# E6.1 DEMOSTRADORES Y REPORTES

ESPACIO DE DATOS LINGÜÍSTICO (SER 15/23 OTT)

## Resumen

Este documento describe los demostradores del espacio de datos que implementan casos de uso definidos en el proyecto. Estos casos de uso se apoyan en el espacio de datos para obtener los datos necesarios e incluyen el análisis de afirmaciones científicas, el análisis de historias clínicas y la evaluación de factualidad de los modelos del lenguaje. En esta primera versión del documento se presenta una versión completa del análisis de afirmaciones científicas, mientras que el resto demostradores solo son descritos y se espera que estén implementados en la siguiente versión del documento. Además, se presenta la primera versión del reporte de estadísticas del catálogo del espacio de datos que incluye el número de activos por tipo y lenguaje.

Andrés García Silva Raúl Ortega, Cristian Berrío, Flavio Merenda, José Manuel Gómez Pérez

22 de noviembre 2024 Expert.ai Expert.ai Language Technology Research Lab

Calle Poeta Joan Maragall, 3-5, Escalera Izquierda, Planta 1<sup>a</sup>, Derecha, 28020, Madrid CIF: B-66425513, Inscrita en el Registro Mercantil de Madrid, en el Tomo 44.538, Folio 74, Hoja Número M-784613, Inscripción 1<sup>a</sup>.



# **Revision History**

Revision	Date	Description	Author (Organisation)
0.1	26/11/24	Primera versión completa del documento	Andrés García-Silva
0.1.1	12/12/24	Descarga de datos de conector ELG	Andrés García-Silva
0.1.2	30/12/24	Corrección url repositorio de la demo de afirmaciones	Andrés García-Silva



## Contenido

1	Introd	Introducción4		
2	Demo	ostradores	4	
	2.1	Análisis de afirmaciones científicas	. 5	
	2.1.1	Desarrollo del caso de uso	. 5	
	2.2	Diagnóstico médico	13	
	2.3	Evaluación de factualidad en dominio de seguros	14	
3	Repo	rtes	14	



# 1 Introducción

Este documento describe los resultados iniciales del paquete de trabajo 6 Demostración y reportes. Los objetivos del paquete de trabajo son desarrollar demostradores que usen los servicios y datos del catálogo del espacio de datos lingüístico EDL<sup>1</sup>, y generar reportes periódicos con el estado de los recursos almacenados en catálogo del espacio de datos.

Los demostradores permiten la interacción de los usuarios con el catálogo del EDL por ejemplo, consultando activos disponibles, contratar y negociar activos o descargarlos para utilizarlos en una tarea determinada. Estos demostradores se apoyan en los resultados del proyecto KG4LLM<sup>2</sup> en el marco del proyecto en INESData, donde se está investigando en técnicas para el análisis de diagnóstico médico, evaluación de factualidad de los modelos del lenguaje masivos LLM en dominios como el de los seguros, o el análisis de afirmaciones científicas.

En este entregable se presenta la descripción completa del demostrador de análisis de afirmaciones científicas que incluye el uso del espacio de datos lingüístico para buscar, negociar y descargar los datos y su posterior procesamiento en un notebook de Jupyter con la ayuda de un LLM. Los demostradores de la evaluación de factualidad de LLM y el de diagnóstico médico solo se especifican. El desarrollo de estos dos demostradores se incluirá en la siguiente versión de este entregable.

Los reportes están disponibles en la interfaz del conector de EDL permitiendo que todos los miembros del espacio de datos cuenten con la información actualizada. Estos reportes permiten visualizar en un solo lugar las estadísticas del catálogo del EDL incluyendo el número de activos disponibles por idioma, y por tipología de activos incluyendo Corpora, Recursos léxicos o modelos.

