



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

## **Visualización de datos con perspectiva de género**

Rocío Ladrón de Guevara Jara<sup>a</sup>, Valentina Campaña Herrera<sup>b</sup>, Caua Terra Santiago Paz<sup>c</sup>

Jocelyn Dunstan Escudero<sup>d</sup>, Daniela Moyano Dávila<sup>e</sup>

<sup>a</sup> *Mayor ingeniería Matemática, Escuela ingeniera, Pontificia Universidad Católica. Cuarto año, rocioghj@uc.cl*

<sup>b</sup> *Mayor Ingeniería de Software, Escuela de Ingeniería, Pontificia Universidad Católica. Cuarto año, valentinacamph@uc.cl*

<sup>c</sup> *Mayor Ingeniería de Software, College ciencias naturales y matemáticas, Pontificia Universidad Católica. csantiagopaz@uc.cl*

<sup>d</sup> *Departamento de Ciencia de la Computación (DCC), Escuela ingeniería, Pontificia Universidad Católica. Profesora asistente, jdunstan@ing.puc.cl.*

<sup>e</sup> *Departamento de Diseño, Escuela de Diseño UC, Pontificia Universidad Católica. Profesora adjunta, dpmoyano@uc.cl.*

---

### **Resumen**

La evidencia de EE. UU. y Europa muestra que mujeres y hombres difieren significativamente en cuanto a accidentes y tasas de enfermedades laborales. Sin embargo, se necesitan más datos para los países del sur global. Usando un enfoque interdisciplinario, nuestro proyecto tiene como objetivo aprovechar datos desglosados por género sobre accidentes laborales en Chile. Para ello colaboramos con la Asociación Chilena de Seguridad (ACHS), que brinda salud ocupacional a la mitad de los empleados contratados en el país. Accidentes laborales y de trayecto se registran utilizando texto libre y sexos binarios. Dentro de estos textos se pueden detectar dos entidades: el agente que causó la enfermedad o accidente (por ejemplo, una superficie para caminar o escaleras) y el **MECANISMO** (por ejemplo, una caída o golpe). En esta investigación mostramos los resultados para la detección automática de mecanismos en 350.471 documentos correspondientes al periodo 2022-2023 utilizando la API ChatGPT. Nuestros resultados muestran que en el 57,3% de los accidentes se ve afectado un hombre, mientras que en el 42,7% es una mujer. La ocupación más común entre los accidentes protagonizados por hombres es el de operador de máquina, mientras que para las

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

mujeres se relaciona con tareas de limpieza. El trabajo futuro incluye comparar los resultados con otros Large Language Model (LLM) y anotar un corpus para evaluar el rendimiento de los modelos en cuanto a extracción comprensiva de información. Además, planeamos colaborar con científicos sociales para sopesar nuestros resultados en a la luz de los trabajadores sin contrato en Chile. Finalmente, estamos trabajando en aplicar una perspectiva feminista para visualizar estos datos sensibles.

**Palabras clave:** NLP, texto libre, LLM, agente, mecanismo, embeddings.

---

## **1. Introducción**

Las revisiones sistemáticas actuales muestran que hombres y mujeres pueden tener diferentes riesgos de lesiones y/o discapacidad ocupacional, ya sea para la misma ocupación o para ocupaciones diferentes, especialmente para algunas exposiciones físicas, pero también para exposiciones químicas y biológicas. La mayoría de los estudios que analizaron las exposiciones psicológicas no encontraron diferencias en el riesgo entre hombres y mujeres en diferentes ocupaciones (Biswas et al., 2022).

En Chile, sin información laboral desagregada por género y lista para usar, no podemos estimar las diferencias en las que los trabajadores y trabajadoras sufren enfermedades y accidentes profesionales. Nuestra hipótesis de trabajo es que mediante el uso del Procesamiento del Lenguaje Natural (**NLP**), podemos considerar grandes volúmenes de informes de accidentes y detectar el mecanismo del accidente, aprovechando así información desagregada por género.

Durante este informe mostraremos los métodos utilizados para agrupar ocupaciones y cómo extrajimos automáticamente el mecanismo del accidente. Luego pasamos a la sección de resultados, donde se describen las ocupaciones más comunes vistas en el conjunto de datos para mujeres y hombres. A continuación se resumen los mecanismos más comunes para el periodo analizado. Finalizamos con una discusión donde, entre otros temas, planificamos trabajos futuros

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

para sistematizar la evaluación de la extracción automática de información e introducimos la situación de los trabajadores sin contrato o en trabajo no remunerado.

