

LASAI 2026 · Lenguaje, Educación y Explicabilidad

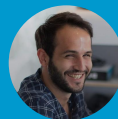
Extracción multilingüe de habilidades

para un mercado laboral impulsado por IA*

Multilingual Skill Extraction for an AI-Driven Job Market

Laura Vásquez-Rodríguez, Bertrand Audrin (EHL, HES-SO), Samuel Michel (Idiap), Samuele Galli (Arca24), Julneth Rogenhofer (EHL, HES-SO), Jacopo Negro Cusa (Arca24), Lonneke van der Plas (USI)

*Trabajo realizado en el Instituto de Investigación Idiap en Suiza.



De un CV a una decisión

¿Qué pasa con un CV cuando lo lee un sistema?





¿A cuál persona contratarías?



CONTACTO
 Lima, Perú
 +51 9XX XXX XXX
 valeria.mora@example.com

HABILIDADES
 Python
 SQL
 Power BI
 Excel

FORTALEZAS
 Trabajo en equipo
 Comunicación
 Resolución de problemas

IDIOMAS
 Español
 Inglés
 Portugués

VALERIA MORA
ANALISTA DE DATOS
PERFIL
 Analista con 4 años en el sector hotelero. Transformo datos en decisiones con análisis de datos y disfruto el trabajo en equipo.

EXPERIENCIA
Analista de Datos
 Grupo Hotelero Pacífico · 2022 – hoy
 • Creé tableros en Power BI para 12 hoteles.
 • Automatiqué reportes con Python y SQL.
 • Coordiné un equipo de 3 practicantes.
Asistente de Recepción
 Hotel Los Olivos · 2019 – 2022
 • Atención al cliente en español e inglés.
 • Reservas en Opera PMS.

EDUCACIÓN
Lic. en Administración Hotelera
 Instituto Superior de Turismo · 2019
Curso: análisis de datos con Python
 Plataforma en línea · 2021

VALERIA MORA

Analista de Datos | Lima, Perú | valeria.mora@example.com | +51 9XX XXX XXX

PERFIL

Analista de datos con 4 años de experiencia en el sector hotelero. Transformo datos en decisiones con análisis de datos, reportes y trabajo en equipo.

EXPERIENCIA

Analista de Datos, Grupo Hotelero Pacífico, 2022 – hoy

- Creé tableros en Power BI para 12 hoteles.
- Automatiqué reportes con Python y SQL.
- Coordiné un equipo de 3 practicantes.

Asistente de Recepción, Hotel Los Olivos, 2019 – 2022

- Atención al cliente en español e inglés.
- Gestión de reservas en Opera PMS.

EDUCACIÓN

Licenciatura en Administración Hotelera, Instituto Superior de Turismo, 2019

Curso de Análisis de Datos con Python, 2021

HABILIDADES

Técnicas: Python, SQL, Power BI, Excel, Opera PMS

Blandas: trabajo en equipo, comunicación, resolución de problemas

IDIOMAS

Español (nativo), inglés (C1), portugués (A2)

¿Qué habilidades ven en este CV?¹



VM

CONTACTO
Lima, Perú
+51 9XX XXX XXX
valeria.mora@example.com

HABILIDADES

Python

SQL

Power BI

Excel

FORTALEZAS

Trabajo en equipo

Comunicación

Resolución de problemas

IDIOMAS

Español ●●●●●

Inglés ●●●●○

Portugués ●●○○○

VALERIA MORA

ANALISTA DE DATOS

PERFIL

Analista con 4 años en el sector hotelero. Transformo datos en decisiones con análisis de datos y disfruto el trabajo en equipo.

EXPERIENCIA

Analista de Datos
Grupo Hotelero Pacifico · 2022 – hoy

- Creé tableros en Power BI para 12 hoteles.
- Automatice reportes con Python y SQL.
- Coordiné un equipo de 3 practicantes.

Asistente de Recepción
Hotel Los Olivos · 2019 – 2022

- Atención al cliente en español e inglés.
- Reservas en Opera PMS.

EDUCACIÓN

Lic. en Administración Hotelera
Instituto Superior de Turismo · 2019

Curso: análisis de datos con Python

Plataforma en línea · 2021

Habilidad técnica

Python, SQL, Power BI, Excel, Opera PMS, análisis de datos

Habilidad blanda

trabajo en equipo, comunicación, atención al cliente, resolución de problemas

- Algunas habilidades están **escritas tal cual**: «Python», «SQL».
- Otras hay que **inferirlas**: «Coordiné un equipo» → liderazgo.

