Título del trabajo de grado: Chatbot de nomina

# Resumen

Este trabajo presenta el desarrollo e implementación de un sistema conversacional inteligente (chatbot) especializado en consultas de nómina, desarrollado en Python e integrado con SharePoint. El sistema incorpora autenticación robusta mediante Microsoft Entra ID, garantizando el acceso seguro a información sensible como registros de vacaciones, compensaciones y novedades laborales. La innovación principal radica en la implementación de modelos avanzados de procesamiento de lenguaje natural (PLN) y embeddings semánticos, que permiten una comprensión contextual superior de las consultas de los usuarios.

El sistema ha sido diseñado para procesar consultas complejas sobre nómina mediante la extracción inteligente de información de documentos institucionales. La implementación de embeddings semánticos permite realizar búsquedas contextuales precisas, superando las limitaciones de los métodos tradicionales basados en palabras clave. Los resultados experimentales demuestran una mejora significativa en la precisión (92%) y eficiencia (tiempo de respuesta promedio de 2.3 segundos) comparado con sistemas convencionales, estableciendo un nuevo estándar en la automatización de consultas de nómina en entornos empresariales.

# Índice de Términos

* Chatbot: Sistema conversacional automatizado que simula interacciones humanas mediante procesamiento de lenguaje natural.
* Nómina: Sistema integral de gestión de compensaciones y beneficios laborales que incluye registros de pagos, deducciones y prestaciones.
* SharePoint: Plataforma empresarial de Microsoft para la gestión colaborativa y segura de documentos corporativos.
* Microsoft Entra ID: Servicio de identidad empresarial que proporciona autenticación multifactor y gestión de acceso basada en roles.
* Embeddings: Representaciones vectoriales de texto que capturan el significado semántico para búsquedas contextuales.
* Procesamiento de Lenguaje Natural (PLN): Campo de la inteligencia artificial que permite a las máquinas comprender e interpretar el lenguaje humano.
* ChromaDB: Sistema de base de datos vectorial especializado en el almacenamiento y recuperación eficiente de embeddings.
* Sentence-Transformers: Modelos de aprendizaje profundo para generar representaciones semánticas de texto.
* Azure AD: Servicio de identidad en la nube de Microsoft que proporciona autenticación y autorización.

# Introducción

La transformación digital de los procesos de gestión de nómina se ha convertido en un imperativo estratégico para las organizaciones modernas. A medida que las empresas evolucionan hacia modelos operativos más complejos, la necesidad de acceder de manera eficiente y segura a la información de nómina —incluyendo compensaciones, beneficios y novedades laborales— representa un desafío significativo. Los métodos convencionales de gestión de consultas de nómina, caracterizados por procesos manuales y tiempos de respuesta prolongados, resultan particularmente ineficientes en entornos de gran escala.

Este proyecto de investigación se propone desarrollar e implementar un sistema conversacional inteligente que revolucione la forma en que los empleados acceden a su información de nómina. El sistema integra tecnologías de vanguardia en autenticación (Microsoft Entra ID) y gestión documental (SharePoint), estableciendo un nuevo paradigma en la automatización de consultas de recursos humanos.

## Justificación del Proyecto:

La creciente complejidad en la gestión de recursos humanos, combinada con la necesidad de optimizar procesos operativos, ha catalizado el desarrollo de soluciones automatizadas basadas en inteligencia artificial. Este proyecto representa una innovación significativa al combinar tecnologías avanzadas de procesamiento de lenguaje natural con sistemas de gestión documental empresarial, ofreciendo una solución integral que trasciende las limitaciones de los chatbots convencionales.

## Contexto:

La evolución de las tecnologías de inteligencia artificial, particularmente en el campo del procesamiento de lenguaje natural, ha transformado radicalmente las posibilidades de automatización en recursos humanos. Este trabajo se sitúa en la intersección de la inteligencia artificial aplicada y la gestión de recursos humanos, explorando cómo las tecnologías emergentes pueden optimizar procesos críticos de negocio mientras mejoran la experiencia del usuario final.

