

Improving accuracy and reducing hallucinations in artificial intelligence (AI) with linked data

Rafael Castillo Guerrero
Francisco Garrido Sandoval



UNIVERSIDAD DE CHILE
Dirección de Servicios de
Información y Bibliotecas (SISIB)

30
AÑOS

 CC BY-NC-SA 4.0



4TH ARTIFICIAL INTELLIGENCE AND
LIBRARIES SYMPOSIUM: TRANSFORMING
INFORMATION ACCESS AND DISCOVERY

Part 1: Introduction



AI is the theory and development of computer systems able to perform tasks normally requiring **human intelligence**.

*La **IA** es la teoría y el desarrollo de sistemas informáticos capaces de realizar tareas que normalmente requieren **inteligencia humana**.*

Google Cloud Tech. (mayo 2023). Introduction to Generative AI [Video]. YouTube.
<https://www.youtube.com/watch?v=G2fqAlgmoPo>

Definitions of Artificial Intelligence

Artificial Intelligence (AI) is the field of study focused on the design and **development of algorithms** and computational systems, **capable of emulating human cognitive abilities**, such as machine learning, natural language processing, sensory perception, and decision-making. It draws upon disciplines like computer science, logic, psychology, and robotics, aiming to create machines that can operate autonomously or assist humans in complex tasks, enhancing their efficiency and ability to perform specific activities.

La inteligencia artificial (IA) es el campo de estudio que se centra en el diseño y desarrollo de algoritmos y sistemas computacionales capaces de emular capacidades cognitivas humanas, como el aprendizaje automático, el procesamiento del lenguaje natural, la percepción sensorial y la toma de decisiones. Se basa en disciplinas como la ciencia de la computación, la lógica, la psicología y la robótica, y busca crear máquinas que puedan operar de manera autónoma o asistir a los humanos en tareas complejas, mejorando su eficiencia y capacidad para realizar actividades específicas.

Importance of addressing these issues

Addressing accuracy and hallucinations in artificial intelligence (AI) is crucial for **ensuring the trust**, reliability, and fairness of these technologies. It **enhances user experience**, has a positive economic impact, and encourages the advancement of AI research. By solving these issues, we ensure the safe and effective integration of AI into society, maximizing its benefits while minimizing ethical and practical risks.

Importancia de abordar estos problemas.

*Abordar la precisión y las alucinaciones en la inteligencia artificial (IA) es esencial para **garantizar la confianza**, la fiabilidad de estas tecnologías. **Mejora la experiencia del usuario**, tiene un impacto económico positivo y fomenta el avance de la investigación en IA. Al solucionar estos problemas, se asegura una integración segura y efectiva de la IA en la sociedad, maximizando sus beneficios y minimizando los riesgos éticos y prácticos.*

Specific challenges AI faces in terms of accuracy and hallucinations.

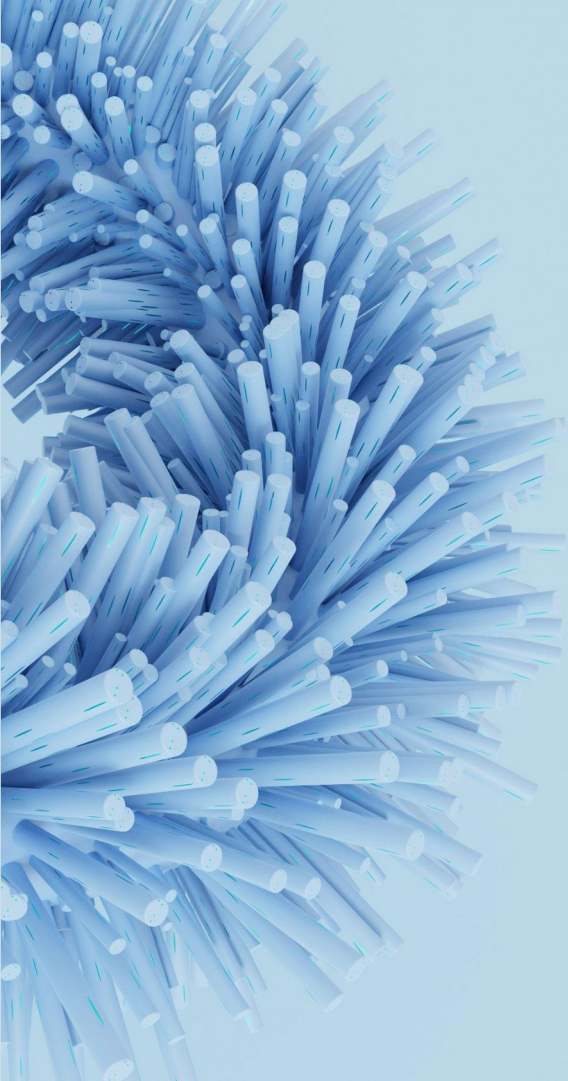
Accuracy in AI indicates how accurate the results of a model are compared to what is expected, essential for reliable predictions and classifications. It is crucial because of:

- **Reliability:** It ensures reliability in critical areas such as medicine, autonomous navigation, and finance.
- **Decision making:** It significantly impacts the effectiveness of decisions in business, judicial, and public sectors.
- **Safety:** It is vital to avoid errors in applications that compromise public and personal safety.
- **Efficiency and cost:** It influences operational efficiency and can reduce costs by decreasing the need for human corrections.
- **User trust:** It increases acceptance and use of AI technologies by improving their perceived

Desafíos específicos que enfrenta la IA en términos de precisión y alucinaciones.

La precisión en IA indica cuán acertados son los resultados de un modelo respecto a lo esperado, esencial para predicciones y clasificaciones confiables. Es crucial por:

- **Confiabilidad:** Garantiza la fiabilidad en áreas críticas como la medicina, la navegación autónoma y las finanzas.
- **Toma de decisiones:** Impacta significativamente en la efectividad de las decisiones en sectores empresariales, judiciales y públicos.
- **Seguridad:** Es vital para evitar errores en aplicaciones que comprometen la seguridad pública y personal.
- **Eficiencia y costo:** Influencia la eficiencia operacional y puede reducir costos al disminuir la necesidad de correcciones humanas.
- **Confianza del usuario:** Eleva la aceptación y uso de tecnologías de IA al mejorar su percepción de fiabilidad.



Examples of AI hallucinations

AI hallucinations can take many different forms. Some common examples include:

Incorrect predictions: An AI model may predict that an event will occur when it is unlikely to happen. For example, an AI model that is used to predict the weather may predict that it will rain tomorrow when there is no rain in the forecast.

False positives: When working with an AI model, it may identify something as being a threat when it is not. For example, an AI model that is used to detect fraud may flag a transaction as fraudulent when it is not.

False negatives: An AI model may fail to identify something as being a threat when it is. For example, an AI model that is used to detect cancer may fail to identify a cancerous tumor.

Examples of AI hallucinations

R You
wich was the first feaure-length animated film?

ChatGPT
The first feature-length animated film is "Snow White and the Seven Dwarfs," produced by Walt Disney Productions and released in 1937. It's widely considered a groundbreaking film in the history of animation, as it was the first full-length cel-animated feature in motion picture history, as well as the first full-length animated feature film ever made. The film was a major success, both critically and commercially, setting the stage for the future of animated feature films.

R You
what do you know about the early El Apostol film?

ChatGPT
"El Apóstol" (1917) is considered the first animated feature film in the history of cinema, predating Disney's "Snow White and the Seven Dwarfs" by two decades. Created by Italian-Argentinian cartoonist Quirino Cristiani, "El Apóstol" was a satirical film that used cutout animation to tell its story, a technique where characters and scenes are crafted from paper cutouts and moved frame by frame to create the illusion of motion.

Challenges in accuracy

1. **Data quality:** can affect the accuracy of AI, leading to errors in applications such as recommender systems or medical diagnostics, for example, raising issues of fairness, among others, due to false results.
2. **Overfitting and generalization:** Overfitting occurs when a model is trained too well on the training data, losing the ability to generalize to new data.
3. **Practical implication:** Systems suffering from overfitting may fail in real-world situations, where they encounter scenarios not seen during training.