## **2. Metodología.**

En primer lugar, hicimos un análisis exploratorio de los datos, viendo el tipo, cantidad y formato de los datos entregados. En específico para nuestro objetivo, observamos cual era la distribución por sexo de las ocupaciones existentes en la base de datos, los datos sin rellenar (nulos) e identificar cuál era el problema al analizar los datos sin hacer un procesamiento de la columna ocupaciones.

Como mencionamos, el desafío que tienen los actuales datos de la ACHS para hacer un análisis de cómo afectan los accidentes laborales a cada sexo, es que las ocupaciones están escritas en texto en formato libre, es decir, no tiene un esquema claro y regular. Por ejemplo, en la base de datos en la columna de ocupación se encuentran las palabras ‘profesor’, ‘docente’ y educador. Tres palabras con significados distintos pero que se pueden clasificar como el mismo tipo de ocupación. Por ello, como segundo paso buscamos una manera de clasificar y/o agrupar las ocupaciones, encontrando el ISCO-08-CL (Clasificador Chileno Uniforme de Ocupaciones) elaborado por el INE (Instituto Nacional de Estadísticas), el cual organiza las ocupaciones dependiendo a las tareas a realizar en cada trabajo con un estándar internacional propuesto por la OIT (Organización Internacional del Trabajo).

Como el contenido de este clasificador estaba en un archivo PDF, como tercer paso, computacionalmente, utilizando librerías de Python especializadas para extraer información de archivos PDF a estructuras de datos que podíamos usar para nuestro modelo de NLP, se extrajo la información del clasificador y la almacenamos en **DATAFRAMES**.

Finalmente, con los datos ya preparados, usamos un Modelo de Lenguaje de Gran Tamaño (**LLM**, por sus siglas en inglés) para obtener los **EMBEDDINGS** de clasificaciones estandarizadas provenientes

--	--	--



# PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

## ESCUELA DE INGENIERÍA

### Dirección de Investigación e Innovación

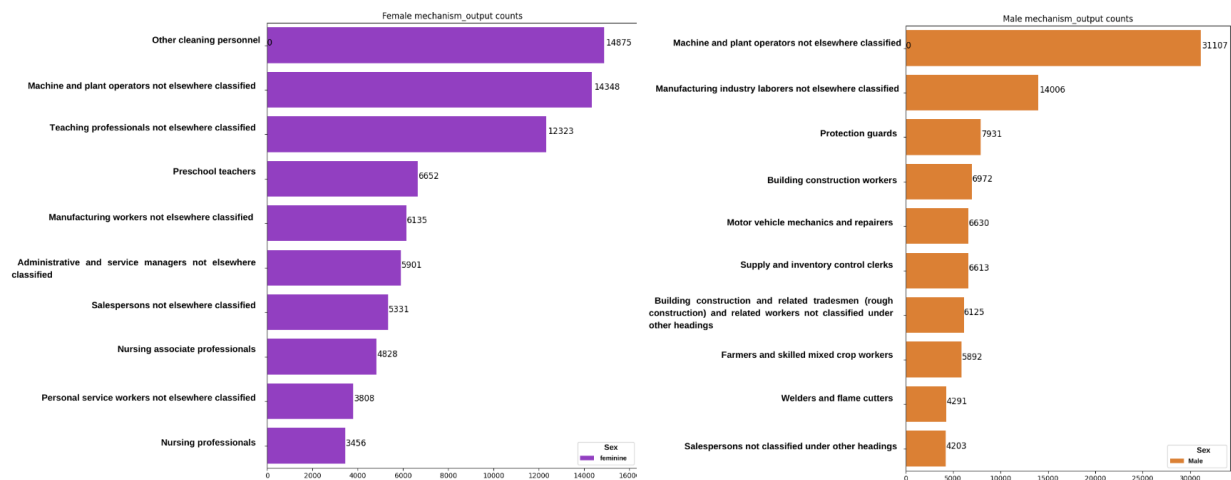
### Programa IPre de Investigación en Pregrado

de ISCO-08-CL y de los datos en la columna de ocupaciones. Posteriormente, con un algoritmo de que, en base a los criterios de similitud como distancia de los vectores, agrupaban ocupación con su categoría más cercana.

Ya con el procesamiento acabado procedíamos a nuevamente analizar los datos respondiendo la hipótesis principal del trabajo, ¿Existen diferencias entre los accidentes laborales por sexo?

### 3. Resultados y discusión

Dado los resultados, fue posible entender mejor la relación social de los accidentes laborales con la ocupación y género de los trabajadores con los datos proporcionados por la ACHS, figura 1.