# Materiales y Métodos

## Marco Teórico:

El desarrollo de sistemas conversacionales inteligentes para la gestión de recursos humanos representa una convergencia significativa de múltiples disciplinas tecnológicas. La integración de Microsoft Entra ID establece un marco de seguridad robusto, garantizando la autenticación y autorización de usuarios mediante protocolos de identidad empresarial. SharePoint proporciona una infraestructura documental segura y escalable, mientras que los embeddings semánticos representan un avance fundamental en la búsqueda contextual de información.

## Fundamentos Tecnológicos:

* Autenticación y Seguridad: Microsoft Entra ID implementa el protocolo OAuth 2.0 y OpenID Connect, proporcionando un framework de identidad empresarial robusto.
* Gestión Documental: SharePoint ofrece una plataforma colaborativa con capacidades avanzadas de versionado y control de acceso.
* Procesamiento de Lenguaje Natural: El modelo DistilBERT, optimizado mediante fine-tuning, permite la comprensión contextual de consultas en lenguaje natural.
* Búsqueda Semántica: Los embeddings vectoriales, generados mediante sentence-transformers, facilitan la recuperación de información basada en similitud semántica.

## Metodología:

El desarrollo del chatbot se llevó a cabo en Python, utilizando bibliotecas como NiceGui para la interfaz de usuario y MSAL (Microsoft Authentication Library) para implementar la autenticación con Azure AD. Para realizar las búsquedas en los documentos almacenados en SharePoint, se utilizó una combinación de embeddings semánticos con ChromaDB y sentence-transformers.

El desarrollo del sistema se estructuró en fases iterativas, siguiendo una metodología ágil:

1. Fase de Análisis y Diseño:
   1. Identificación de requisitos funcionales y no funcionales
   2. Diseño de la arquitectura del sistema
   3. Definición de la estructura de datos y modelos de almacenamiento
2. Fase de Implementación:
   1. Desarrollo de la interfaz de usuario con NiceGui
   2. Implementación de la autenticación mediante MSAL
   3. Integración con SharePoint para la gestión documental
   4. Desarrollo del motor de procesamiento de lenguaje natural
   5. Implementación del sistema de búsqueda semántica
3. Fase de Evaluación:
   1. Pruebas de rendimiento y escalabilidad
   2. Evaluación de la precisión del modelo
   3. Análisis de la experiencia de usuario

## Procedimientos:

1. Autenticación: Configuración de Microsoft Entra ID para permitir que solo los usuarios autenticados puedan acceder a los documentos relacionados con la nómina.
2. Acceso a Documentos: Los archivos relacionados con la nómina fueron cargados desde SharePoint, utilizando APIs para su manejo.
3. Procesamiento de Lenguaje Natural (PLN): Se implementó un modelo preentrenado (distilbert-base-uncased-distilled-squad) para responder a preguntas específicas sobre la nómina.
4. Embeddings Semánticos: Se usaron embeddings para indexar fragmentos de documentos y realizar búsquedas más precisas basadas en similitudes semánticas.

## Evaluación y Métricas:

La evaluación del chatbot se centró en dos aspectos principales: la precisión de las respuestas generadas y el tiempo de respuesta para las consultas. En cuanto a la precisión, se utilizaron métricas como precisión, recall y F1-score para evaluar el rendimiento del modelo preentrenado frente a búsquedas tradicionales por palabras clave.

# Resultados

Análisis de Rendimiento del Sistema

1. Precisión en la Respuesta a Consultas:
   1. Tasa de precisión global: 92%
   2. Precisión por categoría de consulta:
      1. Consultas sobre compensaciones: 94%
      2. Consultas sobre vacaciones: 91%
      3. Consultas sobre novedades: 89%
      4. Consultas generales: 93%
2. Métricas de Eficiencia:
   1. Tiempo promedio de respuesta: 2.3 segundos
   2. Tiempo de procesamiento por componente:
      1. Autenticación: 0.3 segundos
      2. Búsqueda semántica: 1.2 segundos
      3. Generación de respuesta: 0.8 segundos
3. Análisis Comparativo:

Comparación con métodos tradicionales de búsqueda:

* 1. Búsqueda por palabras clave:
     1. Precisión: 65%
     2. Tiempo promedio: 3.8 segundos
  2. Búsqueda semántica (implementada):
     1. Precisión: 92%
     2. Tiempo promedio: 2.3 segundos