CV ficticio, creado para esta charla.

¹ Vázquez-Rodríguez et al. (2025). A Human Perspective to AI-based Candidate Screening. HICSS-58.

Dos miradas a un mismo CV²

PERSPECTIVA HUMANA

Recursos humanos (RR. HH.)



- Un reclutador revisa un currículum (CV) en **unos 10 segundos.**¹
- Busca el **ajuste con el puesto**, no habilidades sueltas.
- Las habilidades blandas se expresan de muchas formas: requieren **interpretación**.

PERSPECTIVA COMPUTACIONAL

Sistemas automáticos



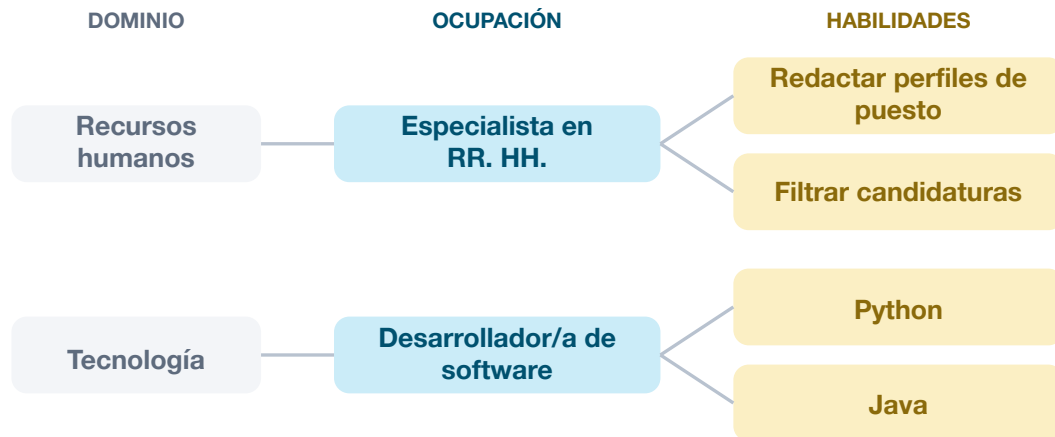
- Un **ATS** (sistema de seguimiento de candidatos) extrae habilidades del CV y de la oferta, y las compara.
- Taxonomías: **muy precisas**, pero encuentran poco.
- Modelos supervisados: **generalizan mejor**, pero dependen de las anotaciones.

¹ Hangartner et al. (2021). Monitoring Hiring Discrimination through Online Recruitment Platforms. Nature, 589.

² Vásquez-Rodríguez et al. (2025). A Human Perspective to AI-based Candidate Screening. HICSS-58.

¿Qué es una taxonomía?¹

Un mapa de conceptos: qué ocupaciones existen y qué habilidades necesitan.



Ejemplo ilustrativo, inspirado en la estructura dominio → ocupación → habilidades de la taxonomía interna del proyecto.

¹Comisión Europea. ESCO: clasificación europea de capacidades, competencias y ocupaciones. esco.ec.europa.eu.

La contratación es multilingüe¹

La mayoría de los métodos se limitan a ofertas de empleo en inglés. No reflejan la contratación real, que es multilingüe.

01 · EL IDIOMA

Misma idea, otras palabras.

DE Teamfähigkeit

ES trabajo en equipo

DE Belastbarkeit

ES capacidad de trabajar bajo presión

Una palabra en alemán, varias en español: ¿dónde empieza la habilidad?

02 · EL PAÍS Y LA CULTURA

Detrás del idioma hay una cultura de reclutamiento.

Foto: habitual en Suiza; se evita en Estados Unidos.

Trato: formal o informal. «Sie» o «du»; «usted» o «tú».

Datos personales: en Suiza, nacionalidad y permiso de trabajo.

03 · NUESTRO RETO

Un sistema que identifique habilidades en **CVs y ofertas de empleo**, en **seis idiomas**, teniendo en cuenta estas variaciones.

EN

FR

DE

IT

ES

PT

¹ Senger et al. (2024). Deep Learning-based Computational Job Market Analysis: A Survey on Skill Extraction and Classification from Job Postings. NLP4HR 2024.

Proyecto SEM24^{1,2}

Investigadores en RR. HH. e IA, junto con la industria, para emparejar mejor candidatos y ofertas en varios idiomas.