# 2 Demostradores

Los demostradores se apoyan en el espacio de datos lingüístico para negociar y descargar los datos necesarios para los casos de uso. En cada caso de uso los datos son procesados en Jupyter Notebooks para alcanzar el objetivo del caso de uso. Si los datos no forman parte aún del espacio de datos, serán cargados usando la interfaz de usuario de un conector que actúe como un proveedor. Posteriormente el usuario negociará el contrato que aplica a los datos usando un conector que actúa como consumidor de datos. Los datos se descargarán en el entorno local usando la interfaz de usuario del espacio de datos o directamente en el entorno de ejecución del notebook usando el API del conector correspondiente. Al menos un conjunto de datos por tarea se obtiene del EDL y si como resultado de la tarea se genera un conjunto de datos este se subirá y publicará en el EDL.

Para poder interactuar con el EDL es necesario que el usuario tenga las credenciales de acceso a un conector desplegado en el EDL. Con estás credenciales el usuario podrá acceder a la interfaz de usuario del espacio de datos o invocar el servicio de autenticación del EDL usando el servicio RESTful correspondiente desde el notebook. Este servicio retorna un token de acceso que debe pasarse en todas las llamadas que se realicen al API del conector. El API del conector soporta todas las funcionalidades disponibles en la interfaz de usuario del EDL, incluyendo la consulta de activos en el catálogo, y la negociación, transferencia y descarga de activos.

Además, en los demostradores se incluirá el uso de datos disponibles y almacenados en el European Language Grid ELG. Para esto se usará el conector de ELG desarrollado en el paquete de trabajo 5 Servicios de integración. Este conector pone a disposición de los usuarios del EDL los corpus y recursos léxicos disponibles alojados en ELG que tengan una licencia de uso permisiva.

<sup>&</sup>lt;sup>1</sup> https://labdemos.expertcustomers.ai/edl

<sup>&</sup>lt;sup>2</sup> https://labdemos.expertcustomers.ai/kg4llm



A continuación, se describe los demostradores que se van a implementar en el proyecto. El demostrador "Análisis de afirmaciones científicas" está implementado en una primera versión. Los otros dos demostradores se implementarán en la segunda versión de este entregable.

# 2.1 Análisis de afirmaciones científicas

El objetivo de este demostrador es detectar afirmaciones científicas en un texto de entrada y verificarlas con literatura científica de referencia. Para verificar las afirmaciones es necesario descargar e indexar un corpus científico de referencia. La detección y verificación de las afirmaciones se apoya en un LLM.

Caso de uso	Análisis de afirmaciones científicas	
Entradas		
datos en ELG	Pubmed abstracts: Pubmed Corpus for verification (lite) https://live.european-language-grid.eu/catalogue/corpus/12116	
LLM en hugging face	Mejor modelo en <u>https://huggingface.co/spaces/la-leaderboard/la-leaderboard</u> con menos de 4 billones de parámetros.	
Requisitos	<ul> <li>Conector de EDL integrado con ELG</li> <li>Acceso a Hugging Face</li> <li>LLM disponible en Hugging Face</li> </ul>	
Proceso		
<ul> <li>Descargar los conjuntos de datos en ELG desde EDL para español.</li> <li>Indexar los conjuntos de datos siguiendo el algoritmo de ranking BM25.</li> <li>Instruir al LLM para que detecte las afirmaciones en un texto de entrada</li> <li>Para cada afirmación detectada:         <ul> <li>Obtener abstracts similares que se hayan indexado</li> <li>Instruir a un LLM para que verifique la afirmación con los abstracts recuperados El resultado de la verificación es: "Soportado" para los casos en que la afirmación se puede derivar del abstract, "Refutado" en caso contrario, "Información Insuficiente" cuando no se puede tomar una decisión. También se incluye una explicación para la elección del resultado de verificación, y las frase más relacionadas del documento del conjunto de datos indexado.</li> </ul> </li> </ul>		

### 2.1.1 Desarrollo del caso de uso

#### Búsqueda y descarga del recurso lingüístico en el EDL.

Este caso de uso inicia buscando y descargando el recurso lingüístico "Pubmed Corpus for verification (lite)" usando el conector de ELG en el EDL. El usuario debe iniciar sesión en el espacio de datos usando la URL y las credenciales que el promotor del espacio de datos le ha enviado al registrarse en la plataforma. Una vez ha ingresado al sistema, debe buscar el activo correspondiente al recurso lingüístico en el catálogo del EDL. Para esto primero debe acceder a la opción "Corpora" en el menú izquierdo y realizar una búsqueda por una palabra clave (ver Figura 1).