Desafíos en la precisión

1. **Calidad** de los datos: puede afectar a la precisión de la IA, provocando errores en aplicaciones como los sistemas de recomendación o los diagnósticos médicos, por ejemplo, planteando problemas de imparcialidad, entre otros, debido a los falsos resultados.
2. **Sobreajuste y generalización:** se produce cuando un modelo se entrena demasiado bien en los datos de entrenamiento, perdiendo la capacidad de generalizar a nuevos datos.
3. **Sobreajuste y generalización:** El sobreajuste se produce cuando un modelo se entrena demasiado bien en los datos de entrenamiento, perdiendo la capacidad de generalizar a nuevos datos.

Challenges in hallucinations

1. **Generation of unexpected content:** Generative models, like those used for creating text, images, and audio, sometimes "hallucinate" irrelevant or nonexistent content.
2. **Practical implication:** This can lead to the generation of fake news, inappropriate content on social media, and errors in customer support and assistance systems, undermining trust in these technologies.
3. **Lack of deep understanding:** AI often lacks a true understanding of the content it generates or analyzes, which can lead to significant hallucinations in its outputs.
4. **Practical implication:** In legal or financial domains, this can result in the generation of documents or financial analyses with serious errors, impacting critical decisions.

Desafíos en Alucinaciones

Generación de contenido inesperado: Los modelos generativos, como los utilizados para la creación de texto, imágenes y audio, a veces "alucinan" contenido irrelevante o inexistente.

Implicación práctica: Esto puede conducir a la generación de noticias falsas, contenido inapropiado en medios de comunicación social, y errores en sistemas de asistencia y soporte al cliente, socavando la confianza en estas tecnologías.

Falta de comprensión profunda: La IA a menudo carece de una verdadera comprensión del contenido que genera o analiza, lo que puede llevar a alucinaciones significativas en sus salidas.

Implicación práctica: En el ámbito legal o financiero, esto puede resultar en la generación de documentos o análisis financieros con errores graves, impactando decisiones críticas.

General implications

- **Ethics and trust:** Accuracy errors and hallucinations can undermine public trust in AI, raising ethical questions about its use
- **Regulation and legislation:** These challenges drive the need for stricter regulations and ethical considerations in the development and deployment of AI technologies, which in turn can influence the innovation and adoption of these technologies.
- **Technological development:** There is constant pressure on researchers and developers to improve AI methodologies, including model training techniques, data processing, and evaluation methods to address these issues of accuracy and hallucinations.
- **Ética y confianza:** Los errores de precisión y las alucinaciones pueden minar la confianza del público en la IA, lo que plantea cuestiones éticas sobre su uso.
- **Regulación y legislación:** Estos retos impulsan la necesidad de una normativa más estricta y de consideraciones éticas en el desarrollo y despliegue de las tecnologías de IA, lo que a su vez puede influir en la innovación y adopción de estas tecnologías.
- **Desarrollo tecnológico:** Existe una presión constante sobre los investigadores y desarrolladores para que mejoren las metodologías de IA, incluidas las técnicas de entrenamiento de modelos, el procesamiento de datos y los métodos de evaluación para abordar estos problemas de precisión y alucinaciones.



Part 2: Linked open data and Artificial intelligence



What is Linked Data?


Linked Data is an approach and methodology for **publishing, connecting, and linking data on the web in a structured and meaningful way**. It was proposed by Tim Berners-Lee, the creator of the World Wide Web, as a way to extend the principles of the web so that not only humans can access documents, but also data in a smarter and automated manner.

Linked Data is a method of publishing **RDF data** on the Web and **interconnecting data** between different **data sources**.

¿Qué son los datos enlazados?

Linked Data es un enfoque y una metodología para **publicar, conectar y enlazar datos en la web de forma estructurada y significativa**. Fue propuesto por Tim Berners-Lee, creador de la World Wide Web, como una forma de ampliar los principios de la web para que no solo los humanos puedan acceder a los documentos, sino también a los datos de una manera más inteligente y automatizada.

Linked Data es un método de **publicación de datos RDF en la Web** y de **interconexión** de datos entre distintas fuentes de datos.



