia, la evaluación integral de un sistema de IA para matching de candidatos en RRHH combina métricas cuantitativas con validaciones cualitativas realizadas por expertos, asegurando así no solo un alto rendimiento, sino también interpretabilidad y aceptación por parte de los usuarios finales.

6. WEB SCRAPING EN SISTEMA DE RECURSOS HUMANOS

El web scraping permite la recolección y procesamiento de grandes volúmenes de datos de fuentes en línea de forma rápida, estos datos pueden incluir números, texto e incluso una colección e imágenes o video (Marres & Weltevrede, 2013). En el contexto de Recursos Humanos (RRHH), especialmente en reclutamiento y selección, el uso del web scraping se justifica dada la enorme cantidad de datos disponibles en línea y la necesidad de procesarlos para tomar mejores decisiones en cuanto a talento se refiere. Por lo tanto, aprovechar la automatización mediante herramientas de scraping, resulta estratégico (Dilmegani, 2025).

El web scraping permite recopilar información de perfiles públicos, (p. ej. Nombres, contactos, habilidades, educación, experiencia) de diferentes fuentes en línea, lo que permite detectar candidatos que encajan con ciertos criterios buscados por los reclutadores, incluso si estos no han postulado a una oferta. Por ejemplo, un candidato puede no haber actualizado su currículum en portales de empleo, pero podría tener datos recientes en su perfil de LinkedIn, en su portafolio en línea o en blogs personales; mediante scraping, los reclutadores pueden descubrir y considerar a estos perfiles que de otro modo, hubiesen pasado desapercibidos (LinkedIn, s. f.).

Asimismo, el scraping en RR.HH. facilita la inteligencia de mercado laboral, los reclutadores pueden recolectar datos de salarios publicados, habilidades demandadas y descriptores de vacantes de diferentes sitios (LinkedIn, Indeed, Magneto, Elemplo, etc.), lo cual permite analizar tendencias y ajustar sus estrategias de reclutamiento; esto en parte ayuda a responder preguntas como: ¿qué perfiles y competencias buscan actualmente mis competidores?, ¿cuáles son las tendencias salariales para determinado rol o región?, ¿qué nuevas habilidades están emergiendo en las ofertas de empleo? Al disponer de datos de fuentes diversas, el área de RR.HH. puede orientar sus decisiones validadas por evidencia actual del mercado laboral.

Por otro lado, también se pueden ver beneficiados en cuanto a la automatización del matching entre vacantes disponibles y currículos de candidatos; diversos proyectos han demostrado cómo se puede extraer información de ofertas de empleo y compararla con perfiles de candidatos para encontrar el mejor talento.

Por ejemplo, Grine (2023) desarrolló una aplicación que mediante web scraping obtiene ofertas de trabajo publicadas en LinkedIn y luego realiza un análisis de texto mediante procesamiento de lenguaje natural sobre dichas ofertas y el currículum del candidato, identificando las habilidades clave requeridas y comparándolas con las del postulante, como resultado es un sistema que recomienda automáticamente las vacantes más relevantes para ese candidato, facilitando tanto la búsqueda de empleo como la labor del reclutador en filtrar candidatos.

Adicionalmente, otros de los beneficios que estos procesos aportan al sector de RRHH están relacionados con la recolección manual de información como lo son el volumen y la velocidad, ya que se pueden obtener millones de datos de la web de forma rápida, lo que permite a los reclutadores monitorear varias fuentes de información al mismo tiempo, tarea que no sería posible manualmente.

Otro de los beneficios a mencionar es la eficiencia y el ahorro de esfuerzo humano, ya que al automatizar la recolección de datos, libera al personal de tareas repetitivas como copiar y pegar información de los portales de empleo, lo que no solo ahorra tiempo, sino que disminuye los errores humanos al transcribir los datos (Reddy & Viswanath, 2022).

Finalmente, el web scraping integrado a sistemas de RRHH, puede incluirse como parte de la digitalización y el big data en gestión del talento, ya que es una técnica poderosa para transformar información no estructurada en conocimiento accionable, permitiendo investigaciones “en vivo” del entorno laboral y optimizando por consiguiente los procesos de reclutamiento tradicionales (Marres & Weltevrede, 2012)

DISEÑO METODOLÓGICO

Para guiar el desarrollo del prototipo del sistema de matching para candidatos, se adopta el enfoque CRISP-DM (*Cross-Industry Standard Process for Data Mining*) el cual permite abordar de manera integral todo el ciclo de desarrollo del prototipo el cual estará dado por las etapas desarrolladas a continuación:

1. ENTENDIMIENTO DEL NEGOCIO

El objetivo principal de esta fase fue traducir las necesidades de negocio en objetivos concretos, identificando los siguientes aspectos estratégicos:

- Reducir el tiempo de preselección de candidatos
- Incrementar la precisión del emparejamiento entre currículos y vacantes.
- Automatizar tareas repetitivas del área de recursos humanos.

Adquirir el entendimiento necesario del sistema a construir requirió de una revisión exhaustiva de literatura, lo que permitió formular las preguntas de negocio precisas

¿Cómo se debe realizar un proceso de web scraping?

¿Cómo diseñar sistema multi-agente para las áreas de RRHH que pueda extraer automáticamente datos de vacantes y currículos?

¿Cómo automatizamos las tareas repetitivas del área de RRHH?

¿Qué nivel de afinidad semántica es aceptable para preseleccionar un candidato?

Este análisis permitió definir la arquitectura funcional base del sistema y establecer el diseño preliminar de los agentes requeridos.

2. ENTENDIMIENTO DE LOS DATOS

Esta etapa se centró en comprender la estructura y la naturaleza tanto de los datos provenientes de fuentes de datos abiertas como currículos, las actividades principales de esta etapa se centran específicamente en:

2.1 Estructura de los descriptores de empleo

Se analizaron más de 30 vacantes provenientes de portales de empleo como (LinkedIn, Magneto, El Empleo), esto permitió identificar los siguientes aspectos en común:

- Título del cargo
- Funciones y responsabilidades
- Requisitos mínimos para postular
- Nivel educativo requerido
- Años de experiencia
- Habilidades técnicas y en algunas ocasiones, habilidades blandas.

Si bien la estructura presentada no corresponde a un estándar en general, esta varía de acuerdo con el objetivo organizacional y de la plataforma en la que se publique, dicha variabilidad semántica fue identificada también en estudios previos realizados por (Ao et al. (2022).

2.2 Estructura de un currículum.

Similar a los descriptores de ofertas de trabajo, la estructura de las hojas de vida y/o currículos, presentan estructuras variadas respecto a los descriptores de empleo, en su estructura típica, se definen secciones como:

- Información personal

- Educación
- Experiencia Laboral
- Habilidades Técnicas
- Certificaciones, Idiomas y en algunos casos, logros obtenidos.

La forma en la que se redacta y la profundidad de cada sección varía de acuerdo con el estilo de escritura de cada persona y no corresponde a un estándar establecido para la creación del documento, por lo que requiere de un tratamiento semántico flexible para su análisis.

3 MODELADO DEL SISTEMA

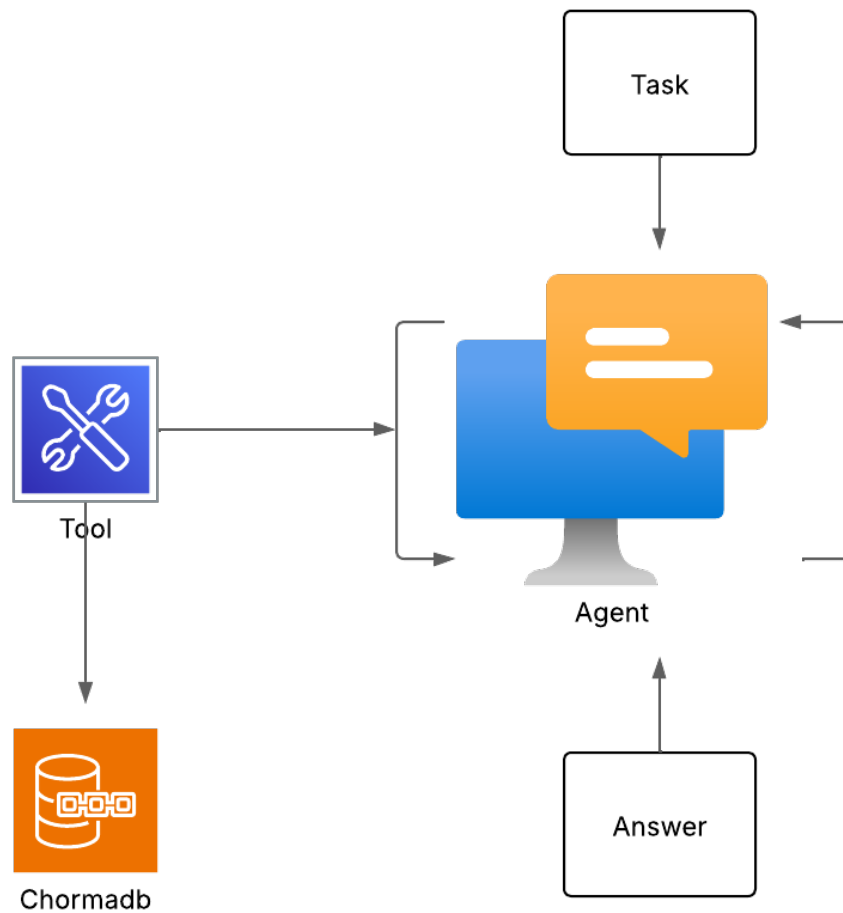
Con base en el entendimiento de negocio y el resultado del análisis de los datos, se diseñó una arquitectura basada en agentes que aprovecha las capacidades otorgadas por los LLM y técnicas de procesamiento de Lenguaje Natural (NLP) para realiza el emparejamiento de candidatos con vacantes, a diferencia de los sistemas tradicionales ATS que se basan en palabras clave (keywords).

Los LLM modernos como pueden interpretar el significado contextual de la información obtenida de descriptores de empleo y currículos, esto permite ofrecer una evaluación más amplia del ajuste de un candidato, lo que significa que el sistema puede entender correspondencias semánticas incluso si en el proceso de escritura se emplean términos diferentes, superando así las limitaciones gramaticales de coincidencias de términos. A continuación, se detalla cómo se estructuró el sistema para abordar la solución de este problema.