INESData		E Corpora List of all corpora offered in t	he in the INES	SData langu	age data sp	bace catalogue			
	CREATE	Pubmed				٩	CREATE		
	Corpora	Pubmed Corpu						F⊐ Languages	~
Ļ,	Models	Verification Corpus to						≜ Categories	~
E	Lexical resources	English     Science     The http://connector- cl:19194/protocol     Created on 12/19/2024							
			<	1	>				
0	Account								
	My console								

Figura 1. Búsqueda y selección de activo en el catálogo del espacio de datos lingüístico.

Una vez se ha identificado el recurso, el usuario debe negociar el contrato que ha definido el proveedor del activo, en este caso ELG. En la vista del catálogo del EDL se debe clicar sobre el activo a negociar para que se despliegue la información detallada del activo y las opciones de contratación (ver Figura 2). El usuario puede revisar las condiciones de los diferentes contratos y negociar el que más le interese. La negociación se inicia al clicar sobre el botón negociar del contrato correspondiente.

INESData	← BACK				
CREATE	Pubmed Corpus fo Verification Corpus to check claims in biom	r verification (lite) medicine containing abstracts from pubmed.			
🖹 Corpora	CO Originator: http://connector-c1:19194/protocol				
🙏 Models	Content Type: application/json				
Lexical resources	CONTRACTS				
	Contract Offer 1	Contract Offer 2			
	Permissions •	Permissions			
	Prohibitions	Prohibitions			
Account	Obligations	Obligations			
🖺 My console	NEGOTIATE	NEGOTIATE			

Figura 2. Negociación del contrato asociado al recurso lingüístico



Inmediatamente el sistema lleva al usuario a la vista de las negociaciones que ha iniciado en la opción "Mi consola" (Figura 3). Inicialmente la negociación estará en estado pendiente, pero en cuestión de menos de 1 minuto debería cambiar a estado completado, ya que las políticas definidas en el EDL son automáticamente evaluadas como satisfechas por definición. Para actualizar la vista de negociaciones es necesario clicar en otra opción y regresar a la vista negociaciones.

INESData	My console	Negotiations Here you can review all your contract offer negotiations.				
CREATE	Overview	PROPOSALS SENT PROPOSALS RECE	IVED			
Corpora	Assets	ld	Asset name	Creation date	Status	
🙏 Models	Contracts	c94a0e5b-acb5-432c-87df- 427c84e10477		12/12/2024, 11:22:50 AM	Pending	
E Lexical resources	Negotiations	5fefe558-8712-4781-825d- e30c7819fa5b	space-ideas-plus (Corpora)	12/10/2024, 12:20:34 PM	Completed •	
	Transfer Histo	ce6fed21-7f60-44dd-a8c4- 48a8fea04692	European Clinical Case Corpus (Corpora)	12/5/2024, 12:35:49 PM	Completed	
		224f3bda-e405-42fd-8685- 020309ea647f	space-ideas-plus (Corpora)	12/5/2024, 12:31:35 PM	Completed <b>O</b>	
		6bdd0783-a34c-41af-aa66- b9446f62d41e	space-ideas-plus (Corpora)	12/5/2024, 12:17:31 PM	Completed	
		ee95071a-0655-44f5-acc1- 9efad79f19d0	European Clinical Case Corpus (Corpora)	12/4/2024, 5:54:41 PM	Completed <b>O</b>	
		Oec4d9b3-fce8-4f17-906f-f87e370181c7	SQuAD2.0 (Corpora)	12/4/2024, 5:49:30 PM	Completed <b>O</b>	
Account		80c9a26e-23a1-4a00-bdba- 184368d693eb	Space Ideas (Corpora)	12/4/2024, 5:43:11 PM	Completed <b>O</b>	
My console			Items per page:	10 👻 1-8 of 8		

Figura 3. Estado de la negociación

Cuando la negociación se ha completado, se puede iniciar el proceso de transferencia del activo. Este proceso tiene por objetivo generar un endpoint para acceder al recurso lingüístico. Para solicitar la transferencia del activo hay que clicar en la opción ver (icono de ojo en la última columna de la tabla de negociaciones) que corresponde a la negociación. En la pantalla que se despliega se clica en el botón Transfer.