**Figura 1.** Gráfico de ocupaciones desagregadas por género, donde las labores de limpieza, operarios de planta y máquinas, y profesionales de la educación son las tres ocupaciones más comunes para el género femenino, mientras que operarios de planta y máquinas, trabajadores de la industria manufacturera y guardias de seguridad son las más comunes para el género masculino.

Como podemos observar, las ocupaciones más comunes del sexo femenino y masculino solo comparten una ocupación en común, que es la de operarios de maquinaria. Es así que nos releva que existe una diferencia entre los sectores de ocupación en que las mujeres y hombres se accidentan.

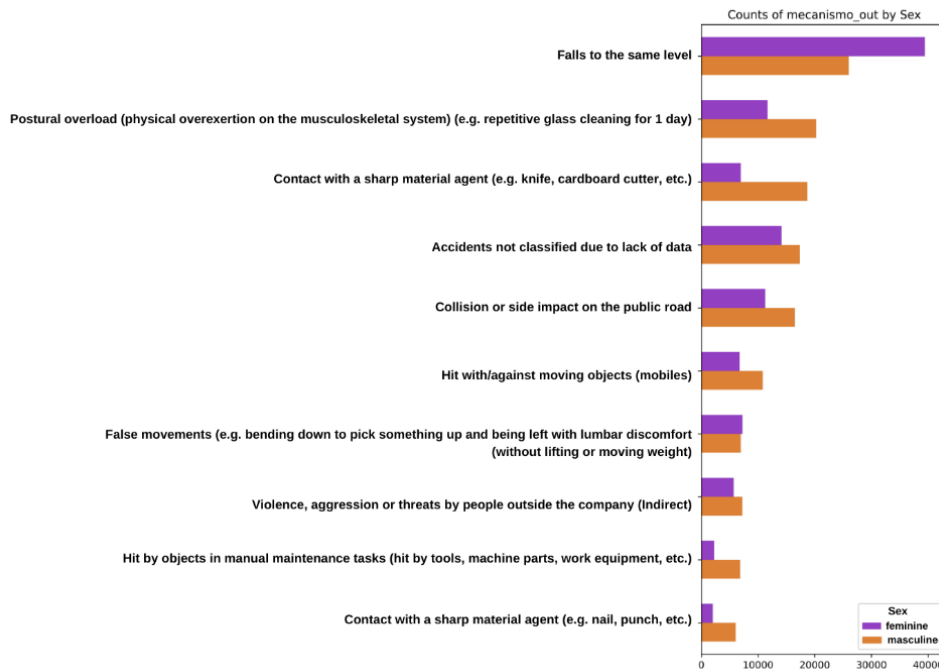
--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

Destacamos que las tareas de limpieza se concentran significativamente en el género femenino. Y, lamentablemente, esto es solo la punta del iceberg. En Chile gran parte de estos trabajos ni siquiera cuentan con un contrato laboral. De esta forma, los números y proporciones encontrados pueden ser mucho mayores al considerar mujeres trabajando sin un contrato formal o sin pago.

Ahora analizando los mecanismos de los accidentes, buscamos los más frecuentes desagregados por sexo, como se muestra en la Figura 2. En nuestro conjunto de datos, los hombres representan el 57% del total, mientras que, analizando solamente caídas, encontramos que 39.000 casos son de mujeres versus 25.000 de hombres, aun cuando hay una mayor predominancia del sexo masculino en el conjunto de datos.



**Figura 2.** Gráfico de los 10 mecanismos de accidentes más comunes de la base de datos desagregando por sexo.

**Discusión y Trabajo Futuro.**

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

Debido a que los datos fueron recolectados por la ACHS con **TEXTO LIBRE**, existe una similitud en muchas entradas, que son categorizadas distintamente, lo que dificulta un análisis más profundo. Esto hace necesario un procesamiento previo para mejorar la categorización y el agrupamiento de ocupaciones similares.

Sobre los datos de ocupaciones semánticamente equivalentes, tuvimos que tomar una decisión al respecto. Esto es debido a que nos encontramos con más de 42.000 entre los 350.000 accidentes registrados, por lo que fue necesario emplear técnicas de Aprendizaje Profundo para poder clasificarlas, resultando en un agrupamiento mucho más conciso y que mejora la forma de poder visualizar la diferencia de género en distintas ocupaciones, ya que el mismo gráfico antes y después de este proceso cambió drásticamente. Es importante recalcar que mismo con el nuevo agrupamiento aún hay trabajo por hacer, ya que algunas ocupaciones quedaron sin clasificar, debido a errores de ingreso de los datos por personal de la ACHS.