Because linked data is a potential solution

Enriched context: Linked data provide **additional context** around the information being processed. This means that AI models can have a better understanding of the relationships and interdependencies among the data, which reduces the likelihood of generating responses or content that have no basis in reality.

Improved machine learning: By having access to a richer and more structured set of interconnected data, **machine learning algorithms can be trained more effectively**. This results in more accurate models that are capable of making more accurate inferences based on complex patterns and connections in the data.

Validation and verification: Linked data can facilitate the validation and verification of information generated by AI systems. By **comparing the outputs of AI with a vast repository of interconnected knowledge**, it's possible to identify and correct errors or hallucinations before they are presented to users.

Integration of linked data in AI : prototype



Datos abiertos enlazados en repositorio académico

<https://doi.org/10.7764/tesisUC/ING/2213>

⁹The study applied semantic techniques to data from the University of Chile's Repository, specifically in the forestry area, to enhance the visualization and quality of academic metadata through a semantic system. It includes a review of the state of the art in semantic technology and the development of a proprietary data model for Linked Open Data, using Open Refine to transform metadata into RDF structures.



Prototype based on thesis project - 2017



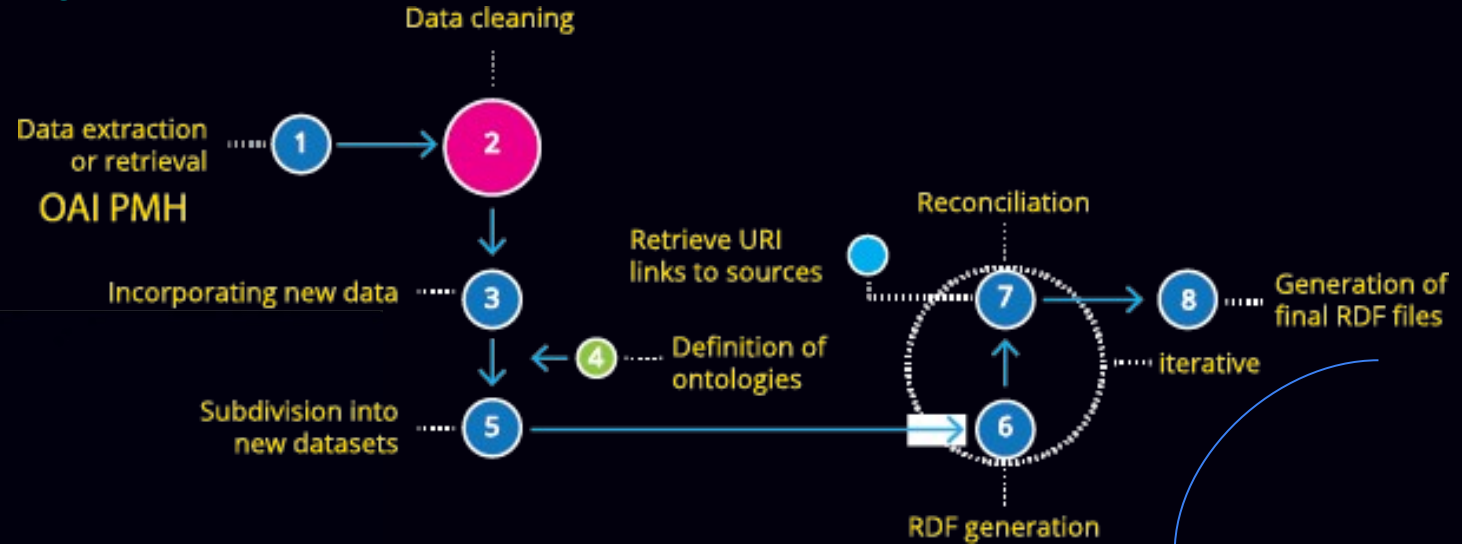
Integration of linked data in AI : prototy

Retrieval Augmented Generation (RAG) boosts AI text generation by integrating pre-trained language models with real-time data search from external sources. Before responding or generating content, the AI queries a database for relevant information, which it then incorporates to enhance its output. This approach results in more accurate, informative, and fact-based responses, improving upon the limitations of models trained only on initial datasets.