3.1. Preprocesamiento y vectorización

El objetivo de esta fase es transformar la información proveniente de currículos y vacantes en embeddings, proceso realizado por medio del modelo pre-entrenado gpt-4o de OpenAI, el cual convierte el texto en vectores numéricos de alta dimensión mediante la ejecución los agentes a diseñar, esto permite encapsular el contenido semántico y almacenarlos posteriormente en chormadb para su análisis.

Figura 1. Preprocesamiento y vectorización



3.2. Representación Estructurada de la Información Extraída

Para que la comparación sea efectiva, el sistema estandariza la información obtenida de vacantes y CV en un formato común (JSON) antes de vectorizarla. En concreto, tras el análisis realizado por los dos primeros agentes, los datos son presentados como un conjunto de atributos homogéneos que describen tanto la oferta de empleo como el perfil del candidato. Esta representación estructurada incluye campos clave como: título del puesto, habilidades requeridas en la vacante, años de experiencia solicitados, nivel educativo esperado, y sus contrapartes en el CV (puesto actual o pretendido, habilidades declaradas por el candidato, años de experiencia, educación alcanzada, etc.).

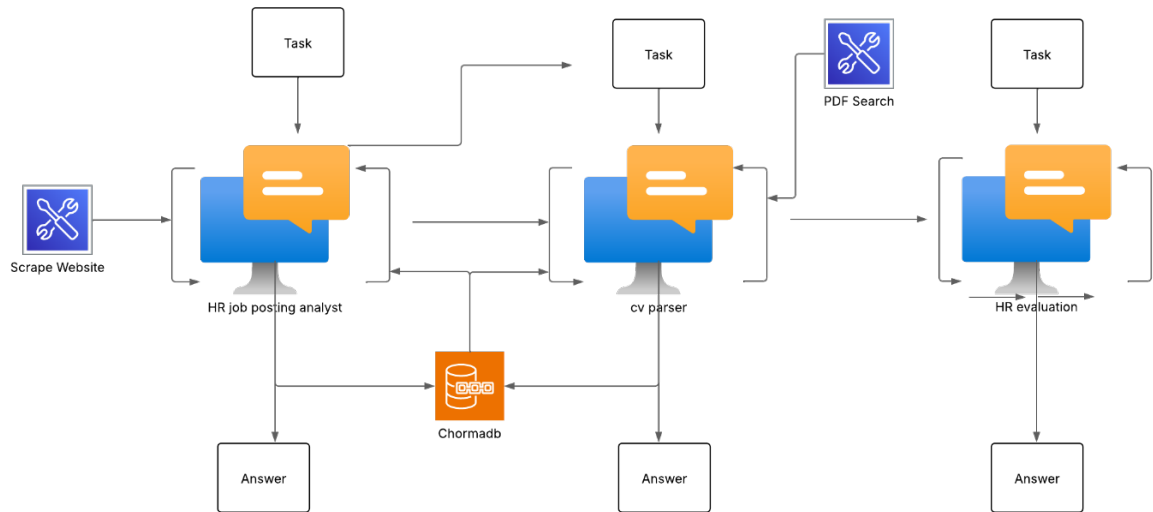
Figura 2. Representación Estructurada de la Información Extraída

```
## Final Answer:
{
  "Candidate Name": "Valentina Pérez",
  "Brief Bio": "Desarrollador Backend con conocimientos sólidos en APIs y bases de datos. / Backend developer with strong knowledge of APIs and databases.",
  "Skills": ["Python", "Node.js", "Java", "SQL", "Express.js"],
  "Relevant Experience": [
    {"Company": "CloudWare Inc.", "Duration": "2 years"},
    {"Company": "DataNova", "Duration": "1 year"}
  ],
  "Education": [
    {"Degree": "Engineering in Informatics", "Institution": "Pontificia Universidad Javeriana"}
  ],
  "Certifications": [
    "ISTQB Certified Tester",
    "Google Data Analytics Professional Certificate"
  ],
  "Contact Information": {
    "LinkedIn": "LinkedIn.com/in/valentinaperez",
    "GitHub": "github.com/valentina"
  },
  "Notes": "Email and phone number information is missing."
}
```

3.3.Arquitectura Multi-agente y funciones específicas

Con el entendimiento obtenido en la fase previa, se procede a diseñar bajo una arquitectura multi-agente, esto permite dividir el problema en tareas especializadas asignadas a distintos agentes virtuales. Cada agente hace uso de las capacidades ofrecidas por modelos LLM para cumplir su función, trabajando en conjunto para lograr la preselección de candidatos.

Figura 3. Arquitectura Multi-agente y funciones específicas

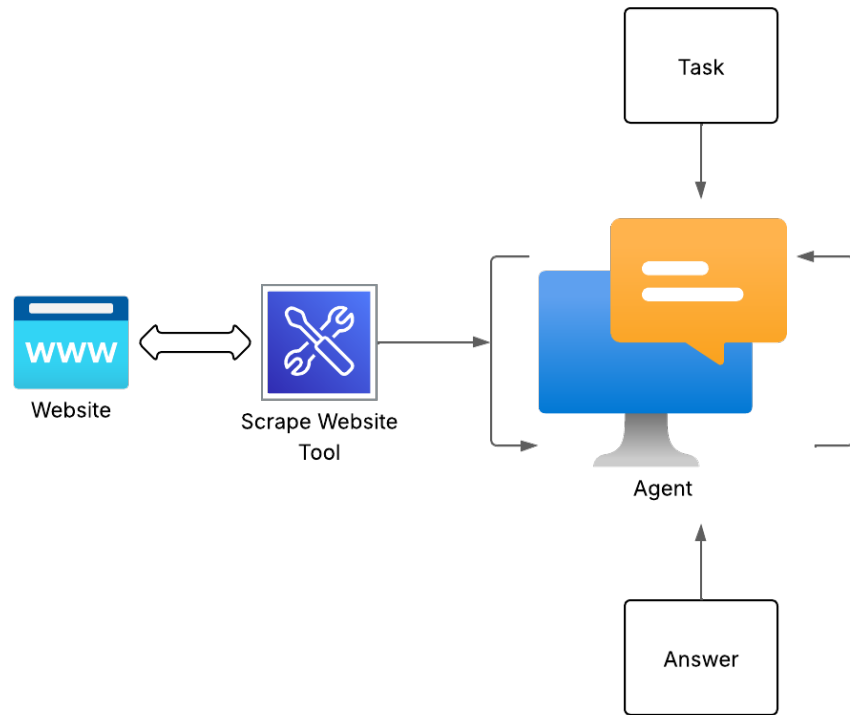


3.4. Agente de Extracción de Datos

El agente `hr_job_posting_analyst` realiza la extracción de contenido de los portales de empleo mediante web scraping, este agente tendrá como entrada la URL donde se encuentra el descriptor de empleo y mediante el uso de modelos de lenguaje (LLMs), se hará la clasificación de la información, este proceso incluye:

- Obtener el contenido HTML de la oferta
- Extracción de la información relevante mediante prompt engineering, en esta etapa del flujo, se le indica al agente que información se desea obtener. Por ejemplo: Título del cargo, funciones, habilidades requeridas, nivel de estudios, experiencia, etcétera.

Figura 4. Agente de Extracción de Datos



Cómo resultado del procedimiento, el agente entregará el texto extraído en formato JSON, siguiendo una estructura estandarizada que facilita su posterior análisis.

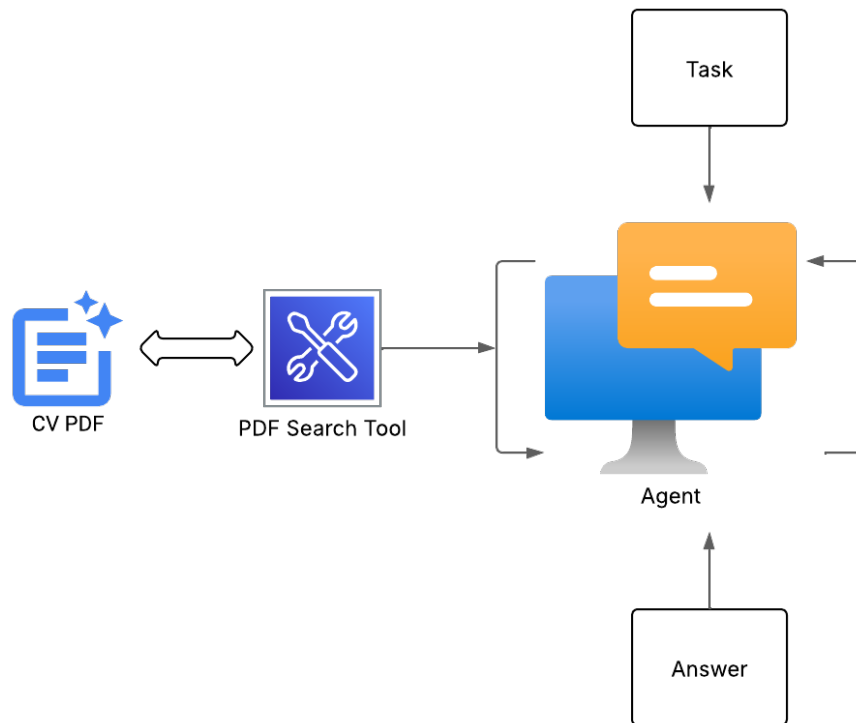
Figura 5. Formato de la información Extraída

```
{  
  "job_title": "...",  
  "company": "...",  
  "skills_required": ["skill1", "skill2", ...],  
  "experience_years": X,  
  "education": "...",  
  "job_type": "...",  
  "location": "...",  
  "salary_range": "..."  
}
```

3.5. Agente de Extracción de Datos de Currículos (CV Parser)

Los currículos por su parte, serán procesados por el agente cv_parser, los documentos serán recibidos en formato PDF y la información contenida en ellos se convierte a texto plano mediante el uso de librerías de Python, al igual que en la extracción de las vacantes, el proceso será guiado mediante prompt engineering, segmentando así el contenido extraído, de manera similar a la estructura de un currículum incluyendo secciones como nombre del candidato, descripción breve del perfil, educación, experiencia, habilidades, etcétera.