Innovation project supported by



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation

Innosuisse – Swiss Innovation Agency

- Evaluamos varios métodos de extracción **en seis idiomas**.
Inglés, francés, alemán, italiano, español y portugués.
- Comparamos **CVs y ofertas de empleo** en ~45 dominios profesionales.
- El objetivo: un marco más completo y **relevante para la industria**.
- **Colaboración:** proyecto **Innosuisse** (agencia suiza de innovación) de 2 años con Arca24, EHL, Idiap Research Institute y USI.

¹ HES-SO. SEM24, proyecto Innosuisse. hes-so.ch/en/recherche-innovation/research-projects/detail-projet-recherche/sem24.

² Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

Los datos

¿Con qué datos se entrena y se evalúa un sistema?



01

El CV

02

Los datos

03

Los sistemas

04

Personas vs. sistemas

05

Las decisiones

06

Tu postulación

¿Pero dónde están los datos?¹

El primer reto del proyecto

01 · LA PROPUESTA

Miles de ofertas de empleo y CVs, **ya anotados** con sus habilidades.



02 · LO QUE ENCONTRAMOS

Las únicas etiquetas venían del **sistema basado en reglas**. Evaluar con ellas es medir al sistema **contra sí mismo**.



03 · LO QUE HICIMOS

Replantear el proyecto: diseñar un **marco de anotación** con especialistas en RR. HH. (marco KSAO)¹

¹ Campion et al. (2011). Doing Competencies Well: Best Practices in Competency Modeling. Personnel Psychology, 64(1).



Anotación guiada por RR. HH.

1,200

DOCUMENTOS EN TOTAL

100 ofertas de trabajo + 100 CVs por idioma

~45

DOMINIOS PROFESIONALES

10,379

HABILIDADES

Taxonomía INTERNA (recopilada previamente)

SEIS IDIOMAS

EN

FR

DE

IT

ES

PT

CON HUMANOS EN EL CICLO

Dos investigadores en RR. HH. escribieron las guías con el marco **KSAO**¹: conocimientos, habilidades, aptitudes y otros atributos.

Cuatro anotadores con experiencia previa anotaron en **Doccano**. Tras una primera ronda, resolvimos juntos los conflictos hasta lograr un **acuerdo total**.

Anotadores: EN-ES · FR-DE · IT-PT

Nativos: IT, FR, ES

Nivel C2: EN, DE, PT

¹Campion et al. (2011). Doing Competencies Well: Best Practices in Competency Modeling. Personnel Psychology, 64(1).

Conjuntos de datos públicos^{1,2}

~10,000

ORACIONES

Green Dataset (ofertas, Reino Unido)

5+

DOMINIOS: TI, FINANZAS, SALUD, VENTAS

13,939

HABILIDADES

Taxonomía europea ESCO, 28 idiomas

UN IDIOMA

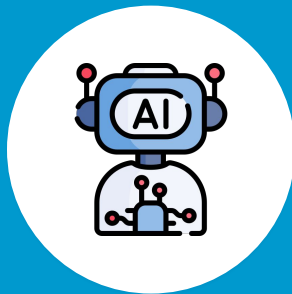
EN

CON HUMANOS EN EL CICLO

Oraciones anotadas en Amazon Mechanical Turk (AMT) por 39 trabajadores muy calificados + posprocesamiento automático.

Los sistemas

¿Cómo encuentra un sistema las habilidades en un texto?