Retrieval Augmented Generation (RAG) potencia la generación de textos por IA integrando modelos lingüísticos preentrenados con la búsqueda de datos en tiempo real a partir de fuentes externas. Antes de responder o generar contenido, la IA consulta una base de datos en busca de información relevante, que luego incorpora para mejorar su resultado. Este enfoque da lugar a respuestas más precisas, informativas y basadas en hechos, mejorando las limitaciones de los modelos entrenados sólo en conjuntos de datos iniciales.

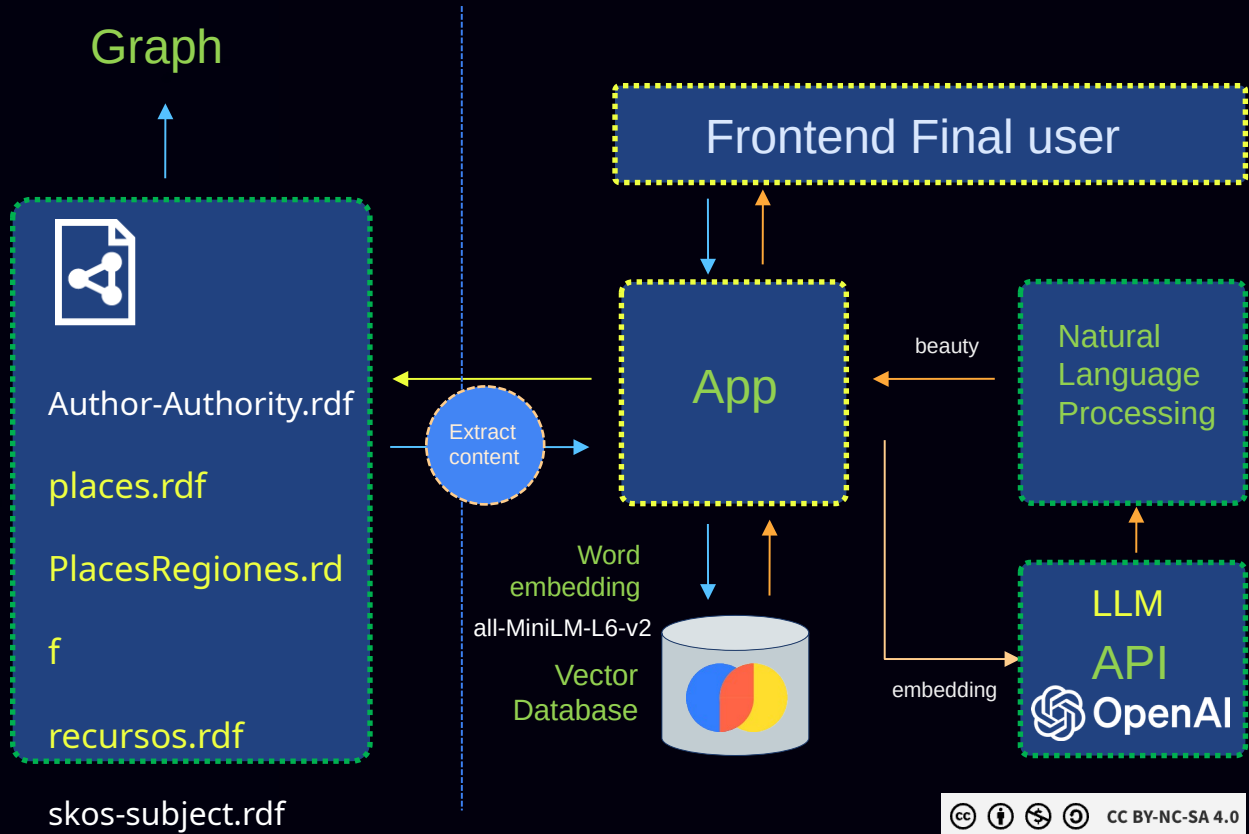
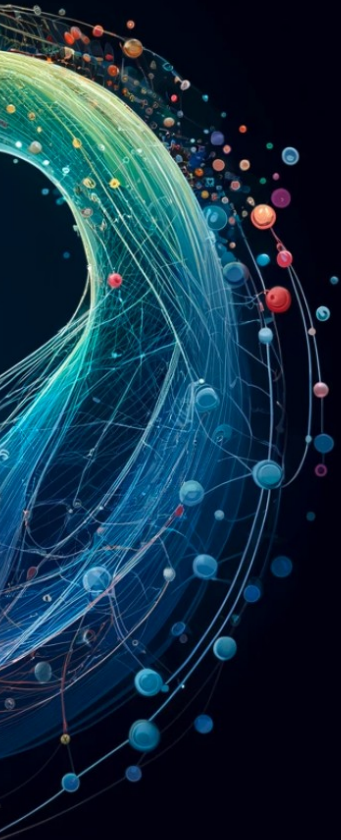