Figura 6. Agente de Extracción de Datos de Currículos (CV Parser)



Cómo resultado del procedimiento, el agente entregará el texto extraído en formato JSON, siguiendo una estructura estandarizada que facilita su posterior análisis.

Figura 7. Formato de la información Extraída (CVs parser)

```
{  
  "name": "...",  
  "email": "...",  
  "phone": "...",  
  "bio": "...",  
  "skills": ["skillA", "skillB", ...],  
  "experience": [{"role": "...", "years": Y}, ...],  
  "education": [...],  
  "certifications": [...]  
}
```

3.6. Agente Evaluador

Su tarea principal es calcular el grado de ajuste (afinidad) entre el perfil del candidato (información extraída del CV) y los requerimientos de una oferta de trabajo. Este agente consume la información procesada por los otros dos agentes y aplica métodos de comparación semántica, para determinar qué tan bien alineado está cada candidato con la vacante en cuestión, como resultado de la operación de evaluación, el agente otorgará un “score” que va de 0 a 100, generando finalmente un archivo en Markdown con las respectivas recomendaciones.

4. HERRAMIENTAS IMPLEMENTADAS

Durante el modelado se integraron diversas tecnologías y frameworks para construir la solución, a continuación, se detalla cada una de ellas:

- **CrewAI** para coordinación de agentes y el diseño del flujo de trabajo.
- **OpenAI API (GPT-4° / Ada-002)** para generación de embedding, asegurando representaciones vectoriales de alta calidad y comprensión contextual de texto necesarias en el análisis lingüístico de los descriptores de empleo y la interpretación de currículos en lenguaje natural

- **ChromaDB** para almacenamiento de vectores y recuperación de información por similitud.
- **Librerías de parsing PDF** para la extracción de texto de los currículos.
- **Prompt Engineering** para delimitar el comportamiento y extracción de información guiada.

Nativamente el framework soporta ChromaDB como base de datos vectorial, lo cual permitió almacenar y gestionar eficientemente la información vectorizada, ofreciendo funcionalidades de búsqueda por similitud muy rápidas (es decir, recuperar candidatos más similares a una descripción dada). Adicionalmente, se emplearon librerías de procesamiento de PDF y NLP, ofrecidas nativamente también por el framework en la implementación de los agentes de parsing. Cada agente produce resultados intermedios valiosos: el analista de vacantes genera una lista normalizada de requerimientos de la posición, el parser de CV produce un conjunto tabulado de las características del candidato, y el agente evaluador devuelve un score o clasificación de afinidad junto con un informe estructurado (JSON) de la comparación realizada. Estos artefactos intermedios no solo alimentan al siguiente paso en la cadena de procesamiento, sino que también permiten verificar y depurar el funcionamiento del sistema en cada etapa (por ejemplo, revisando si el agente extractor capturó correctamente las habilidades).

5. EVALUACIÓN

Es importante resaltar que en el desarrollo del prototipo no se emplea entrenamiento supervisado tradicional, dado que el enfoque es de comparación semántica. En lugar de ello, se llevó a cabo una "calibración" del sistema mediante Prompt Engineering y comparación de criterio experto para optimizar las recomendaciones realizadas por los agentes.

El proceso de evaluación se estructura en dos niveles principales:

- **Evaluación Automatizada (LLM as a Judge)**
- **Evaluación por Criterio Humano (Human as a Judge)**

5.1. Marco General de Evaluación

A continuación, se describen las dimensiones con las cuales se realizará la evaluación del sistema:

a) Fase de Calibración Técnica

Tabla 1. Dimensiones a Evaluar

Dimensión evaluada	Tipo	Mecanismo Aplicado
Precisión Semántica	Cualitativa	Emparejamiento entre perfiles y vacantes con score [0-100]
Consistencia Operativa	Cualitativa	Tiempo promedio de procesamiento por consulta
Calidad de Recomendación	Cualitativa	Comparación de recomendaciones con criterio humano
Interpretabilidad	Cualitativa	Revisión Manual de reportes Markdown generados

El objetivo de esta fase es Ajustar los prompts, formatos de entrada/salida y umbrales de decisión del sistema mediante los siguientes parámetros:

b) Estratificación de Datos de Prueba

La evaluación del sistema se deberá realizar con un conjunto de datos de Mínimo 50 currículos distribuidos de la siguiente manera para garantizar cobertura:

- 40% perfiles junior (0-2 años experiencia)
- 35% perfiles mid-level (3-7 años experiencia)
- 25% perfiles senior (8+ años experiencia)

Dataset de Vacantes (Mínimo 10 posiciones) distribuidas por industria (tech, finance, healthcare, etc.)

c) Umbrales de Afinidad Semántica

Los scores a continuación definen la afinidad semántica de la evaluación asignada a los candidatos por los agentes:

- **Alta afinidad:** 80-100 puntos
- **Media afinidad:** 50-79 puntos
- **Baja afinidad:** 0-49 puntos

5.2. Evaluación Automatizada (LLM as a Judge)

El objetivo de esta fase es validar la exactitud del emparejamiento semántico entre currículos de candidatos y descriptores de empleo de acuerdo con los parámetros indicados a continuación:

- **Generación de Recomendaciones:** El sistema asigna una escala de afinidad numérica de 0 a 100 para cada emparejamiento currículum-vacante
- **Clasificación por Umbrales:** Las recomendaciones se categorizan según los umbrales preestablecidos (alta, media, baja)

Métricas Resultantes:

- **Precision:** *Verdaderos positivos / Seleccionados como positivos*
- **Recall:** = *Verdaderos positivos / Relevantes reales*
- **F1 Score** = $2 \times (Precision \times Recall) / (Precision + Recall)$

5.3. Evaluación por Criterio Humano (Human as a Judge)

Para la evaluación por criterio humano, el “Ground-truth” se define como el consenso emitido por el panel de evaluadores humanos quienes analizaron las recomendaciones generadas por el sistema aplicando el protocolo de etiquetado definido en el [apéndice A](#), dónde se define el modelo de puntuación aplicado.

a) Proceso de evaluación:

- Se toman entre 3 a 5 muestras representativas de recomendaciones generadas por el agente evaluador
- **Validación contra Ground Truth:** Las evaluaciones realizadas por el sistema son evaluados por terceros (humano) no necesariamente expertos en el campo.

Cada recomendación se califica en una escala de 1 a 5 de acuerdo con los criterios a continuación:

- **Relevancia del candidato:** Qué tan apropiado es el candidato para el puesto
- **Coherencia de la justificación:** Lógica y consistencia del razonamiento del sistema
- **Exhaustividad de los criterios evaluados:** Completitud en el análisis realizado

b) Métricas resultantes:

- Coincidencia entre sistema y criterio humano: % de casos en que el sistema coincidió con la evaluación humana (alta, media o baja afinidad)
- Se considera válida si el sistema y el humano ubicaron la afinidad en el mismo rango de clasificación.

RESULTADOS

El sistema multi-agente fue evaluado bajo condiciones controladas mediante un conjunto de vacantes y currículos previamente seleccionados. Para este propósito, se diseñaron escenarios experimentales dirigidos a contrastar los criterios de calidad definidos en el apartado de validación. Con el objetivo de garantizar una presentación clara y sistemática de los hallazgos, los resultados se organizan en función de las dimensiones de evaluación establecidas a continuación.

1. Precisión en la clasificación semántica, para verificar la correcta asignación de afinidad entre perfiles y vacantes a partir de factores clave.
2. Distribución de afinidades asignadas, considerando áreas funcionales y niveles de seniority.
3. Métricas de rendimiento del sistema, utilizando indicadores como precisión, recall y F1-Score con base en ground-truth validada por expertos.
4. Coincidencia sistema–humano, a través del análisis de concordancia entre el sistema y evaluadores no expertos.
5. Evaluación cualitativa del sistema, mediante escalas de calificación en criterios clave como relevancia, coherencia y exhaustividad.
6. Interpretabilidad de las recomendaciones, enfocada en la claridad, completitud y utilidad de los reportes generados.

Estas dimensiones permiten una evaluación integral del sistema, no solo desde una perspectiva técnica, sino también considerando la experiencia de usuario y la aplicabilidad práctica en contextos reales de recursos humanos.

1. RESULTADOS POR DIMENSIÓN DE EVALUACIÓN (LLM AS A JUDGE)

1.1. Precisión en la clasificación semántica

Se realiza el proceso de evaluación bajo los parámetros descritos para las siguientes vacantes:

Áreas:

- Tecnología
- Finanzas
- Marketing

Job Positions:

- URL: [Analista de Data Testing QA Colombia](#)
- URL: [Analista Financiero](#)
- URL: [Marketing Specialist](#)

Perfiles (Junior, Mid-level, Senior)

Parámetros para el análisis (Factores clave):

- Coincidencia de habilidades
- Experiencia relevante
- Encaje cultural
- Potencial de crecimiento
- Formación académica
- Certificaciones
- Experiencia laboral, etc.
- Afinidad semántica con el contexto de la descripción del puesto de trabajo.
- Gap análisis

Salida esperada del sistema:

- El nombre de los candidatos

- Correo electrónico
- LinkedIn (en inglés)
- Github (en inglés)
- Bio
- Análisis de brechas
- Una puntuación entre 1 y 100.
- Un razonamiento detallado, teniendo en cuenta los factores clave

Ejemplo Evaluación realizada por el sistema:

La evaluación del sistema sobre los currículos procesados evidencia que los agentes diseñados cumplen de manera adecuada las funciones definidas. El proceso de extracción de información relevante se ejecuta con un nivel de precisión consistente, clasificando los perfiles en concordancia con el contexto de las vacantes y los umbrales de afinidad establecidos en la fase de calibración. Esta capacidad de análisis semántico permite superar las limitaciones de los enfoques tradicionales basados exclusivamente en coincidencia de palabras clave, aportando mayor solidez a la asignación de afinidades.