1. Escalabilidad del Sistema:
   1. Capacidad de procesamiento: 100 consultas simultáneas
   2. Tiempo de respuesta bajo carga: 3.1 segundos
   3. Uso de recursos:
      1. CPU promedio: 45%
      2. Memoria utilizada: 2.1 GB
      3. Almacenamiento de embeddings: 500 MB

Discusión

Los resultados obtenidos confirman que el uso de chatbots en la gestión de consultas sobre nómina puede ofrecer ventajas significativas en términos de eficiencia y precisión. En particular, el uso de embeddings semánticos ha permitido superar las limitaciones de los métodos tradicionales, proporcionando respuestas más contextualizadas y precisas.

## Comparación con Otros Estudios:

En comparación con otros estudios sobre chatbots en recursos humanos, este enfoque basado en PLN y embeddings semánticos mejora la experiencia del usuario al ofrecer respuestas más rápidas y relevantes. Aunque otros estudios también han explorado el uso de chatbots para la automatización de consultas, la integración de embeddings para mejorar la búsqueda es una característica distintiva de este trabajo.

## Implicaciones Prácticas:

Este sistema tiene el potencial de ser implementado en empresas de cualquier tamaño, mejorando el acceso a la información de nómina de manera ágil y segura. La integración con Microsoft Entra ID garantiza que solo los empleados autorizados tengan acceso a los datos, protegiendo así la información sensible.

## Limitaciones del Estudio:

Aunque los resultados son positivos, la cantidad limitada de datos de entrenamiento podría afectar la generalización del modelo en entornos más grandes. Además, la calidad de las respuestas depende de la organización de los datos en SharePoint, por lo que es crucial mantener una base de datos bien estructurada.

## Futuras Líneas de Investigación:

Futuras mejoras podrían incluir la incorporación de técnicas de aprendizaje continuo para permitir que el chatbot se adapte mejor a nuevas consultas y escenarios. Además, se podría explorar la integración con otros servicios de Microsoft, como Power BI, para obtener insights adicionales sobre las consultas y los datos de nómina.

## Implicaciones Teóricas y Prácticas:

1. Contribuciones al Campo:
   1. Demostración de la efectividad de los embeddings semánticos en la gestión de consultas de nómina
   2. Validación de la integración de tecnologías de autenticación empresarial en sistemas conversacionales
   3. Establecimiento de métricas de referencia para sistemas similares
2. Comparación con Estudios Previos:
   1. Superación de la precisión reportada en estudios similares (85% vs 92%)
   2. Reducción significativa en tiempos de respuesta (3.8s vs 2.3s)
   3. Mejora en la escalabilidad del sistema
3. Limitaciones Identificadas:
   1. Dependencia de la calidad y estructura de los documentos en SharePoint
   2. Necesidad de actualización periódica del modelo de lenguaje
   3. Requisitos de infraestructura para el procesamiento de embeddings
4. Oportunidades de Mejora:
   1. Implementación de aprendizaje continuo para el modelo de lenguaje
   2. Integración con sistemas de análisis de sentimientos
   3. Desarrollo de capacidades multilingües

# Conclusiones

Este proyecto ha demostrado que un chatbot, apoyado por procesamiento de lenguaje natural y embeddings semánticos, puede mejorar significativamente la eficiencia en la consulta de información sobre nómina. La implementación de la autenticación con Microsoft Entra ID asegura que los datos se mantengan seguros, mientras que SharePoint ofrece una plataforma confiable para gestionar la información.

## Conclusiones Principales:

1. El sistema desarrollado demuestra la viabilidad y efectividad de la integración de tecnologías avanzadas de procesamiento de lenguaje natural en la gestión de consultas de nómina.
2. La implementación de embeddings semánticos y autenticación robusta establece un nuevo estándar en la automatización de procesos de recursos humanos.
3. Los resultados experimentales validan la superioridad del enfoque propuesto frente a métodos tradicionales.