Integration of linked data in AI: process for generation of RDF file



Author-Authority.rdf
places.rdf
PlacesRegiones.rdf
recursos.rdf
skos-subject.rdf

Integration of linked data in AI





Tokens, english v/s spanish

The future of artificial intelligence (AI) alongside linked data presents a promising development field for libraries, positioning them as nerve centers of advanced knowledge and personalized information services. By integrating AI with linked data, libraries will be able to offer profoundly enriched user experiences, facilitating more intuitive and contextualized access to vast collections of resources.

El futuro de la inteligencia artificial (IA) junto con los datos vinculados abre un campo de desarrollo prometedor para las bibliotecas, posicionándolas como centros neurálgicos de conocimiento avanzado y servicios de información personalizados. Integrando la IA con datos vinculados, las bibliotecas podrán ofrecer experiencias de usuario profundamente enriquecidas, facilitando un acceso más intuitivo y contextualizado a vastas colecciones de recursos.

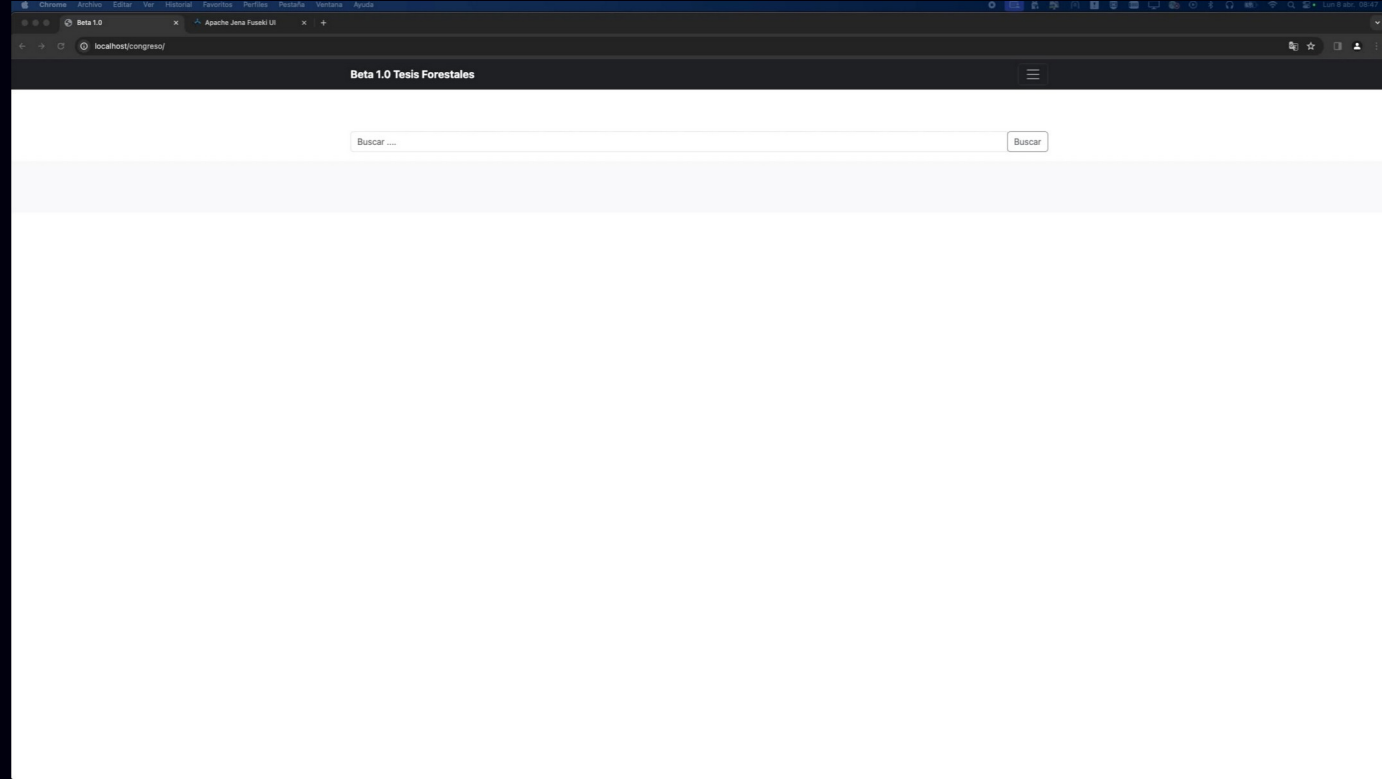
Word embeddings



By default, Chroma converts the text into the embeddings using all-MiniLM-L6-v2



Prototyping case studies. Practical example



The background features a glowing Earth with a digital network overlay of nodes and lines. The text is centered over the image. The left side of the image has a blue and green color scheme, while the right side has a warmer, yellow and orange glow.

Parte 3:
Superación de
obstáculos
y visión de futuro

Part 3:
Overcoming
obstacles
and future vision

Challenges and considerations: how to overcome obstacles

1. **Interdisciplinary collaboration:** Bringing together experts from various fields such as computer science, data science, AI, and domain-specific areas to ensure a comprehensive understanding of both the technical and practical aspects of linked data and AI.
2. **Quality data management:** Ensuring the integrity, accuracy, and relevance of the data through robust data management practices. This includes regular data cleaning, validation, and enrichment processes.
3. **Standardization and interoperability:** Adopting and adhering to established standards for data representation, storage, and exchange, such as RDF (Resource Description Framework), to enhance interoperability between different systems and datasets.
4. **Scalability and performance optimization:** Designing systems that are scalable and can efficiently process and analyze large volumes of interconnected data. This may involve optimizing algorithms, utilizing cloud computing resources, and leveraging distributed computing techniques.