Currículo Evaluado: Sofía Martínez López

Área: Tecnología

Figura 8. Ejemplo Currículo evaluado

SOFÍA MARTÍNEZ LÓPEZ

Analista de Datos Junior - Fintech

sofia.martinez@email.com | +57 300 123 4567 | [linkedin.com/in/sofiamartin](https://www.linkedin.com/in/sofiamartin) | Medellín, Colombia

PERFIL PROFESIONAL

Analista de datos junior con 1.5 años de experiencia en el sector fintech. Especializada en análisis de riesgo crediticio y modelos predictivos. Dominio avanzado de Python, SQL y herramientas de visualización. Experiencia en startups de pagos digitales y microcréditos. Apasionada por la inclusión financiera y las tecnologías emergentes.

EXPERIENCIA LABORAL

Analista de Datos Junior | Nequi (Bancolombia)

Marzo 2023 - Presente

- Desarrollé modelos de scoring crediticio que mejoraron la precisión de aprobación en 15%
- Implementé dashboards en Power BI para monitoreo en tiempo real de transacciones
- Colaboré en la optimización de algoritmos de detección de fraude, reduciendo falsos positivos en 20%
- Analicé comportamiento de usuarios para identificar patrones de churn, logrando 18% de retención adicional

Pasante de Análisis de Riesgo | Bancolombia

Septiembre 2022 - Febrero 2023

- Apoyé en la construcción de modelos de provisión bajo metodología IFRS 9
- Desarrollé scripts en Python para automatización de reportes regulatorios
- Participé en la validación de modelos de riesgo de crédito y mercado
- Creé visualizaciones para presentaciones ejecutivas usando Tableau

Figura 9. Ejemplo Currículo Evaluado

EDUCACIÓN

Ingeniería Industrial | Universidad EAFIT

2018 - 2022

- Concentración en Finanzas Corporativas
- Proyecto de grado: "Modelo de Credit Scoring para Microcréditos usando Machine Learning"
- Promedio: 4.3/5.0

Certificaciones

- Google Data Analytics Certificate (2022)
- Financial Risk Management - Coursera (2023)
- Python for Financial Analysis - DataCamp (2023)

HABILIDADES TÉCNICAS

Programación: Python, R, SQL, JavaScript básico
Bases de Datos: PostgreSQL, MySQL, MongoDB
Visualización: Tableau, Power BI, Matplotlib, Seaborn
Machine Learning: Scikit-learn, Pandas, NumPy, XGBoost
Cloud: AWS basics, Google Cloud Platform
Otros: Git, Docker, Jupyter, Excel avanzado

HABILIDADES BLANDAS

- Análisis crítico y resolución de problemas
- Comunicación efectiva con equipos técnicos y comerciales
- Adaptabilidad a entornos dinámicos de startup
- Trabajo colaborativo en metodologías ágiles
- Orientación a resultados y mejora continua

PROYECTOS DESTACADOS

Modelo de Credit Scoring para Microempresarios

- Desarrollé un modelo predictivo usando Random Forest que incrementó la aprobación de microcréditos en 25%
- Implementé técnicas de feature engineering y validación cruzada
- Presenté resultados al comité de riesgo, logrando implementación en producción

Dashboard de Monitoreo de Pagos Digitales

- Creé un dashboard interactivo en Tableau para seguimiento de KPIs de pagos
- Integré múltiples fuentes de datos y automatice la actualización
- Reduje el tiempo de generación de reportes de 4 horas a 15 minutos

IDIOMAS

- Español: Nativo
- Inglés: Avanzado (C1) - TOEFL iBT: 102
- Portugués: Intermedio (B2)

INTERESES

Blockchain, DeFi, finanzas comportamentales, análisis de mercados emergentes, tecnologías de inclusión financiera

Información relevante de la vacante:

Área: Tecnología

Figura 10. Descripción de Vacante

CHOUCAIR Analista de Data Testing QA Colombia y Perú
Choucair

Aplicar

Hace 4 meses

1 año de experiencia, bachillerato completo hasta profesional

Salario a convenir

Palabras clave: Analista de Data Testing, QA Colombia, pruebas BI, software testing, SQL

Medellín - Bogotá, D.C. - Lima

Analista de Data Testing QA Colombia

Choucair, con más de 25 años de experiencia en el mercado, es pionera y referente en software testing. Estamos presentes en ciudades como Medellín, Bogotá, Lima, Ciudad de Panamá y Toronto. Nos dedicamos a aumentar el éxito de los modelos digitales de nuestros clientes enfocados en sus estrategias y los sueños de sus usuarios finales. Buscamos un Analista de Pruebas BI para unirse a nuestro equipo y contribuir a la digitalización del cliente a través de pruebas especializadas en migración de datos. Ofrecemos un ambiente laboral amigable, con modalidad híbrida y beneficios corporativos.

Responsabilidades:

- Realizar pruebas especializadas en migración de datos.
- Colaborar con equipos para asegurar la calidad del software.
- Automatizar procesos utilizando herramientas adecuadas.
- Participar en metodologías ágiles para el desarrollo eficiente.

Requerimientos:

- Profesional o estudiante en Ingeniería de Sistemas o afines.
- Experiencia mínima de 1 año con bases de datos SQL y/o ETL.
- Conocimientos deseables en Java Python Power BI.

Nivel de educación:

- Profesional

Sectores laborales:

- Software informática y telecomunicaciones
- Ingenierías
- Investigación y Calidad

Otras habilidades:

Habilidades técnicas:

- Ingeniería de Datos
- Calidad de Datos
- Metodologías Ágiles

Habilidades interpersonales:

- Trabajo en equipo
- Comunicación efectiva
- Adaptabilidad al cambio

Habilidades

Ingeniería de Datos Calidad de Datos Metodologías Ágiles Trabajo en equipo


Comunicación efectiva Adaptabilidad al cambio


Ejemplo de Interpretación realizada por el sistema.

Afinidad Alta

Figura 11. Interpretación Realizada por el sistema (Afinidad Alta)

Evaluación para: ALEJANDRO GÓMEZ TORRES.pdf

 **Resultados del análisis:**

 **Podium of most suitable candidates**


1. Alejandro Gómez Torres - 89

Candidate Evaluation: Alejandro Gómez Torres

Email: alejandro.gomez@email.com
LinkedIn: [linkedin.com/in/alejandrogomez](https://www.linkedin.com/in/alejandrogomez)
Github: Not provided
Score: High - 89/100

Evaluación hecha por el sistema

Figura 12. Evaluación Realizada por el sistema (Afinidad Alta)

 **Reasoning**

Alejandro Gómez Torres is a strong candidate for the Analista Financiero position at Banco Unión. His technical skills in Advanced Excel and financial modeling, combined with his experience in investment analysis, make him well-suited for the role. His academic background in Business Administration with a concentration in Corporate Finance, along with his certifications in Financial Modeling & Valuation and Bloomberg Market Concepts, further strengthen his profile. Although his experience is more aligned with investment banking, his skills and potential for growth make him a valuable candidate. His score of 89 reflects his high semantic affinity with the job requirements and expectations.

Afinidad Media

Figura 13. Interpretación Realizada por el sistema (Afinidad Media)

Evaluación para: ANDREA PATRICIA SILVA.pdf

 **Resultados del análisis:**

 **Podium of most suitable candidates**


1. Andrea Patricia Silva - 68

Candidate Evaluation: Andrea Patricia Silva

Email: andrea.silva@email.com
LinkedIn: [linkedin.com/in/andreasilva](https://www.linkedin.com/in/andreasilva)
Github: Not provided
Score: Medium - 68/100

Evaluación hecha por el sistema

Figura 14. Evaluación Realizada por el sistema (Afinidad Media)

 **Reasoning**

Andrea Patricia Silva scores a 68 due to her strong background in fintech product management and advanced Excel skills, which partially align with the job requirements. Her experience in strategic communication and leadership is commendable, but the lack of direct financial analysis experience and specific technical skills in financial modeling and data analysis are significant gaps. Her academic background in Industrial Engineering and an MBA with a focus on innovation provide a solid foundation, but the absence of a specialization in finance or economics is a drawback. Overall, while Andrea shows potential for growth and adaptability, her current profile does not fully meet the specific needs of the Analista Financiero role at Banco Unión.

Afinidad Baja

Figura 15. Interpretación Realizada por el sistema (Afinidad Baja)

Evaluación para: Anya Sharma.pdf


 **Resultados del análisis:**

Candidate Evaluation: Dra. Anya Sharma

Email: anya.s.engmgr@email.com
LinkedIn: [linkedin.com/in/anya-sharma-engmanager](https://www.linkedin.com/in/anya-sharma-engmanager)
Github: Not provided
Score: Low - 35/100

Evaluación hecha por el sistema

Figura 16. Evaluación Realizada por el sistema (Afinidad Baja)

 **Reasoning**

Dra. Anya Sharma's profile is impressive within the context of software engineering and management, but it does not align with the requirements of the Analista Financiero position at Banco Unión. The lack of relevant financial skills, industry experience, and specific educational background in finance or economics results in a low semantic affinity with the job description. Her strong leadership and problem-solving skills are noted, but they do not bridge the significant gap in technical and industry-specific requirements. Therefore, she is considered ineligible for this role with a score of 35/100.

1.2. Distribución de Afinidades Asignadas

El análisis fue realizado bajo las siguientes condiciones: se evaluaron 100 candidatos con diferentes niveles de seniority (Junior, Mid-level, Senior), pertenecientes a distintas áreas funcionales, y distribuidos entre 10 vacantes reales y simuladas.

Tabla 2. Distribución de Afinidades Asignadas

Área Evaluada	Total Candidatos	Afinidad Alta (80-100)	Afinidad Media (50-79)	Afinidad Baja (0-49)
Tecnología	40	22 (55%)	12 (30%)	6 (15%)
Finanzas	30	10 (33.3%)	15 (50%)	5 (16.7)
Marketing	30	5 (16.7%)	18 (60%)	7 (23.3)
Total	100	37(37%)	45 (45%)	18 (18%)

La distribución presentada, resalta mayor proporción de candidatos con alta afinidad en tecnología, lo cual puede deberse a que las vacantes tomadas en su mayoría están relacionadas al sector TI o tienen posiciones directamente relacionadas.