	My console	Here you can review	all your contract offer negotiations.			
		PROPOSALS SENT	PROPOSALS RECEIVED			
CREATE	Overview	Id	Asset name Creat	ion date Statu	JS	
Corpora	Negotiation 79e3e8e	a-d5b5-47d2-8	3be8-7cb39cb52408	×	ending	0
E Lexical reso	connector-c1 Asset owner Pubmed Corpus for verifica	ation (lite)	Contract		ompleted	0
	Asset name 12/19/2024, 4:30:53 PM Proposal creation date		Policy     Contract policy	•	ompleted	0
	COMPLETED Status			2	ompleted	0
	-	TRANSFE	R	2	ompleted	o
			(Corpora)			
Account		224f3bda-e405- 42fd-8685- 020309ea647f	space-ideas- 12/5/2 plus (Corpora) 12:31:	2024, 35 PM	ompleted	0
My console		6bdd0783-a34c- 41af-aa66- b9446f62d41e	space-ideas-         12/5/2           plus (Corpora)         12:17:3	2024, 31 PM	ompleted	O

Figura 4. Iniciar transferencia del activo



Finalmente, sí la transferencia ha sido exitosa se habilita el botón download (ver Figura 5) que permite descargar el recurso lingüístico al entorno local del usuario.

INESData	My console	Here you can review	all your contract offer nego	tiations.		
	,	PROPOSALS SEN	T PROPOSALS RECEIV	ED		
CREATE	Overview	Id	Asset name	Creation date	Status	
🖃 Corpora	Negotiation 79e3e	8ea-d5b5-47d2-8	8be8-7cb39cb524	08 ×	Pending	0
Lexical resou	connector-c1 Asset owner Pubmed Corpus for veri	ification (lite)	Contract		Completed	Ø
	Asset name 12/19/2024, 4:30:53 PM Proposal creation date		Policy Contract policy	•	Completed	0
	COMPLETED Status				Completed	Ø
	TRANSFE	R	DOWNLO	DAD	Completed	0
			(Corpora)			
Account		224f3bda-e405- 42fd-8685- 020309ea647f	space-ideas- plus (Corpora)	12/5/2024, 12:31:35 PM	Completed	0
My console		6bdd0783-a34c- 41af-aa66-	space-ideas- plus (Corpora)	12/5/2024, 12:17:31 PM	Completed	o

Figura 5. Descarga del recurso lingüístico.

#### Notebook para el análisis de afirmaciones científicas.

En este demostrador incluye un notebook que describe el flujo de trabajo para el análisis de afirmaciones en textos con contenido científico enfocado en el ámbito biomédico (ver Figura 6). Estas afirmaciones serán inferidas a partir del texto y reformuladas mediante un LLM con el objetivo de funcionar como unidades atómicas de información, para a continuación ser analizadas por ese mismo modelo, generando un reporte de factualidad de las mismas. Todas las afirmaciones serán comparadas con una colección de documentos confiables, en este caso abstracts procedentes de Pubmed.

El notebook está disponible en el repositorio de demostradores del proyecto, en la siguiente url:

https://github.com/oeg-upm/inesdata-espacio-linguisticodemostradores/blob/main/health\_claims.ipynb



#### **Health Claims**



In this notebook we will show a workflow to obtain a report analyzing collection of biomedicine claims stated in different type of texts using a single instance of a

First we will use the LLM to extract the claims from the objetive text. After that, we will find related abstracts from a dataset of trustworthy sources, such as Pubmed. Finally, we will ask again the LLM for a report for each pair of claim-abstract, giving a label (SUPPORT, CONTRADICT or NOT ENOUGH INFORMATION - NEI), the rationale for that decision and the most related sentence from the verified abstract.