Challenges and considerations: how to overcome obstacles

- 
5. **Ethical considerations and privacy:** Addressing ethical concerns and ensuring the privacy and security of the data being used. This includes implementing appropriate data protection measures and considering the ethical implications of AI decisions and predictions.
 6. **Continuous Learning and adaptation:** Incorporating mechanisms for continuous learning and adaptation to new data, evolving standards, and emerging technologies. This can be achieved through machine learning models that adapt over time and systems designed for easy integration of new data sources and technologies.
 7. **User-centric design:** Focusing on the needs and experiences of the end-users of the technology. This includes developing intuitive interfaces, providing clear explanations of AI decisions, and ensuring that the technology addresses real-world problems effectively.
 8. **Regulatory compliance:** Ensuring that the solution complies with all relevant laws and regulations, particularly those related to data protection and privacy, such as GDPR in Europe.

The future of AI with linked data and libraries.

1. The integration of AI and linked data offers libraries the opportunity to become **advanced knowledge centers**, enhancing user access to information with personalized services.

2. This technology enables **better information management and discovery** in libraries through intelligent recommendation systems tailored to user preferences, facilitating **deeper research connections**.

3. AI and linked data will improve **digital preservation and collection curation**, automating the classification of materials and uncovering historical patterns in topics and disciplines.

4. Adopting AI and linked data technologies **will strengthen libraries** roles in promoting democracy and education by ensuring equitable information access and supporting a more **informed society**.

1. La integración de la IA y los datos enlazados ofrece a las bibliotecas la oportunidad de convertirse en **centros avanzados de conocimiento**, mejorando el acceso de los usuarios a la información con servicios personalizados.

2. Esta tecnología permite **mejorar la gestión y el descubrimiento de información** en las bibliotecas a través de sistemas inteligentes de recomendación adaptados a las preferencias de los usuarios, lo que facilita unas conexiones de **investigación más profundas**.