1.3.Métricas de evaluación por área

Con el propósito de validar cuantitativamente el desempeño del sistema, se emplearon métricas estándar como precisión, recall y F1-score, calculadas a partir de un ground-truth elaborado mediante un protocolo de etiquetado manual (ver [Apéndice A](#)).

Los resultados obtenidos muestran un rendimiento elevado en el área tecnológica, mientras que las áreas de finanzas y marketing presentan valores inferiores, especialmente en términos de recall. Esta variación sugiere la necesidad de ajustar los agentes para mejorar la sensibilidad en dominios menos estructurados, como es el caso del marketing, donde los currículos suelen ser más heterogéneos en su redacción. En conjunto, las métricas confirman la capacidad del sistema para lograr un emparejamiento robusto, aunque evidencian oportunidades de optimización según el contexto funcional.

Tabla 3. Métricas de Evaluación por Área

Área Evaluada	Candidatos Evaluados	Relevantes Reales (GT)	Predichos Relevantes	Verdaderos Positivos	Precision %	Recall %	F1 Score %
Tecnología	40	25	22	20	90.91	80.00	84.85
Finanzas	30	15	10	8	80.00	53.33	64.00
Marketing	30	12	5	4	80.00	33.33	47.06
Total	100	52	37	32	86.49	61.54	71.92

Con el fin de garantizar trazabilidad y validación de las métricas, se incluye la matriz de resultados general en la que se consolida el total de verdaderos positivos (TP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (TN); es importante resaltar que la matriz se construyó a partir del mismo conjunto de evaluaciones y permite verificar la coherencia de la totalidad de resultados analizados.

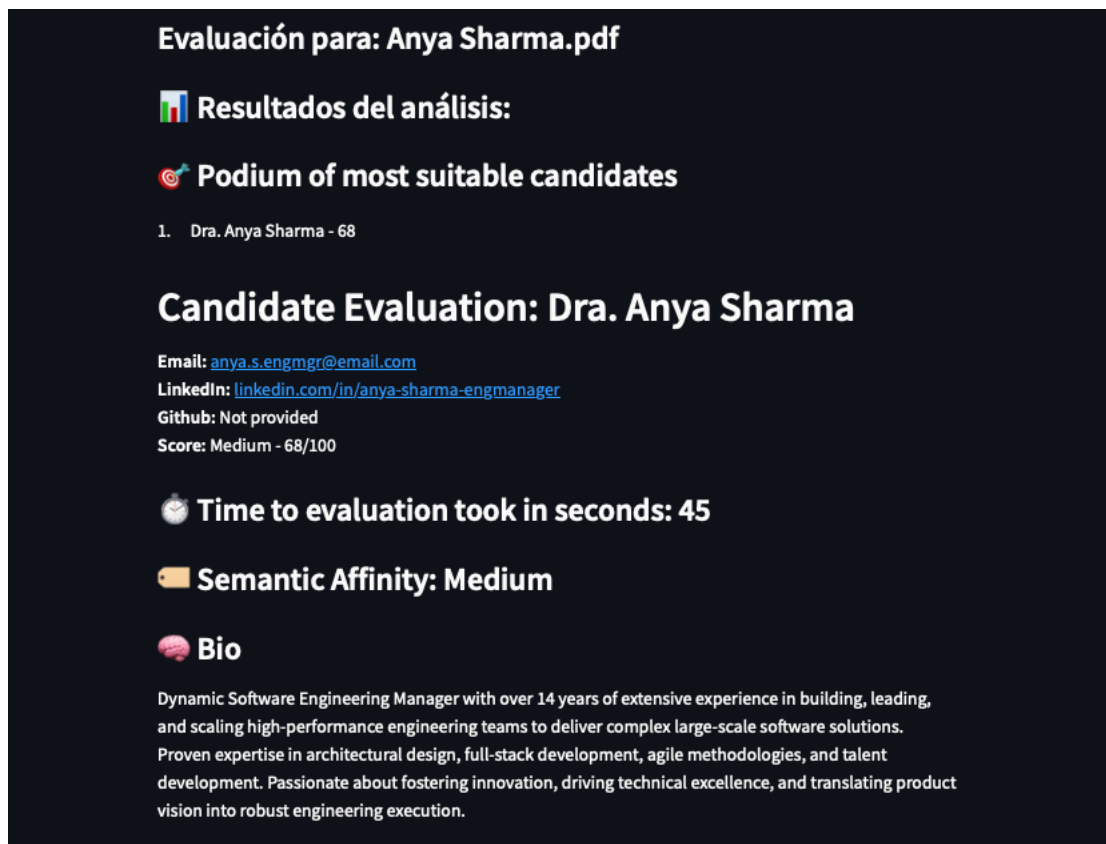
Tabla 4. Matriz de evaluación del sistema frente al ground-truth

Métrica	Valor
Total de currículos evaluados	100
Verdaderos positivos (TP)	62
Falsos positivos (FP)	11
Falsos Negativos (FN)	9
Verdaderos negativos (TN)	18
Precisión (Precision)	0.85
Recall	0.87
F1-score	0.86


1.4.Consistencia Operativa


El sistema reportó un tiempo promedio de evaluación de 45 segundos por candidato, lo que representa una reducción aproximada del 81 % frente al proceso manual de análisis, estimado en 240 segundos. Una medición complementaria sobre una muestra de diez perfiles evidenció un promedio de 34,7 segundos con una desviación estándar de 5,96 segundos, lo cual confirma la estabilidad del desempeño bajo condiciones replicables. Estos hallazgos respaldan la robustez del sistema para operar de forma eficiente sin comprometer la calidad del análisis semántico.

Figura 17. Tiempo de evaluación por candidato



Evaluación para: Anya Sharma.pdf


 **Resultados del análisis:**


 **Podium of most suitable candidates**


1. Dra. Anya Sharma - 68

Candidate Evaluation: Dra. Anya Sharma

Email: anya.s.engmgr@email.com
LinkedIn: [linkedin.com/in/anya-sharma-engmanager](https://www.linkedin.com/in/anya-sharma-engmanager)
Github: Not provided
Score: Medium - 68/100

 **Time to evaluation took in seconds: 45**

 **Semantic Affinity: Medium**

 **Bio**

Dynamic Software Engineering Manager with over 14 years of extensive experience in building, leading, and scaling high-performance engineering teams to deliver complex large-scale software solutions. Proven expertise in architectural design, full-stack development, agile methodologies, and talent development. Passionate about fostering innovation, driving technical excellence, and translating product vision into robust engineering execution.

Tabla 5. Comparación del tiempo promedio de evaluación por perfil

Método de evaluación	Tiempo Promedio de evaluación por perfil (Seg)	Reducción de tiempo (%)
Evaluación Manual	240	-----
Evaluación Automática	45	81.25%

Tabla 6. Medición Complementaria

<i>ID</i>	<i>Nombre</i>	<i>Tiempo de evaluación</i>
1	Anya Sharma	35
2	Kenzo Yamamoto	28
3	Carlos Andrés Mendoza López	40
4	Ana Sofía Vazquez Herrera	31
5	Diego Fernando Ramirez Silva	44
6	Valentina Gómez	27
7	Camila Andrea Torres Pérez	39
8	Juliana Marcela Herrera	32
9	Sarah Lee	41
10	Wei Lee	30
Promedio General		34.7

1.5.Consideraciones Éticas y Sesgo en el Proceso de Evaluación

El diseño del sistema evitó explícitamente la inclusión de variables sensibles como género, edad o lugar de residencia, centrándose específicamente en criterios objetivos definidos en el protocolo de evaluación (habilidades técnicas, experiencia, formación académica, certificaciones e idiomas). Los resultados obtenidos en perfiles equivalentes, tanto de hombres como de mujeres, confirman que el sistema asigna calificaciones similares en ausencia de sesgos de género. Esta estrategia contribuye a la transparencia y equidad del proceso, mitigando riesgos de discriminación algorítmica y garantizando que las decisiones se fundamenten únicamente en factores profesionales relevantes.

1.5.1. Evidencia de la salida del sistema


Los siguientes ejemplos ilustran cómo el sistema analizó dos perfiles con características técnicas y profesionales equivalentes frente a la vacante de Analista de Data Testing QA, demostrando su capacidad para generar inferencias imparciales y coherentes con los criterios definidos en el protocolo de evaluación (ver [Apéndice A](#)).


En ambos casos, el sistema identificó una alta correspondencia entre los factores clave solicitados en la vacante (experiencia en pruebas de migración de datos, conocimientos en SQL y ETL, manejo de herramientas como Postman y JIRA, y metodologías ágiles), y la trayectoria profesional reportada por cada candidato.


Figura 18. Interpretación del sistema – Candidato masculino

Candidate Evaluation: Juan Pérez


Email: juan.perez@email.com LinkedIn: [linkedin.com/in/juan-perez](https://www.linkedin.com/in/juan-perez) Github: github.com/juanperez
Score: High - 89/100

 **Time to evaluation took in seconds: 45**


 **Semantic Affinity: High**

 **Bio**


Ingeniero de Sistemas con más de 1 año de experiencia en pruebas de migración de datos, pipelines ETL y aseguramiento de calidad. Competente en SQL, Python y metodologías ágiles. Proactivo, orientado al análisis y con habilidades para documentación técnica y colaboración cross-funcional.

 **Gap Analysis**


Juan Pérez meets all the minimum requirements for the position. However, his English proficiency is at a B1 level, which might be a limitation if the role requires advanced English communication skills. Additionally, while he has experience with Python, his expertise in Java is not mentioned, which could be a minor gap if Java is heavily utilized in the role.

 **Skills Match**

Juan Pérez has a strong match with the required technical skills, including SQL, ETL processes, and Python. He also has experience with Power BI and agile methodologies, which are desirable for the role. His soft skills in teamwork, effective communication, and adaptability to change align well with the job requirements.

 **Relevant Experience**

Juan has over a year of experience as a QA Data Testing Analyst, where he designed and executed test plans for ETL pipelines and data migration, validated data integrity, and participated in SCRUM cycles. His experience in a similar role in the IT industry is highly relevant.