#### Figura 6. Cabecera del notebook.

Después de cargar las librerías y funciones auxiliares necesarias para la ejecución del notebook, se inicializa el modelo, el tokenizador (ver Figura 7). La selección del modelo se ha realizado teniendo en cuenta las limitaciones de hardware con las que cuentan entornos colaborativos como Google Colab, donde solo es posible instanciar en inferencia LLMs de no más de 4 billones de parámetros. Es por ello que la elección fue Llama3.2-3B-Instruct<sup>3</sup>, con 3 billones de parámetros y capacidades similares a las de modelos de un tamaño superior.

Por su parte, el conjunto de datos de verificación ha sido generado usando una colección reducida de abstract procedentes de Pubmed, junto a sus títulos y DOIs. El archivo jsonl con el que se carga este conjunto de datos se encuentra en el activo "Pubmed Corpus for verification (lite)" en el espacio de datos lingüístico. Este fichero se descargó en la sección anterior desde el espacio de datos al entorno local del usuario. Es necesario transferir el fichero a la misma ubicación donde se esté ejecutando el notebook.

<sup>&</sup>lt;sup>3</sup> https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct



### 2. Initialize model, tokenizer and verification data

First, we instantiate the model, the tokenizer and the sampling parameters we will use in execution time to obtain the claim analysis report. You can tune the top\_p and temperature values as you please.

m	[6]	seed = 42			
		<pre>model = LLM(model="meta-llama/Llama-3.2-3B-Instruct", dtype="half", max_model_len=16384) tokenizer = AutoTokenizer.from_pretrained("meta-llama/Llama-3.2-3B-Instruct") sampling_params = SamplingParams(     temperature=0.7,     top_p=1.0,     max_tokens=1024,     logprobs=0, )</pre>			
	₹	config.json: 100% 878/878 [00:00<00:00, 51.3kB/s]			
		WARNING 12-12 17:46:34 config.py:1865] Casting torch.bfloat16 to torch.float16. INFO 12-12 17:46:45 config.py:350] This model supports multiple tasks: {'generate', 'embedding'}. Defaulting to 'generate'. INFO 12-12 17:46:45 llm_engine.py:249] Initializing an LLM engine (v0.6.4.post1) with config: model='meta-llama/Llama-3.2-3B-Instr			
		tokenizer_config.json: 100% 54.5k/54.5k [00:00<00:00, 1.82MB/s]			
		tokenizer.json: 100% 9.09M/9.09M [00:00<00:00, 23.1MB/s]			
		special_tokens_map.json: 100% 296/296 [00:00<00:00, 5.34kB/s]			
		generation_config.json: 100% 189/189 [00:00<00:00, 3.42kB/s]			
		INFO 12-12 17:46:50 selector.py:261] Cannot use FlashAttention-2 backend for Volta and Turing GPUs. INFO 12-12 17:46:50 selector.py:144] Using XFormers backend. INFO 12-12 17:46:52 model_runner.py:1072] Starting to load model meta-llama/Llama-3.2-3B-Instruct INFO 12-12 17:46:53 weight_utils.py:243] Using model weights format ['*.safetensors']			
		model-00001-of-00002.safetensors: 100% 4.97G/4.97G [01:58<00:00, 42.7MB/s]			
	model-00002-of-00002.safetensors: 100% 1.46G/1.46G [00:34<00:00, 46.3MB/s]				
		model.safetensors.index.json: 100% 20.9k/20.9k [00:00<00:00, 904kB/s]			
		Loading safetensors checkpoint shards: 100% Completed   2/2 [00:32<00:00, 14.88s/it]			
		INFO 12-12 17:49:25 model_runner.py:1077] Loading model weights took 6.0160 GB			

Figura 7. Inicialización del modelo y tokenizador.

A continuación, se selecciona el texto que se desea analizar (ver Figura 8). En este caso, hemos elegido un texto relacionado con el factor genético de la obesidad proveniente de una revista científica.