3. La IA y los datos enlazados mejorarán la preservación digital y la conservación de colecciones, automatizando la clasificación de materiales y **descubriendo patrones históricos en temas y disciplinas**.

4. La adopción de tecnologías de IA y datos enlazados **reforzará el papel de las bibliotecas** en la promoción de la democracia y la educación, garantizando un acceso equitativo a la información y apoyando una sociedad más informada.

The integration of linked open data and artificial intelligence in libraries shines in four key areas:

1. **AI and machine learning:** Enhance personalization and efficiency of library services through **recommendation systems and advanced analytics**.
 2. **Privacy and ethics:** Are crucial for responsibly **handling data, protecting user privacy**.
 3. **Open science and access:** Facilitate equitable access to academic and scientific information, **promoting transparency in research**.
 4. **Research services:** Enrich research support with improved data analysis and visualization powered by AI and linked data.
1. **IA y aprendizaje automático:** Mejora la personalización y la eficiencia de los servicios bibliotecarios mediante sistemas de recomendación y análisis avanzados.
 2. **Privacidad y ética:** Son cruciales para el manejo responsable de los datos, protegiendo la privacidad del usuario.
 3. **Ciencia abierta y acceso:** Facilitar el acceso equitativo a la información académica y científica, fomentando la transparencia en la investigación.
 4. **Servicios de investigación:** Enriquecer el apoyo a la investigación con análisis y visualización de datos mejorados mediante IA y datos enlazados.

"We invite researchers, developers, innovators, and all those committed to advancing knowledge to dive into the world of linked open data and artificial intelligence. Actively participate in creating solutions that leverage the power of these resources to tackle **complex challenges**, **improve decision-making**, and promote an environment of transparency and open access. Collaborate on projects that transform the way we access and use information, contributing to a more informed, fair, and **connected society**.

Invitamos a investigadores, desarrolladores, innovadores y a todos aquellos comprometidos con el **avance del conocimiento** a sumergirse en el mundo de los **datos abiertos enlazados** y la **inteligencia artificial**. Participen activamente en la creación de soluciones que aprovechen el poder de estos recursos para **abordar desafíos complejos**, mejorar la toma de decisiones y **promover un entorno de transparencia y acceso abierto**. Colaboren en proyectos que transformen la manera en que accedemos y utilizamos la información, contribuyendo a una sociedad más informada, justa y conectada.



Some interesting links

 [Hugging Face](#) [Modelo para crear embeddings con all-MiniLM-L6-v2](#)


 [Embeddings](#)

 [Vector Database: Chroma](#)

 [Apache Jena](#)

 [Tokenizer](#)

 [OpenAI API](#)

 [Datos abiertos enlazados en repositorio académico](#)
[o](#)

Thanks to

Photo by [Philip Oroni](#) on [Unsplash](#)

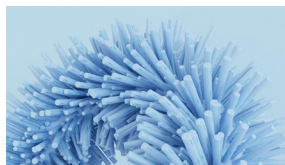


Photo by [Google DeepMind](#) on [Unsplash](#)

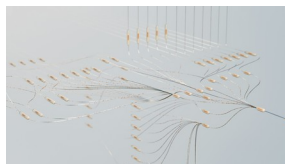


Photo by [SIMON LEE](#) on [Unsplash](#)

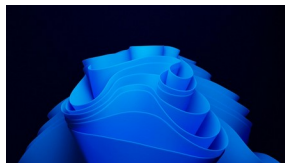


Imagen by DALL.E, prompt:
Rafael Castillo Guerrero
and Francisco Garrido



[De PICOL.](#)

Pictorial Communication Language -

[http://picol.org/.](http://picol.org/)

CC BY 3.0,





Rafael Castillo Guerrero

rafael.castillo@uchile.cl

<https://orcid.org/0009-0004-2990-6364>

<https://www.linkedin.com/in/rafaelcastilloguerrero/>



Francisco Garrido Sandoval

francisco.garrido@uchile.cl

<https://www.linkedin.com/in/fgarridosandoval/>