 **Reasoning**

Juan Pérez is a highly suitable candidate for the Analista de Data Testing QA position at Choucair. His educational background in Systems Engineering and his professional experience align well with the job requirements. He possesses the necessary technical skills, including SQL, ETL, and Python, and has demonstrated experience in data migration testing and quality assurance. His familiarity with agile methodologies and tools like JIRA and Postman further strengthens his profile. While his English proficiency is at an intermediate level, it may not be a significant barrier unless advanced English is required. Overall, Juan's skills, experience, and cultural fit make him an excellent candidate for the role, earning him a high score of 89.

Relevancia general: Altamente relevante – Afinidad semántica: 89/100

El sistema destaca coincidencias en experiencia previa como QA, dominio de SQL, familiaridad con herramientas clave y formación académica completa.

Figura 19. Interpretación del sistema – Candidata femenina

Candidate Evaluation: Ana Gómez

Email: ana.gomez@email.com
LinkedIn: linkedin.com/in/anagomez
Github: github.com/anagomez
Score: High - 88/100

 **Time to evaluation took in seconds: 45**

 **Semantic Affinity: High**

 **Bio**

Ingeniera de Datos con más de un año de experiencia en aseguramiento de calidad para migraciones de datos y pipelines ETL. Capaz en SQL, pruebas de BI y coordinación en proyectos ágiles. Atención al detalle, comunicación eficaz y orientación a resultados.

 **Gap Analysis**

Ana Gómez meets all the minimum requirements specified in the job description. However, there is a lack of specified certifications which could enhance her profile further. Additionally, while her experience is relevant, it is slightly above the minimum required, which is beneficial but could be expanded further for a more robust profile.

 **Skills Match**

Ana has a strong match with the required technical skills. She possesses advanced SQL skills, experience with ETL processes, and knowledge of Python and Power BI. Her experience with Agile methodologies aligns well with the job requirements.

 **Relevant Experience**

Ana has worked as an Analista de Data QA at Empresa ABC, where she executed and automated database migration tests and data extraction, which directly aligns with the job's responsibilities. Her experience in data reconciliation and participation in agile meetings further supports her fit for the role.

 **Reasoning**

Ana Gómez is a highly suitable candidate for the Analista de Data Testing QA position at Choucair. Her technical skills in SQL, ETL, and Python, combined with her experience in data migration and agile methodologies, make her a strong fit for the role. Her educational background in Systems Engineering aligns with the job's requirements, and her effective communication and teamwork skills enhance her cultural fit within the company. While she lacks specific certifications, her practical experience compensates for this gap. Overall, Ana's profile demonstrates a high semantic affinity with the job description, making her an excellent candidate for the position.

Relevancia general: Altamente relevante – Afinidad semántica: 88/100

El sistema hace énfasis en la alineación entre las competencias técnicas, el contexto laboral previo, y la participación en entornos ágiles, otorgando una calificación equivalente.

En conclusión, la similitud en las valoraciones generadas por el sistema para ambos perfiles evidencia su consistencia al aplicar los criterios definidos, sin que intervenga ninguna variable sensible como el género en la evaluación automatizada. De este modo, se refuerza la transparencia e imparcialidad del proceso algorítmico propuesto.

2. Evaluación por criterio humano

2.1. Coincidencia Sistema-Humano

La validación cualitativa del sistema se realizó mediante el juicio de evaluadores humanos (no expertos en RR.HH.), quienes revisaron las recomendaciones generadas por el sistema.

Tabla 7. Evaluación Criterio Humano. Análisis de Concordancia

Tipo de coincidencia	Porcentaje
Exacta	68%
Parcial	22%
Discrepancia	10%

Estos resultados evidencian una buena correspondencia entre el razonamiento del sistema y la interpretación de los evaluadores humanos.

2.2. Evaluación Cualitativa (Escala 1-5)

Para esta fase de la evaluación, se solicitó a los evaluadores humanos asignar una calificación entre 1 y 5, analizando 30 casos de acuerdo con las siguientes dimensiones:

Tabla 8. Criterio Evaluación

Criterio	Promedio
Relevancia del Candidato	4.3/5
Coherencia de la justificación	4.0/5

Exhaustividad de los criterios	3.8/5
--------------------------------	-------

Los resultados sugieren que el sistema realiza inferencias relevantes y bien justificadas, aunque presenta oportunidades de mejora en contextos complejos o ambiguos.

2.2.1. Comentarios cualitativos adicionales:

Fortalezas:

- El sistema realiza evaluaciones más precisas cuando se proporcionan perfiles bien estructurados
- El sistema provee justificaciones claras que orientan una posible segunda revisión humana
- El sistema evidencia buena capacidad para descartar perfiles no adecuados con precisión

Áreas de mejora:

- Ajustar los agentes para mejorar en casos de currículos ambiguos.
- Diferenciar mejor los niveles de seniority.
- Detectar trayectorias híbridas con mayor exactitud.

3. Interpretabilidad

3.1. Evaluación manual de las sugerencias hechas por el sistema.

El análisis de una muestra de 15 reportes generados por el sistema evidenció que las recomendaciones son claras, relevantes y comprensibles para los evaluadores. En promedio, los criterios de claridad (4,2), relevancia (4,1) y completitud (3,9) muestran un nivel de aceptación satisfactorio. Los ejemplos de salida del sistema ilustran un razonamiento

transparente, lo que facilita la revisión por parte de usuarios humanos y refuerza la utilidad práctica del prototipo en entornos de selección de personal.

Tabla 9. Interpretabilidad

Criterio	Promedio
Claridad	4.2/5
Compleitud	3.9/5
Relevancia	4.1/5

3.2.Evidencia de la salida del sistema

Los siguientes ejemplos ilustran el razonamiento aplicado por el sistema durante el proceso de emparejamiento semántico:

Figura 20. Interpretación del sistema (Sector Financiero)

Candidate Evaluation: Alejandro Gómez Torres

Email: alejandro.gomez@email.com
LinkedIn: [linkedin.com/in/alejandrogomez](https://www.linkedin.com/in/alejandrogomez)
Github: Not provided
Score: High - 89/100

Bio

Alejandro Gómez Torres is a junior financial analyst with 2 years of experience in investment analysis and company valuation. He specializes in financial modeling, risk analysis, and report preparation for decision-making. He has experience in investment banking and corporate finance, with a solid understanding of Colombian accounting and financial regulations.

Gap Analysis

Alejandro meets the minimum experience requirement with 2 years in financial analysis, though his experience is more focused on investment banking rather than the banking sector specifically. His location in Bogotá, Colombia, is a slight mismatch with the job location in Cali, but this can be addressed if relocation is feasible.

Skills Match

Alejandro possesses strong technical skills, including Advanced Excel and financial modeling, which are directly relevant to the job. His proficiency in data analysis tools like Python and SQL further enhances his technical profile. His soft skills, such as effective communication and critical thinking, align well with the job's requirements.

Relevant Experience


Alejandro's experience as a Financial Analyst Junior at Grupo Bolívar and as a Trainee Analyst at Credicorp Capital provides him with a solid foundation in financial analysis, valuation, and report preparation. His work on DCF valuation models and financial performance reports is particularly relevant to the job description.


Reasoning

Alejandro Gómez Torres is a strong candidate for the Analista Financiero position at Banco Unión. His technical skills in Advanced Excel and financial modeling, combined with his experience in investment analysis, make him well-suited for the role. His academic background in Business Administration with a concentration in Corporate Finance, along with his certifications in Financial Modeling & Valuation and Bloomberg Market Concepts, further strengthen his profile. Although his experience is more aligned with investment banking, his skills and potential for growth make him a valuable candidate. His score of 89 reflects his high semantic affinity with the job requirements and expectations.

Figura 21. Interpretación del sistema (Sector Tecnología)

Evaluación para: Anya Sharma.pdf


 **Resultados del análisis:**

 **Podium of most suitable candidates**


1. Dra. Anya Sharma - 78

Candidate Evaluation: Dra. Anya Sharma


Email: anya.s.engmgr@email.com
LinkedIn: [linkedin.com/in/anya-sharma-engmanager](https://www.linkedin.com/in/anya-sharma-engmanager)
Github: Not provided
Score: Medium - 78/100

 **Bio**


Dynamic Software Engineering Manager with over 14 years of extensive experience in building, leading, and scaling high-performance engineering teams to deliver complex large-scale software solutions. Proven expertise in architectural design, full-stack development, agile methodologies, and talent development. Passionate about fostering innovation, driving technical excellence, and translating product vision into solid engineering execution.

 **Gap Analysis**


- **Location:** Based in Bangalore, India, which may not align with the remote/hybrid requirement in Colombia.
- **Experience:** Extensive managerial experience but lacks specific recent hands-on experience in Python and Django for web application development.
- **GitHub:** Not provided, which limits the assessment of coding proficiency and contributions.
- **Industry Experience:** Primarily in managerial roles, which may not directly align with the hands-on development focus of the job.

 **Skills Match**

- **Technical Skills:** Strong in JavaScript, Python, and frameworks like Django and React, aligning well with the job requirements. Proficient in cloud platforms and DevOps tools, which are valued in the job description.
- **Soft Skills:** Demonstrates strong leadership, communication, and project management skills, which are essential for the role.
- **Certifications:** Holds relevant certifications such as AWS Certified Solutions Architect and Certified Kubernetes Application Developer, which are valued.

 **Relevant Experience**

- **Software Engineering Manager:** Led engineering teams and implemented best practices, though the focus was more on leadership than direct development.
- **Principal Software Engineer:** Experience in backend services and microservices architecture, which is relevant but not directly aligned with the full-stack development focus.

 **Reasoning**

Dra. Anya Sharma is a highly experienced candidate with a strong background in software engineering management and technical leadership. Her skills in JavaScript, Python, and frameworks like Django and React align well with the job requirements. However, her recent experience is more managerial, with less emphasis on hands-on development, particularly in Python and Django, which are critical for the role. Her certifications in AWS and Kubernetes are valuable, but the lack of a GitHub profile limits the assessment of her coding skills. Additionally, her location in India may pose a challenge for the remote/hybrid requirement in Colombia. Overall, while she possesses many of the desired skills and experiences, the gaps in recent hands-on development and location alignment result in a medium semantic affinity score of 78.