01

El CV

02

Los datos

03

Los sistemas

04

Personas vs. sistemas

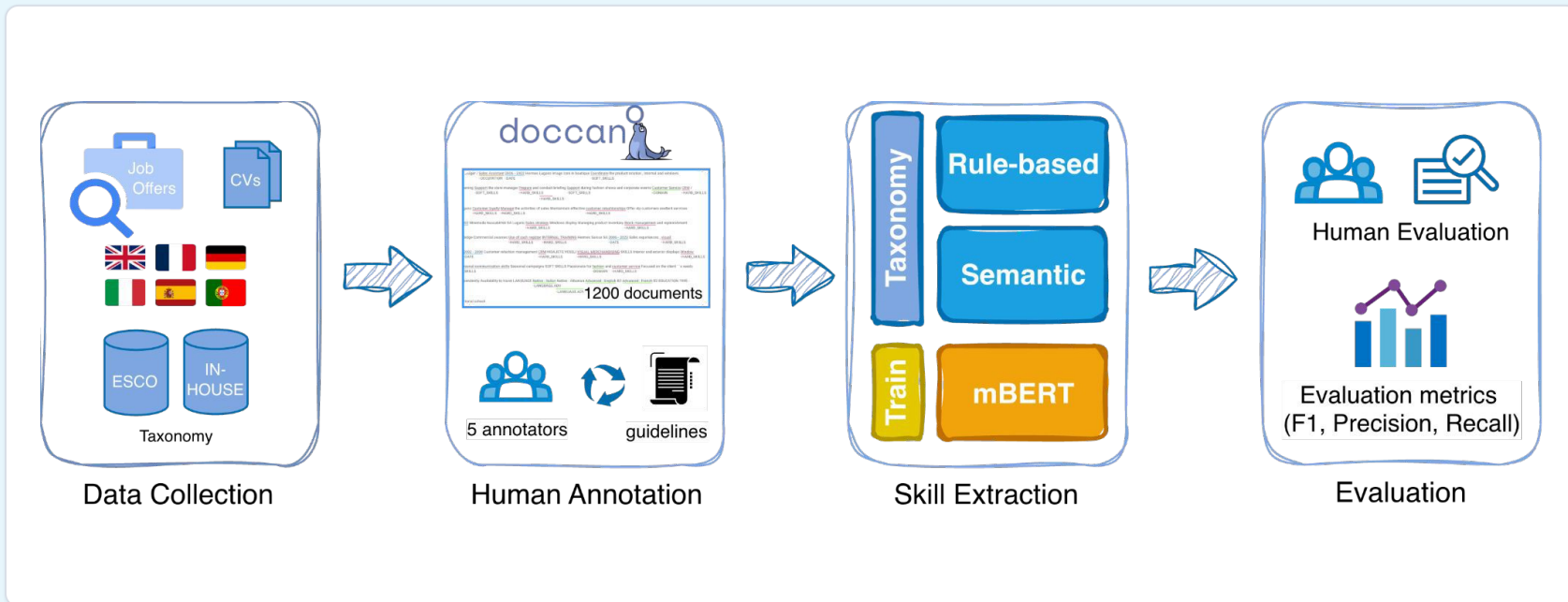
05

Las decisiones

06

Tu postulación

De los documentos a la evaluación¹



¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

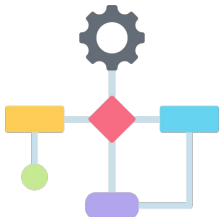
Tres métodos de extracción¹

Basado en reglas

Coincidencia exacta

Coincidencia estricta con las taxonomías interna y ESCO.

Transparente y fácil de actualizar, pero maneja mal la variación.



Semántico

Similitud semántica

Encuentra términos cercanos.

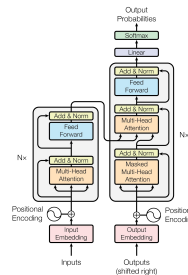
Generaliza mejor y sigue siendo trazable, pero no aprenden conceptos nuevos.



mBERT supervisado

Modelo supervisado

BERT multilingüe (mBERT) entrenado con anotaciones manuales. Capta el contexto, pero es una caja negra.



¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

Método 1: basado en reglas

Busca en el texto los términos exactos de una taxonomía.

1 · TEXTO DE ENTRADA

Especialista en datos: **analiza grandes conjuntos de datos**,
construye modelos predictivos y **comunica resultados**.

2 · NORMALIZAR LAS PALABRAS

analiza → analizar

construye → construir

comunica → comunicar

3 · BUSCAR EN LA TAXONOMÍA

Si el término normalizado está en la taxonomía, es una habilidad.

TAXONOMÍA

Inteligencia artificial

Científico/a de datos

✓ analizar grandes conjuntos de datos

✓ construir modelos predictivos

✓ comunicar resultados

+ Transparente: cada habilidad viene de la taxonomía. **– Rígido:** si el término no está en la lista, no se encuentra.

Método 2: búsqueda semántica^{1,2}

Compara significados, no palabras exactas.

«lideraré un equipo de 5 personas»

Modelo multilingüe (SBERT)¹: texto → vector de números

TÉRMINOS MÁS CERCANOS EN LA TAXONOMÍA (SIMILITUD DEL COSENO)



Términos con significado parecido quedan cerca en el espacio de embeddings.

+ Menos estricto que las reglas y sigue siendo trazable. **– No aprende** conceptos nuevos.