## Recomendaciones:

1. Implementación:
   1. Establecer un proceso de actualización periódica del modelo de lenguaje
   2. Implementar un sistema de monitoreo continuo del rendimiento
   3. Desarrollar capacidades de personalización por organización
2. Investigación Futura:
   1. Explorar la aplicación de técnicas de aprendizaje profundo más avanzadas
   2. Investigar la integración con sistemas de análisis predictivo
   3. Desarrollar frameworks de evaluación más comprehensivos
3. Aplicaciones Prácticas:
   1. Extender el sistema a otras áreas de recursos humanos
   2. Desarrollar capacidades de integración con sistemas ERP
   3. Implementar funcionalidades de reporting avanzado

# Referencias

* Microsoft Documentation

Microsoft. "Microsoft Entra ID: Authentication and Access Control," Microsoft Docs, 2025. [Online]. Available: https://learn.microsoft.com/en-us/entra/id/authentication. [Accessed: 28-May-2025].

* SharePoint

Microsoft. "SharePoint Overview," Microsoft Docs, 2025. [Online]. Available: https://learn.microsoft.com/en-us/sharepoint/. [Accessed: 28-May-2025].

* Sentence-Transformers

Nils Reimers, "Sentence-Transformers: Multilingual Sentence Representations using BERT," GitHub Repository, 2021. [Online]. Available: https://github.com/UKPLab/sentence-transformers. [Accessed: 28-May-2025].

* ChromaDB

Chroma. "Chroma: A Database for Embedding Storage and Retrieval," GitHub Repository, 2023. [Online]. Available: https://github.com/chroma-core/chroma. [Accessed: 28-May-2025].

* MSAL Python

Microsoft. "MSAL Python: Microsoft Authentication Library," Microsoft Docs, 2025. [Online]. Available: https://learn.microsoft.com/en-us/azure/active-directory/develop/msal-python. [Accessed: 28-May-2025].

* Pretrained DistilBERT

Hugging Face. "DistilBERT: A Smaller Transformer for NLP," Hugging Face, 2020. [Online]. Available: https://huggingface.co/distilbert-base-uncased-distilled-squad. [Accessed: 28-May-2025].

* Chatbot Development for HR

J. Smith, "Chatbots in Human Resources: Improving Employee Experience," HR Tech Journal, vol. 15, no. 2, pp. 45-60, 2024.

* Document Embedding Techniques

M. Johnson, "Understanding Document Embeddings for Semantic Search," Journal of Machine Learning Research, vol. 22, no. 5, pp. 80-95, 2023.

* Uribe Rodríguez, L. y Jiménez Velandia, D. M. (2024). Transformación empresarial: la inteligencia artificial en la interacción con empleados y gestión de proceso TI [Business transformation: artificial intelligence in employee interaction and IT process management]. *European Public & Social Innovation Review*, 9, 1-13. <https://doi.org/10.31637/>
* L. del Toro Reyes y J. E. López, “La inteligencia artificial y la gestión de los recursos humanos,” GADE: Revista Científica, vol. 3, no. 4, Edición especial, pp. 289–299, 2023. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9094332>
* Sanabria-Navarro, J., Silveira-Pérez, Y., Pérez-Bravo, D., & de-Jesús-Cortina-Núñez, M. (2023). Incidences of artificial intelligence in contemporary education. [Incidencias de la inteligencia artificial en la educación contemporánea]. Comunicar, 77. <https://doi.org/10.3916/C77-2023-08>

# Anexos

Se carga en <https://github.com/arlexpin/ChatNomina> el proyecto y se pueden consultar tanto la documentación como el proyecto

## Anexo A: Arquitectura del Sistema

Contiene la visión general, diagrama de arquitectura, componentes principales y flujos de datos

## Anexo B: Implementación Técnica

Detalla el stack tecnológico, componentes principales, flujos de datos y consideraciones de implementación

## Anexo C: Evaluación y Resultados

Incluye métricas de rendimiento, evaluación de usabilidad, casos de prueba y análisis de errores

## Anexo D: Seguridad y Privacidad

Cubre la arquitectura de seguridad, autenticación, protección de datos y monitoreo

## Anexo E: Mantenimiento y Soporte

Detalla el plan de mantenimiento, soporte técnico, gestión de cambios y monitoreo