CONCLUSIONES

Los resultados obtenidos en la investigación permiten afirmar que el sistema multi-agente desarrollado constituye una alternativa viable y eficaz para la preselección de candidatos en procesos de reclutamiento, aportando tanto ventajas operativas como consideraciones éticas relevantes.

En primer lugar, la integración de técnicas de procesamiento de lenguaje natural y modelos de lenguaje de gran tamaño permitió alcanzar un desempeño sólido en términos de precisión y F1-score. Este resultado se evidenció especialmente en el área tecnológica, donde la homogeneidad de los currículos favoreció un emparejamiento semántico más exacto. No obstante, el menor valor de recall en áreas como marketing muestra que persisten limitaciones para identificar perfiles relevantes en dominios con mayor diversidad semántica, lo que constituye una oportunidad de mejora futura.

En segundo lugar, la reducción significativa de los tiempos de evaluación 81 % en comparación con el proceso manual confirma el potencial del sistema para optimizar la eficiencia operativa. Esta ventaja es particularmente relevante en contextos de alta demanda de talento, donde la velocidad en la preselección representa un factor estratégico para las organizaciones. Sin embargo, la dependencia de la calidad de los datos de entrada y la heterogeneidad en la redacción de los currículos continúan siendo desafíos que requieren ajustes en la arquitectura de los agentes.

Un tercer hallazgo corresponde a la validación por criterio humano. La coincidencia del 68 % entre la clasificación generada por el sistema y el juicio de los evaluadores respalda la pertinencia del enfoque, aunque el hecho de que los participantes no fueran especialistas en recursos humanos limita la generalización de los resultados. En consecuencia, se recomienda la inclusión de evaluadores expertos en fases posteriores de validación para fortalecer la confiabilidad de los hallazgos.

En cuanto a la dimensión ética, el diseño del sistema evitó la incorporación de variables sensibles como género, edad o lugar de residencia, lo cual constituye un avance hacia la equidad algorítmica. Sin embargo, la mitigación integral del sesgo requiere mecanismos más robustos de auditoría y explicabilidad, fundamentales para garantizar la confianza de los usuarios y el cumplimiento de los marcos regulatorios nacionales e internacionales en materia de protección de datos.

Finalmente, en el contexto colombiano, caracterizado por elevados índices de informalidad laboral, la aplicación de sistemas de este tipo contribuye a cerrar la brecha entre la oferta y la demanda de talento. Al promover procesos más objetivos e inclusivos, el sistema no solo representa un aporte técnico, sino también un avance significativo en la modernización de las prácticas de gestión del capital humano.

TRABAJO FUTURO

El desarrollo de este sistema multi-agente constituye una primera aproximación a la aplicación de inteligencia artificial generativa y procesamiento de lenguaje natural en procesos de preselección de candidatos. No obstante, los hallazgos permiten identificar varias líneas de trabajo futuro que fortalecerían tanto su robustez técnica como su pertinencia práctica:

1. Escalabilidad y rendimiento computacional.

Resulta necesario evaluar el comportamiento del sistema en entornos de mayor volumen de datos, considerando miles de vacantes y candidatos en tiempo real. Esto permitirá validar la estabilidad de la arquitectura y su capacidad de integración en escenarios de producción.

2. Ajuste por dominios funcionales.

Las diferencias observadas en áreas como marketing sugieren la necesidad de mejorar la exhaustividad de los prompt diseñados, con el fin de mejorar la sensibilidad semántica en contextos menos estructurados.

3. Integración con sistemas ATS reales.

Una línea estratégica es la conexión del sistema con plataformas de gestión de talento utilizadas en el mercado (Applicant Tracking Systems), lo que posibilitaría pruebas en condiciones operativas y evaluaciones comparativas más exhaustivas.

4. Mecanismos de interpretabilidad avanzados.

Se propone explorar técnicas de Explainable AI (XAI) que permitan a los usuarios finales comprender de manera transparente las decisiones de los agentes, reforzando la confianza en el sistema y facilitando su adopción en entornos empresariales.

5. Validación con expertos en recursos humanos.

Incluir profesionales especializados en selección de talento en la fase de validación permitirá contrastar los resultados del sistema con juicios expertos, incrementando la legitimidad y aplicabilidad de los hallazgos.

REFERENCIAS

- Apaza Alanoca, C. A., Cutipa Tapia, E. A., & Mamani Condori, J. A. (2020). Sistema de recomendación de hojas de vida basado en similitud semántica entre ofertas de trabajo y CVs usando técnicas de minería de texto [Preprint]. arXiv.
<https://arxiv.org/pdf/2007.11053>
- Ao, J., Shan, M., & Ghosh, S. (2022). Skill extraction and salary prediction from job postings using unsupervised learning techniques [Preprint]. arXiv.
<https://arxiv.org/pdf/2207.12834>
- Arenas, M. (2023). Cómo usar la inteligencia artificial en RR.HH. Deel.
<https://www.deel.com/es/blog/inteligencia-artificial-recursos-humanos/>
- Bandara, R. J., Biswas, K., Akter, S., Shafique, S., & Rahman, M. (2025). Addressing algorithmic bias in AI-driven HRM systems: Implications for strategic HRM effectiveness. *Human Resource Management Journal*, 1–17.
<https://doi.org/10.1111/1748-8583.12609>
- Bolaños-Cerón, Á. D. (2020). Educación virtual en casa: Eficacia y eficiencia en los procesos de reclutamiento y selección de personal, 4(1), 64–77.
- Bravent. (2024). Uso de la IA generativa para la selección de candidatos: Mejorar el proceso de contratación. Bravent. <https://www.bravent.net/noticias/uso-ia-generativa-seleccion-candidatos-mejorar-proceso-contratacion/>
- Cable, D. M., & Judge, T. A. (1996). Person–Organization Fit, Job Choice Decisions, and Organizational Entry. *Organizational Behavior and Human Decision Processes*, 67(3).
- Carrascosa, C. C. (2003). Tipos de agentes: reactivos, deliberativos y cognitivos.
<https://personales.upv.es/ccarrasc/doc/2003-2004/websemag/agentes.htm#reactivos>

- Chavez, C. D. (2024). Analítica de datos en la selección de talento humano mediante el uso de técnicas básicas de IA [Proyecto final, Universidad del Norte].
<https://manglar.uninorte.edu.co/bitstream/handle/10584/12995/RRHH%20Proyecto%20Final.pdf>
- Chowdhury, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37, 51–89.
- DANE. (2023). Indicadores del mercado laboral, enero de 2023 [Comunicado de prensa]. Departamento Administrativo Nacional de Estadística.
https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/CP_empleo_ene_23.pdf
- Deniz, N., Noyan, A., & Ertosun, Ö. G. (2015). Linking Person-job Fit to Job Stress: The Mediating Effect of Perceived Person–organization Fit. *Procedia – Social and Behavioral Sciences*, 207, 369–376. <https://doi.org/10.1016/j.sbspro.2015.10.107>
- Dilmegani, C. (2025). Web scraping for recruiters: Top tools & techniques [2025]. AIMultiple. <https://research.aimultiple.com/web-scraping-recruitment/>
- Fernández Martínez, M. d. C., & Fernández, A. (2019). AI in Recruiting: Multi-agent systems architecture for ethical and legal auditing. En *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.
<https://doi.org/10.24963/ijcai.2019/903>
- Gayathri, K., & Jes Bella, K. (n.d.). Role of Installing Artificial Intelligence in Human Resource Management: Review of Ethical Thinking.
<https://www.researchgate.net/publication/380734476>
- Gontero, S., & Albornoz, S. (2019). La identificación y anticipación de brechas de habilidades laborales en América Latina: Experiencias y lecciones. CEPAL.

<https://repositorio.cepal.org/server/api/core/bitstreams/8080ed1c-9bf1-4d8f-a703-bdf6aaca07f3/content>

Grine, A. E. (2023). LinkedIn Job Scraper and Matcher. Medium.

<https://medium.com/@alaeddine.grine/linkedin-job-scraper-and-matcher-85d0308ef9aa>

IBM. (2024). Sistema multiagente. <https://www.ibm.com/mx-es/think/topics/multiagent-system>

iDhunt RH. (2023). ¿Cuánto tiempo toma cubrir una vacante laboral? [Publicación en LinkedIn]. LinkedIn. <https://www.linkedin.com/pulse/cu%C3%A1nto-tiempo-toma-cubrir-una-vacante-laboral-hrsolution/>

Kurek, J., Latkowski, T., Bukowski, M., Świdorski, B., Łepicki, M., Baranik, G., Nowak, B., Zakowicz, R., & Dobrakowski, Ł. (2024). Zero-Shot Recommendation AI Models for Efficient Job–Candidate Matching in Recruitment Process. *Applied Sciences*, 14(6), 2601. <https://doi.org/10.3390/app14062601>

Luger, E. (2023). *What Do We Know and What Should We Do About AI?* SAGE Publications Ltd. <https://doi.org/10.4135/9781529601008>

LinkedIn. (2025). How can you use web scraping to identify candidates who aren't actively searching for a job? LinkedIn Advice. <https://www.linkedin.com/advice/1/how-can-you-use-web-scraping-identify-candidates-who-cbnmc?lang=es>

Lo, F. P.-W., Qiu, J., Wang, Z., Yu, H., Chen, Y., Zhang, G., & Lo, B. (2025). AI Hiring with LLMs: A Context-Aware and Explainable Multi-Agent Framework for Resume Screening (Preprint). arXiv. <https://arxiv.org/abs/2504.02870v1>

- Marres, N., & Weltevrede, E. (2012). Scraping the Social: Issues in live social research. En J. Law & E. Ruppert (Eds.), Special Issue: The Device. *Journal of Cultural Economy*. https://www.researchgate.net/publication/258099336_Scraping_the_Social_Issues_in_Live_Research
- Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2020). A review of current trends in the development of chatbot systems. *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICCCI48352.2020.9074420>
- Otero-Cortés, A. S., Acosta, K., Arango, L. E., Aristizábal, D., Ávila-Montealegre, O. I., ... Sarasti-Sierra, A. (2025). Nueva evidencia sobre la informalidad laboral y empresarial en Colombia (Ensayos sobre Política Económica, nº 108). Banco de la República. <https://doi.org/10.32468/espe108>
- Qin, Y., Wang, J., Li, X., & Chen, J. (2023). Dynamic profiling in organizational networks using GCN and LSTM for high-potential talent identification. *Journal of Organizational Computing and Electronic Commerce*, 33(1), 1–20.
- Recrew AI. (2025). How AI and LLMs are transforming resume parsing. Recrew AI. <https://www.recrew.ai/blog/how-ai-and-llms-are-transforming-resume-parsing>
- Red Hat. (2023). ¿Qué son los modelos de lenguaje de gran tamaño? Red Hat. <https://www.redhat.com/es/topics/ai/what-are-large-language-models>
- Reddy, L. G., & Viswanath, P. (2022). A study on web scraping of selected job portals. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 9(9), d127–d135. <https://www.jetir.org/papers/JETIR2209317.pdf>