Modelo: distiluse-base-multilingual-cased-v1. Valores de similitud ilustrativos.

¹Yang et al. (2020). Multilingual Universal Sentence Encoder for Semantic Retrieval. ACL 2020.

²Reimers y Gurevych (2020). Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. EMNLP 2020.

Extracción multilingüe de habilidades · LASAI 2026

Método 3: mBERT supervisado¹

La extracción como reconocimiento de entidades nombradas (NER).

CADA PALABRA RECIBE UNA ETIQUETA (ESQUEMA IOB)



Resultado: **análisis de datos** · **Python** · **Pandas**, cada una con su fragmento exacto en el texto.

EL MODELO

BERT multilingüe (mBERT), ajustado con las anotaciones de RR. HH. Multilingüe y de bajo costo.

POR QUÉ NER

Devuelve **fragmentos exactos**. Así medimos si el límite es correcto: evaluación **exacta** y **parcial**.

UN HALLAZGO

Modelos **sin conocimiento previo** de extracción de habilidades funcionaron mejor.²

¹Devlin et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019.

²Vásquez-Rodríguez et al. (2024). Hardware-effective Approaches for Skill Extraction in Job Offers and Resumes. RecSys in HR, 2024

¿Y por qué no un LLM conversacional?^{1,2}



Es la pregunta más común.

01 · PREDICTIBILIDAD

La misma entrada puede dar salidas distintas. Ser predecible era un **requisito del proyecto**.

02 · EXPLICABILIDAD

Hay que poder explicar de dónde sale cada habilidad. Eso genera **confianza** en los usuarios.

03 · COSTO

Necesitan GPU (procesador gráfico). No todas las empresas pueden pagarla: buscamos **bajo costo**.

04 · CONTROL

Menos control sobre **sesgos** y **alucinaciones** (respuestas inventadas) que con reglas o modelos supervisados.

¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

² Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Evaluación exacta y parcial

Humano

Data analysis using Python and Pandas



Exacta

Data analysis using Python and Pandas



Parcial

Data analysis using Python and Pandas



En la evaluación parcial damos crédito a habilidades detectadas en parte, como «Data» para «Data analysis».

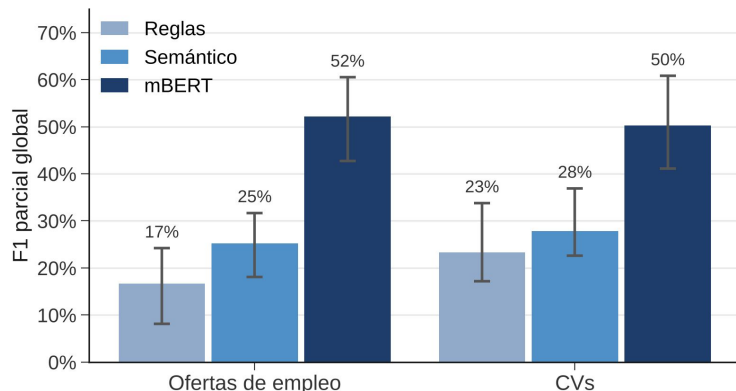
■ Humano — referencia

■ Sistema — modelo

mBERT gana; los límites son difíciles¹

A mBERT gana en rendimiento

F1 parcial medio en 6 idiomas



A · RENDIMIENTO

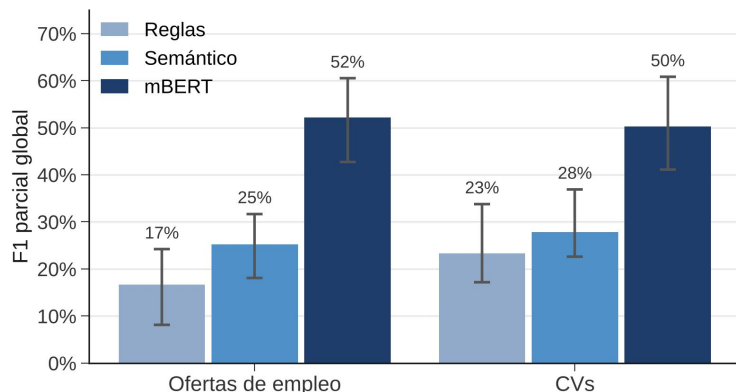
mBERT ajustado supera por amplio margen a reglas y semántico (F1 parcial medio: **52%** en ofertas y **50%** en CVs).

¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

mBERT gana; los límites son difíciles^{1,2}

A mBERT gana en rendimiento

F1 parcial medio en 6 idiomas

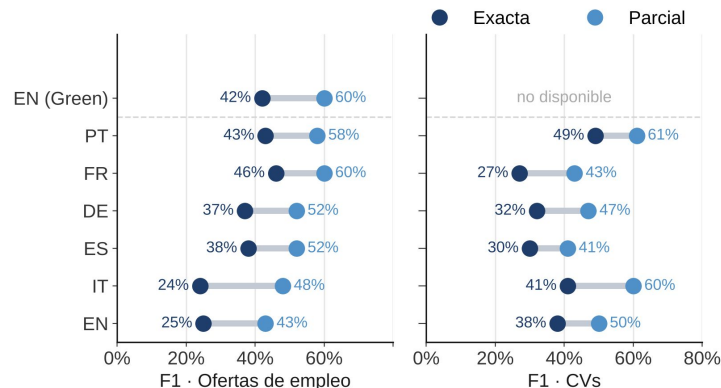


A · RENDIMIENTO

mBERT ajustado supera por amplio margen a reglas y semántico (F1 parcial medio: **52%** en ofertas y **50%** en CVs).

B Exacta vs. parcial

mBERT por idioma — la brecha muestra la incertidumbre en los límites



B · EXACTA VS. PARCIAL

Lo parcial supera a lo exacto. El modelo encuentra la habilidad, pero **no sabe dónde empieza y termina**.

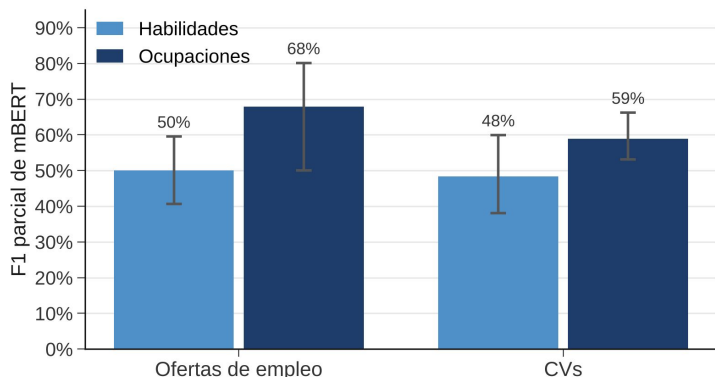
¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

² Green et al. (2022). Development of a Benchmark Corpus to Support Entity Recognition in Job Descriptions. LREC 2022.

Ocupaciones fáciles, habilidades difíciles¹

C Ocupaciones fáciles, habilidades difíciles

F1 parcial medio de mBERT en 6 idiomas



C · OCUPACIONES VS. HABILIDADES

Los puestos son más fáciles: F1 parcial hasta **68%**.

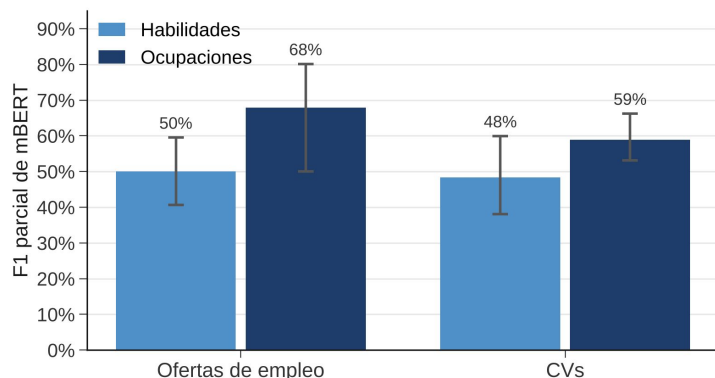
Lo difícil es el **fragmento exacto**.

¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

Ocupaciones fáciles, habilidades difíciles^{1,2}

C Ocupaciones fáciles, habilidades difíciles

F1 parcial medio de mBERT en 6 idiomas

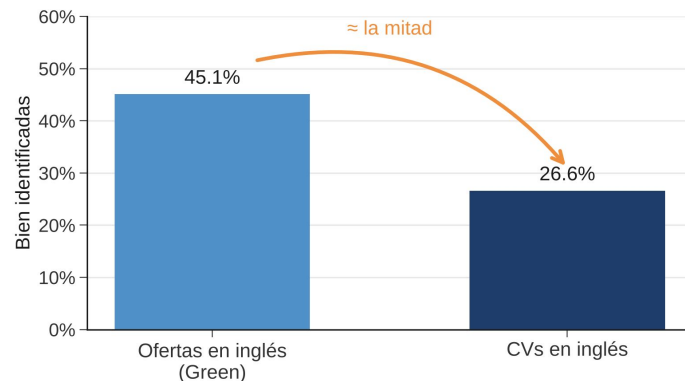


C · OCUPACIONES VS. HABILIDADES

Los puestos son más fáciles: F1 parcial hasta **68%**.
Lo difícil es el **fragmento exacto**.

D Los CVs son más difíciles que las ofertas

Evaluación humana, inglés (habilidades bien identificadas)



D · CVs VS. OFERTAS

Los aciertos **casi se reducen a la mitad** de ofertas a CVs: **45.1% → 26.6%** en inglés.

¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

² Green et al. (2022). Development of a Benchmark Corpus to Support Entity Recognition in Job Descriptions. LREC 2022.

Análisis manual

Ofertas (EN) Semántico (taxonomía interna)

Hum. An **analytical** approach to **problem solving** excellent **team working skills**

Sist. An **analytical approach** to **problem solving** excellent team working skills

CVs (EN) Basado en reglas (taxonomía interna)

Hum. ...different languages **PHP** **HTML** **CSS** **Javascript** **jQuery** **SQL** **Visual Basic** **Linux** **Bash** **FileMaker Script** **C**

Sist. ...different **languages** **PHP** **HTML** **CSS** **Javascript** **jQuery** **SQL** **Visual Basic** **Linux** **Bash** **FileMaker Script** **C**

Supervisado vs. reglas

CVs (EN)

El mismo fragmento, tres lecturas

Humano

...languages PHP HTML CSS Javascript jQuery SQL Visual Basic Linux Bash FileMaker Script C

Reglas

...languages PHP HTML CSS Javascript jQuery SQL Visual Basic Linux Bash FileMaker Script C

mBERT

...languages PHP HTML CSS Javascript jQuery SQL Visual Basic Linux Bash FileMaker Script C

REGLAS

Une términos vecinos («languages PHP», «HTML CSS») y **no encuentra** Linux, Bash ni FileMaker Script: no están en la taxonomía.

MBERT SUPERVISADO

Encuentra **todas** las habilidades. Pero une dos en una («Linux Bash»): es un **error de límite**, no de significado.

 Humano — referencia

 Sistema — modelo

Conclusiones (parte I)¹

01 · RENDIMIENTO

mBERT supera por amplio margen a reglas y semántico, a cambio de un **modelo de «caja negra»** más difícil de controlar.

Alternativa: volver a mapear a la taxonomía y curar los conceptos nuevos.

02 · LAS HABILIDADES SON DIFÍCILES

Los puestos de trabajo son fáciles de encontrar. Saber exactamente dónde **empieza** y **termina** una **habilidad** sigue siendo retador para la IA.

03 · CVs ≠ OFERTAS

Los CVs son personales y poco estandarizados en muchas dimensiones: **personas, dominios, idioma.**

Personas vs. sistemas^{1,2}

De «¿qué sistema acierta más?» a «¿qué ve cada sistema en el mismo CV?»

HASTA AHORA

Tres métodos frente a las anotaciones de especialistas en RR. HH.

1,200 documentos, seis idiomas.

Pregunta: ¿cuál extrae mejor?



AHORA

Un segundo estudio, con **46 CVs** reales.

Humanos y tres sistemas, **incluido un LLM conversacional**.

Pregunta: ¿en qué se diferencian?

01

El CV

02

Los datos

03

Los sistemas

04

Personas vs. sistemas

05

Las decisiones

06

Tu postulación

¹ Vázquez-Rodríguez et al. (2025). A Human Perspective to AI-based Candidate Screening. HICSS-58.

² Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Mismo CV, ¿mismas habilidades?^{1,2}

46

CVs

Estudiantes de posgrado en hotelería (EHL)

116

ANOTACIONES · 41 ANOTADORES, ~2.5 POR CV

~465

PALABRAS POR CV, EN PROMEDIO

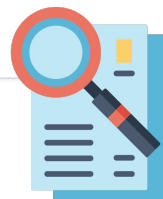
Analizamos solo habilidades técnicas

UN IDIOMA

EN

LA PREGUNTA

Para un mismo CV, ¿qué habilidades extraen los **humanos** y los **sistemas**? Medimos **longitud** y **significado**.