### 3. Input text

In the following cell you can select the text to analyze:

S C	)	text = """Do Genes Have a Role in Obesity?	
		In recent decades, obesity has reached epidemic proportions in populations whose	environments promote physical inactivity and

#### Figura 8. Selección del texto de entrada.

El primer paso del análisis es la extracción de afirmaciones del texto. Esta extracción se realiza mediante dos peticiones al LLM. En primer lugar, se le pide generar a partir del texto de entrada una lista de afirmaciones factuales como frases autocontenidas y relevantes (ver Figura 9). Tras esto, se le pide reformular dichas frases para eliminar correferencias (ver Figura 10). Tras este proceso, tendremos una lista de afirmaciones extraídas del texto analizado (ver Figura 11).



#### Claim generation

The claims generation have two phases: first, we will ask for a list of claims stated on the text; second, we will try to reformulate those claims to solve coreferences.

[] claim\_extraction\_prompt = """Your task is to generate a list with the main factual claims stated in a text. A factual claim makes

```
Arrange your output using the format "-- claim\\n-- claim\\n-- claim"."""
user prompt = f"""
    TEXT: {text}
    OUTPUT:
    ....
messages = [
    sages = [
{"role": "system", "content": claim_extraction_prompt},
{"role": "user", "content": user_prompt},
prompt = tokenizer.apply_chat_template(
    messages, tokenize=False, add generation_prompt=True
outputs = model.generate([prompt], sampling_params)
generated_text = outputs[0].outputs[0].text
```

🕣 Processed prompts: 100% 🚺 🚺 1/1 [00:03<00:00, 3.06s/it, est. speed input: 82.42 toks/s, output: 32.38 toks/s]

Figura 9. Extracción de afirmaciones a partir del texto a analizar.



🖅 Processed prompts: 100%

Figura 10. Reformulación de afirmaciones para eliminar correferencias.

{'claim': 'Variation in body fat distribution is seen in groups of individuals with the same racial or ethnic background and even within families.'}
{'claim': 'Genetic changes in human populations occur too slowly to be responsible for the obesity epidemic.'}
{'claim': 'Not all individuals living in environments that promote physical inactivity and increased consumption of high-calorie foods will become obese.'}
{'claim': 'Genes play a role in the development of obesity.'}
{'claim': 'The variation in how populations respond to the same environment suggests that genes play a role in obesity.'}

Figura 11. Ejemplo de listado de afirmaciones generados por el LLM.

El siguiente paso es la búsqueda de documentos relacionados dentro de la colección de abstracts de verificación. Para ello hacemos uso de la función de ranking BM25, quedándonos con los 3 abstracts más relacionados para cada afirmación. Cada afirmación del listado anterior será enlazada a un nuevo diccionario con el DOI, el título y el abstract de los tres documentos relacionados (ver Figura 12).



Abstract retrieval

Here we will use the bm25 ranking function to obtain the most related abstract from the verified dataset to each claim from the list.

```
/ [13] for i, claim_dict in enumerate(claim_candidates):
    query = claim_dict["claim"]
    tokenized_query = query.split(" ")
    search_results = bm25.get_top_n(tokenized_query, verified_corpus, n=3)
    claim_dict["search_results"] = search_results
```

- - {'claim': 'Variation in body fat distribution is seen in groups of individuals with the same racial or ethnic backgrou
    {'claim': 'Genetic changes in human populations occur too slowly to be responsible for the obesity epidemic.', 'search
    {'claim': 'Not all individuals living in environments that promote physical inactivity and increased consumption of hi
    {'claim': 'Genes play a role in the development of obesity.', 'search\_results': [{'id': 'https://doi.org/10.1046/j.136
    {'claim': 'The variation in how populations respond to the same environment suggests that genes play a role in obesity

Figura 12. Búsqueda de abstracts relacionados con cada una de las afirmaciones generadas.