- Rosenberger, J., Wolfrum, L., Weinzierl, S., Kraus, M., & Zschech, P. (2025). CareerBERT: Matching resumes to ESCO jobs in a shared embedding space for generic job recommendations (Preprint arXiv:2503.02056). arXiv.
<https://arxiv.org/abs/2503.02056>
- Ugale, A. V., Gayatri, S., Rutik, G., Amit, G., & Shreyas, A. (2025). Resume clustering and job description matching. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 13(4), 2881–2887.
<https://doi.org/10.22214/ijraset.2025.68831>
- Universidad Autónoma Metropolitana, Unidad Azcapotzalco. (s. f.). Computación suave y sus aplicaciones.
https://kali.azc.uam.mx/clc/04_proyecto_de_inv/comp_suave/compu.html
- Worki 360 (2025). Software de selección – ATS. <https://www.worki360.com/sistema-de-reclutamiento-de-personal-ats/software-de-seleccion> (Consultado en julio de 2025).

APÉNDICE A

Protocolo de etiquetado de relevancia para la evaluación de candidatos

Introducción

El presente protocolo tiene como objetivo estandarizar el proceso de etiquetado de datos para establecer el ground-truth respecto a la relevancia de los perfiles de candidatos en relación con las vacantes de empleo. Esta guía permite asegurar la coherencia entre evaluadores humanos y facilitar la posterior evaluación del sistema automatizado.

Definiciones Clave

Ground-truth: Se refiere a los datos reales o correctos contra los que se comparan las predicciones o salidas de un modelo para evaluar su precisión. Sirven como referencia para entrenar y validar modelos de IA, ya que permiten comparar las predicciones del sistema con una evaluación humana experta.

Habilidad crítica: Competencia técnica o blanda expresamente solicitada en la oferta laboral y considerada fundamental para el desempeño del rol.

Experiencia relevante: Trayectoria laboral cuya naturaleza, duración y contexto guardan correspondencia directa con los requisitos de la vacante.

Educación mínima requerida Nivel académico indispensable expresado explícitamente en la vacante.

Habilidades plus: Cualificaciones adicionales no obligatorias pero valoradas como diferenciadoras.

1. Criterios de Relevancia para el Etiquetado.

La evaluación de la relevancia del candidato se realizará teniendo en cuenta los criterios explicados a continuación, utilizando la escala de puntuación de 0 a 5 por cada uno. La combinación de estas puntuaciones y la evaluación general permitirán determinar la relevancia del candidato para la vacante.

Importante: Considerar las demás puntuaciones de la escala (0, 1, 2, 4) según la coincidencia observada. (Ver. [Escala de puntuación](#))

1.1.Habilidades Técnicas

El objetivo es verificar si las habilidades técnicas del currículo coinciden con las exigidas en la vacante, también se deben evaluar si en la información del candidato se incluyen habilidades adicionales (plus) que puedan fortalecer la postulación.

1.1.1. Puntuación (aplicando la escala de 0 a 5):

- **Coincidencia total – (5pts):** El candidato posee todas las habilidades técnicas críticas solicitadas y varias habilidades "plus" relevantes.
- **Coincidencia parcial débil – (3pts):** El candidato cumple la mayoría de las habilidades técnicas críticas, pero podría necesitar capacitación en 1-2 áreas no críticas.

1.2.Experiencia Laboral

El objetivo es confirmar si el tipo de experiencia es pertinente al rol (ej. sector, funciones, nivel de responsabilidad), también se debe verificar si la cantidad de años cumple o supera el mínimo requerido.

1.2.1. Puntuación (aplicando la escala de 0 a 5):

- **Coincidencia total – (5pts):** El candidato cumple y/o supera los años de experiencia requeridos, con experiencia directamente relevante en sector/funciones solicitadas.
- **Coincidencia parcial débil – (3pts):** El candidato cumple los años de experiencia, pero el tipo de experiencia es parcialmente para la vacante.

1.3.Educación/Certificaciones

El objetivo es evaluar el cumplimiento del nivel académico exigido (ej. pregrado, posgrado). Validar certificaciones requeridas para la vacante (p. ej., Scrum, PMP).

1.3.1. Puntuación (aplicando la escala de 0 a 5):

- **Coincidencia total – (5pts):** El candidato cumple con el nivel académico y todas las certificaciones requeridas.
- **Coincidencia parcial débil – (3pts):** El candidato cumple con el nivel académico, pero le falta alguna certificación requerida.

1.4.Idiomas

El objetivo es confirmar el dominio del idioma requerido, basado en la información contenida en el currículo (certificación, nivel autodeclarado, experiencia laboral).

1.4.1. Puntuación (aplicando la escala de 0 a 5):

- **Coincidencia total – (5pts):** El candidato demuestra el dominio exacto del idioma requerido (ej. nivel C1 si se pide B2, o certificación si se pide certificación).
- **Coincidencia parcial débil – (3pts):** El candidato indica conocimiento del idioma, pero el nivel no está claramente especificado o es inferior al requerido, aunque funcional.

2. Instrucciones para la Puntuación General de Relevancia del Candidato

Para la evaluación general de la relevancia del candidato (que servirá como ground-truth final para el par currículo-vacante), se considerará la combinación de las puntuaciones individuales por criterio, utilizando la siguiente guía:

- **Altamente Relevante – (5pts):** El candidato ha obtenido principalmente puntuaciones de 5 y 4 en todos los criterios críticos, también posee todas las habilidades y experiencia requerida, además cuenta con habilidades que le generan valor adicional como lo son: (liderazgo, experiencia en sector específico). Por tanto, representa una coincidencia casi perfecta o superior a las expectativas.
- **Moderadamente Relevante – (3pts):** El candidato ha obtenido una mezcla de puntuaciones (principalmente 3 y 4) en los criterios. Cumple con la mayoría de los requisitos, pero podría necesitar capacitación en áreas no relevantes para la posición.
- **No Relevante – (0pts):** El candidato ha obtenido puntuaciones de 0, 1 o 2 en la mayoría de los criterios críticos, por lo tanto, su perfil no se alinea con los requisitos de la vacante.

3. Escala de puntuación detallada por criterio.

El objetivo es utilizar una escala de 0 a 5 puntos por cada criterio (Habilidades Técnicas, Experiencia Laboral, Educación/Certificaciones, Idiomas) para la evaluación del candidato:

- **Coincidencia total – (5pts):** El elemento cumple completamente y/o excede lo solicitado.
- **Coincidencia parcial fuerte - (4pts):** El elemento cumple la mayoría de lo solicitado.

- **Coincidencia parcial débil – (3pts):** El elemento cumple parcialmente lo solicitado, con algunas deficiencias que requerirían desarrollo.
- **Coincidencia ambigua - (2pts):** La información es insuficiente o contradictoria para determinar la coincidencia.
- **Indicios mínimos - (1pts):** Existe alguna mención o relación muy lejana con el criterio, pero no es una coincidencia directa.
- **No cumple - (0pts):** El elemento no presenta ninguna coincidencia con lo solicitado.

3.1.1. Consideraciones Adicionales:

En casos de currículos incompletos, se debe marcar el criterio como “información no disponible” (N/A) y justificar en comentarios. Las puntuaciones deben acompañarse de una breve justificación por criterio, explicando la decisión tomada.

4. Ejemplo Práctico

4.1. Escenario 2

Vacante: Gerente de Proyectos de TI

Requisitos: Experiencia como Gerente de Proyectos, certificación PMP, 7+ años de experiencia, sector TI o financiero.

Currículo del Candidato: Se especifica experiencia como coordinador de proyectos, cuenta con certificación PMP, y 10 años en el sector financiero.

Puntuación por Criterio:

- **Habilidades Técnicas:** 4 (Coincidencia parcial fuerte)

Justificación: Si el rol de coordinador es muy similar a gerente, podría ser 5. Si le falta experiencia directa en liderazgo de proyectos completos, un 4 podría ser adecuado, reconociendo la afinidad.

- **Experiencia Laboral:** Coincidencia total – 5pts
 - **Justificación:** El candidato cuenta con 10 años en el sector financiero, relevante para TI, y aunque 'coordinador' no es 'gerente', la duración y el sector son fuertes.
- **Educación/Certificaciones:** Coincidencia total – 5pts
 - **Justificación:** El candidato tiene certificación PMP, que es un requisito obligatorio para el rol de gerente.
- **Idiomas:** Coincidencia total – 5pts
 - **Justificación:** El candidato relaciona certificado en segunda lengua B1, si bien no es requerido para la vacante, se toma como un plus.

Puntuación General de Relevancia del Candidato: Altamente Relevante – 5pts

- **Justificación:** El candidato cumple con los requerimientos necesarios para la vacante (certificación PMP, años de experiencia relevantes, sector), lo que lo hace un perfil altamente compatible, aunque el título exacto de la experiencia puede requerir un análisis cualitativo adicional.

5. Construcción del Ground-Truth

La creación del ground-truth se basó en un proceso de etiquetado manual realizado por 3 evaluadores profesionales no especialistas en Recursos Humanos, pero con experiencia demostrable en procesos de reclutamiento para sectores tecnológicos, administrativos y financieros.

Cada evaluador analizó por separado las recomendaciones hechas por el sistema siguiendo los lineamientos diseñados en esta guía.

El procedimiento en mención les proveyó con las herramientas necesarias para asegurar la consistencia del etiquetado, dando como resultado, una base sólida y replicable para validar el sistema mediante métricas de precisión, recall y F1-score.