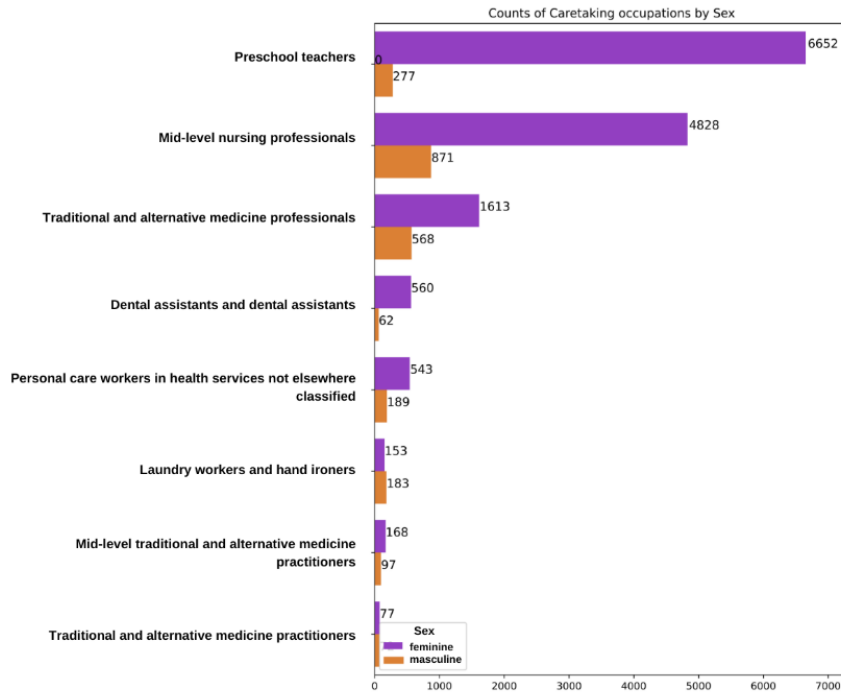
Además de la separación con cuyo puro semántico y categórico, uno de nuestros enfoques fue el trabajo de cuidados, ya que este es uno de los trabajos invisibilizados en nuestra sociedad. Debido a esta invisibilidad del trabajo, gran parte de este se realiza sin remuneración, y nosotros solo tenemos los datos de aquellos que son remunerados. Ejemplos de estas ocupaciones incluyen enfermeras, personal médico etc. Mientras que, de los no remunerados, podemos incluir desde las trabajadoras domésticas que proveen trabajo directo e indirecto de cuidado, hasta personas que tienen que cuidar de miembros de la familia enfermos, niños, etc.

Nuevamente, este tipo de trabajo es mayoritariamente realizado por mujeres, remunerado o no. En la Figura 3, pueden ver un desglose de sexo por las ocupaciones con mayor presencia en el conjunto de datos de accidentes de la ACHS.

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**



**Figura 3.** Gráfico de las 8 ocupaciones de cuidado presentes en la base de datos, desagregadas por sexo.

Esta parte de la revisión fue hecha de manera manual, por lo que es importante tomar en cuenta que puede cumplir con ser parte de la categoría de trabajo de cuidados, pero no fueron consideradas por algún posible error en el análisis. La mayoría de estas ocupaciones corresponde a áreas de la salud, pero la más común es la de profesores/as de preescolar.

Como resultado proporcionado por el conjunto de datos, el 4% de los accidentes registrados estaban relacionados con ocupaciones de cuidado. De este 4%, el 86% de los trabajadores son mujeres. Teniendo en cuenta que estamos analizando solamente ocupaciones pagadas, y que aproximadamente el 46.7% de las mujeres de países de renta media se identifican ocupando un trabajo de cuidado no remunerado (Laura Addati, 2019), la importancia de este análisis radica en el hecho de que, estudiando lo que ocurre en estos grupos de ocupación más precisamente, podemos

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

acercarnos más a entender y resolver problemas que están ocurriendo en mujeres que no reciben pago por su fuerza laboral en Chile.

En cuanto al uso de Aprendizaje Profundo para el mapeo de ocupaciones desde un marco ontológico, en nuestros continuos esfuerzos, hemos trabajado con datos de la ACHS para adaptar y refinar los datos ocupacionales. Para clasificar más de 42,000 ocupaciones escritas en texto libre, emplearemos técnicas de procesamiento de lenguaje natural como el "**WORD EMBEDDING**", que asigna un vector a cada palabra en el vocabulario. Utilizaremos un Modelo de Lenguaje de Gran Tamaño (LLM, por sus siglas en inglés) para obtener los embeddings de clasificaciones estandarizadas y títulos de trabajo. Posteriormente, el algoritmo de similitud del coseno servirá como un criterio matemático, cuantificando la afinidad entre el vector de cada ocupación y las clasificaciones estandarizadas. Al lograr estas puntuaciones de similitud, se empleará un riguroso conjunto de métricas, incluyendo un porcentaje de similitud base, para asegurar la calidad de la asignación ocupacional.