El último paso del proceso será la generación del reporte de factualidad para cada par de afirmación y abstract. Dicho reporte será generado mediante una nueva petición al LLM, que devolverá un documento json con una etiqueta, la base lógica detrás de su decisión y las frases del abstract de verificación que mejor se relacionan con la afirmación emparejada (ver Figura 13). La etiqueta describe la relación entre la afirmación y el abstract, y puede tener tres valores: *SUPPORT* (la afirmación está soportada por el abstract), *REFUTE* (la afirmación está refutada por el abstract) o *NEI* (*Not Enough Information*; no existe relación suficiente entre afirmación y abstract como para asignar una etiqueta al respecto de su factualidad). El *rationale* expone la explicación del modelo detrás de la asignación de la etiqueta. Por último, en *evidence* podemos encontrar las frases del abstract más relacionadas con la afirmación analizada. Además, se incluye un valor de porcentaje de confianza basado en los valores probabilísticos devueltos para la selección de la etiqueta por parte del LLM.

```
"input_text": "Do Genes Have a Role in Obesity?\n\nIn recent decades, obesity has reached epidemic proportions in popul
"claims": [
    {
         "claim": "Variation in body fat distribution is seen in groups of individuals with the same racial or ethnic ba
         "claim_analysis": [
             {
                  "id": "https://doi.org/10.1007/s11892-020-01302-2",
"title": "Physiological and Lifestyle Traits of Metabolic Dysfunction in the Absence of Obesity.",
"abstract": "The prevalence of MUNW varies considerably around the world and among ethnicities, partly
                   "report": {
                       "response": "NEI",
"rationale": "The provided evidence is related to the relationship between obesity and brain activa
                       "evidence": []
                  }
             },
                  "id": "https://doi.org/10.1098/rstb.2006.1853",
                  "title": "Psychophysics of sweet and fat perception in obesity: problems, solutions and new perspective
                   "abstract": "Psychophysical comparisons seem to show that obese individuals experience normal sweet and
                   "report": {
                       "response": "SUPPORT",
"rationale": "Although the evidence suggests that genetic changes are not the primary driver of the
                       "evidence": [
                             The global food system drivers interact with local environmental factors to create a wide vari
                            "Within populations, the interactions between environmental and individual factors, including g
                            "Unlike other major causes of preventable death and disability, such as tobacco use, injuries,
                       ],
"confidence": 54.89
```

Figura 13. Ejemplo de análisis de afirmaciones científicas.



# 2.2 Diagnóstico médico

El objetivo de este demostrador es guiar a un LLM para generar razonamientos clínicos y diagnósticos a partir de datos médicos, ya sea en forma de narrativas clínicas o casos médicos estructurados, incluyendo síntomas y posibles pruebas clínicas. En la instrucción se incluirán ejemplos de narrativas, datos médicos y diagnósticos de referencia si están disponibles. Este demostrador se implementará en la siguiente versión del entregable.

Caso de uso	Diagnóstico médico	
Entradas		
Datos en ELG	European Clinical Case Corpus [Alojado in ELG] [NO anotado con diagnostico] https://live.european-language-grid.eu/catalogue/corpus/7618/overview/	
Datos en Hugging Face	ClinDiagnosES [anotado con diagnóstico] https://huggingface.co/datasets/LenguajeNaturalAI/ClinDiagnosES MedQA https://huggingface.co/datasets/katielink/nejm-medqa-diagnostic- reasoning-dataset [anotado con diagnostico] Casimedicos [anotado con razonamiento clínico, diagnóstico y explicaciones de razonamiento médico] https://huggingface.co/datasets/HiTZ/casimedicos-exp DxBench https://huggingface.co/datasets/FreedomIntelligence/DxBench	
Datos en el EDL	NLICE <u>https://github.com/guozhuoran918/NLICE</u> [anotado con diagnóstico] DDXPlus <u>https://github.com/mila-iqia/ddxplus</u> [anotado con diagnostico] Medbullets <u>https://github.com/HanjieChen/ChallengeClinicalQA/tree/main/medbullets</u> [anotado con razonamiento clínico]	
LLM en hugging face	mejor modelo en <u>https://huggingface.co/spaces/la-leaderboard/la-leaderboard</u> que tiene alrededor de 8B de parámetros	
Requisitos	<ul> <li>Conector de EDL integrado con ELG</li> <li>Acceso a Hugging Face</li> <li>Conector "consumidor" de EDL</li> <li>LLM disponible en Hugging Face</li> </ul>	
Proceso		
<ul> <li>Descargar el conjunto de datos E3C de ELG usando el conector de ELG en EDL</li> <li>Descargar conjuntos de datos ClinDiagnosES, MedQA, Casimedicos y DxBench de HF</li> <li>Descargar conjunto de datos NLICE, DDXPlus, Medbullets del EDL con conecto "consumidor"</li> <li>Para cada caso médico en los conjunto de datos y en un lenguaje de interés:</li> <li>Guiar al LLM para que genere un razonamiento clínico o un diagnóstico a parti</li> </ul>		

- de datos médicos estructurados o narrativas
- Mejorar el prompt inyectando estrategias de razonamiento y resolución de problemas



# Mejorar el prompt con ejemplos tomados de conjunto de datos que incluyen pasos de razonamiento o explicaciones como ClinDiagnosES y Casimedicos

## 2.3 Evaluación de factualidad en dominio de seguros

El objetivo de este demostrador es evaluar la factualidad de los LLM en el dominio de seguros. La factualidad del modelo estará basada en un dataset de referencia, en concreto, se usará un volcado de Wikipedia en inglés. Para calcular la factualidad, el modelo en evaluación responde a un conjunto de preguntas en el dominio de seguros. Las respuestas se procesan, extrayendo los hechos atómicos, y evaluando el soporte o no en base a Wikipedia. Este demostrador se implementará en la siguiente versión del entregable.

Caso de uso	Evaluación de factualidad en dominio de seguros			
Entradas				
Datos en ELG	Insurance QA – test set			
LLM en hugging face	Mejor modelo en <u>https://huggingface.co/spaces/la-leaderboard/la-leaderboard</u> con menos de 4 billones de parámetros.			
Requisitos	<ul> <li>Conector de EDL integrado con ELG</li> <li>Acceso a Hugging Face</li> <li>LLM disponible en Hugging Face</li> </ul>			
Proceso				
<ul> <li>Descargar el conjunto de datos Insurance QA de ELG desde el EDL.</li> <li>Para cada LLM:         <ul> <li>Responder las preguntas con el modelo</li> <li>Extraer los hechos atómicos con <u>Llama-3.2-3B</u></li> <li>Evaluación de factualidad de los hechos, como <i>fact-checker</i> se usará tambié <u>Llama-3.2-3B</u></li> </ul> </li> </ul>				

Para el demostrador, tanto para la parte de extracción de hechos atómicos, como para el *fact-checking*, se usará <u>Llama-3.2-3B</u>. Se usa este modelo porque el notebook estará pensado para ejecutarse en Google Colab, por lo que se deben usar LLMs que puedan ejecutarse con la memoria de GPU disponible. Para mejores resultados, se deberían usar modelos más potentes a través de APIs de pago, como los de OpenAI.

# 3 Reportes

La generación de reportes sobre el estado del catálogo del EDL es en línea de tal manera que provea una visión precisa y actualizada de los recursos funcionales y no funcionales registrados en el catálogo. Para esto se ha incluido un enlace en el icono de InesData en el menú de la izquierda de la interfaz de usuario del conector (ver Figura 14) que despliega una página con el número de activos con contrato asociado disponible en el catálogo del EDL diferencias por tipología de activos e idioma.



nesData			•			
CREATE	Inesdata Catalog					
E. Comm	Sear	ch and negotiate th	e available linguistic assets	in the INESDATA la	nguage data space.	
	Corpus	Corpus 3			Lexical resources O	
J. Models	3					
E Lexical resources						
	Basque	1	Albanian	0	Albanian	0
	Catalan	1	Asturian	0	Asturian	0
	Galician	1	Basque	0	Basque	0
	Spanish	1	Belarusian	0	Belarusian	0
	Albanian	0	Bosnian	0	Bosnian	0
	Asturian	0	Breton	0	Breton	0
	Belarusian	0	Bulgarian	0	Bulgarian	0
	Others	0	Others	0	Others	0
Account						
My console						

Figura 14. Versión Inicial del reporte de estadísticas del catálogo del EDL. Este reporte se despliega al acceder a la interfaz del conector y cuando se da click sobre el logo de INESData en el menú de la izquierda.