Desde una perspectiva de futuro, tendremos una mejor agrupación de ocupaciones, permitiendo una desagregación más comprensiva por género. Esto permitirá la identificación de problemas relacionados con cada ocupación y la toma de medidas dentro de estos accidentes, haciendo que el trabajo ya realizado sea de impacto a la comunidad.

#### **4. Conclusiones**

Para concluir, dentro de los resultados más relevantes de la investigación, cabe destacar que los accidentes laborales desagregados por género, un 57,3% son protagonizados por hombres, por el contrario, un 42,7% son mujeres. A partir de esta diferenciación, también se analizó la relación entre la ocupación laboral y el accidente, donde operario de máquinas es el trabajo más común con accidentes para el género masculino. Por otra parte, para el género femenino la ocupación más común es de labores de limpieza. Por otro lado, el género también guarda relación con el mecanismo del accidente, donde "caídas de un mismo nivel" es el mecanismo con mayor representación en los datos tanto en hombres como mujeres. Otro aspecto relevante, es que del 4% de los accidentes relacionados con una labor de cuidado, 86% son mujeres. A partir de esto, la hipótesis de que la

--	--	--





**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

desagregación de accidentes por género es relevante para entender cómo se accidentan los trabajadores es verificada. Por último, las implicancias de este trabajo radican en que al comprender cómo se accidentan hombre y mujeres, es posible tomar medidas laborales que consideren estas diferencias por género.

### **Agradecimientos**

Agradecemos al Instituto Milenio Fundamento de los Datos por su colaboración en el proceso de la creación de la vídeo presentación del paper.

También agradecer, a la Asociación Chilena de Seguridad por facilitarnos los datos de los accidentes.

### **GLOSARIO**

1. **TEXTO LIBRE:** Textos o escritos sin restricciones de formato, estilo o contenido, utilizados comúnmente en el procesamiento de lenguaje natural.
2. **WORD EMBEDDING:** Representación de palabras en forma de vectores numéricos, facilitando que las máquinas procesen y comprendan el lenguaje humano.
3. **LLM (Large Language Models):** Modelos de lenguaje de gran escala que utilizan algoritmos avanzados para entender, generar y traducir texto en varios idiomas.
4. **EMBEDDINGS:** Término general para técnicas que convierten datos complejos (como texto o imágenes) en vectores numéricos para facilitar su procesamiento por algoritmos.
5. **DATAFRAME:** Un DataFrame es una estructura de datos tabular bidimensional en la que los datos se organizan en filas y columnas, similar a una hoja de cálculo, y se utiliza comúnmente en programación para el análisis de datos.
6. **NLP (Natural Language Processing):** Rama de la inteligencia artificial que se ocupa de la interacción entre las computadoras y el lenguaje humano. Su objetivo es permitir que las máquinas

--	--	--



**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
**ESCUELA DE INGENIERÍA**  
**Dirección de Investigación e Innovación**  
**Programa IPre de Investigación en Pregrado**

comprendan, interpreten y generen texto de manera similar a como lo hacen los humanos, facilitando la comunicación entre humanos y computadoras a través del lenguaje natural.

7. **MECANISMO:** Es la forma en que se produce el contacto entre el accidentado y el agente, es decir la forma cómo ocurre el accidente.

### Referencias

- Biswas A., Harbin S., Irvin E., Johnston H., Begum M., Tiong M., Apedaile D., Koehoorn M., y Smith P. (2022). *Differences between men and women in their risk of work injury and disability: A systematic review. American journal of industrial medicine*, 65(7):576–588.
- Addati L., Cattaneo U., Esquivel V. y Valarino I. 2019. *El trabajo de cuidados y los trabajadores del cuidado para un futuro con trabajo decente.*  
[https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms\\_737394.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_737394.pdf)
- Instituto Nacional de Estadísticas de Chile. (2018). *CIUO 08.CL Clasificador chileno de ocupaciones.* <https://www.ine.gob.cl/docs/default-source/buenas-practicas/clasificaciones/ciuo/clasificador/ciuo-08-cl.pdf>.
- Organización Internacional del Trabajo. (30 de enero de 2005). *Clasificador internacional uniforme de ocupaciones.* <https://www.ilo.org/public/spanish/bureau/stat/isco/index.htm>

--	--	--