¹ Vázquez-Rodríguez et al. (2025). A Human Perspective to AI-based Candidate Screening. HICSS-58.

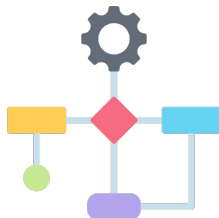
² Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Tres sistemas^{1,2}

Basado en reglas

SkillNER³ + taxonomía

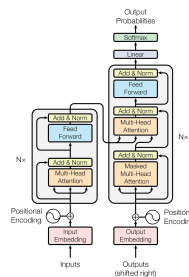
Busca términos de una taxonomía interna de la industria. Salida corta y trazable.



Supervisado

ESCOXLM-R

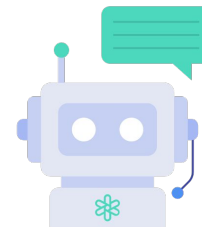
Modelo tipo BERT pre entrenado con ESCO y ofertas de empleo. Sin entrenamiento adicional.



LLM conversacional

Mixtral 8x7B

Modelo de lenguaje grande (LLM) de pesos abiertos, en local por privacidad. Un prompt, sin ejemplos.



¹Zhang et al. (2023). ESCOXLM-R: Multilingual Taxonomy-driven Pre-training for the Job Market Domain. ACL 2023.

²Jiang et al. (2024). Mixtral of Experts. arXiv:2401.04088.

³<https://github.com/AnasAito/SkillNER>

Así le pedimos las habilidades al LLM¹

Prompt

Identify the hard skills present in the following text: "Work on the identity and the architecture of the brand". List each skill, one per line, with no elaboration or clarification in parentheses. If there are no hard skills present in the text, return "None" without further explanation.

Output

"Branding, Identity Design, Architecture (Design)"

LO QUE PEDIMOS

Una habilidad **por línea**, sin aclaraciones entre paréntesis. Si no hay, «None».

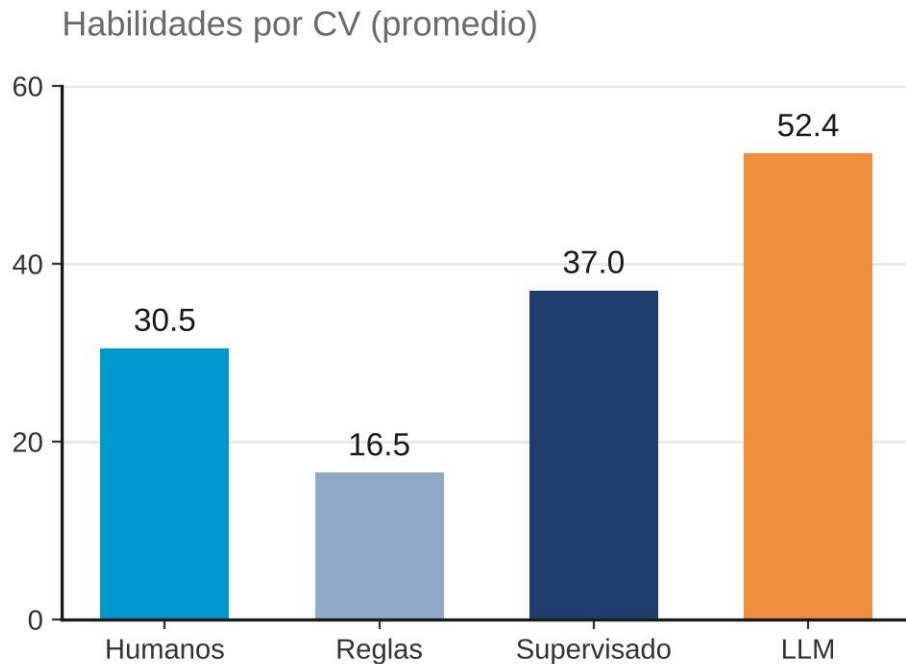
LO QUE OBTUVIMOS

Una lista **con comas y paréntesis**, y un término que **no está en el texto** («Branding»).

Figura 2 del artículo; prompt original en inglés.

¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

¿Cuántas habilidades encuentra cada uno?¹



PROMEDIO POR CV

El **LLM** extrae **1.7 veces** más habilidades que los humanos.

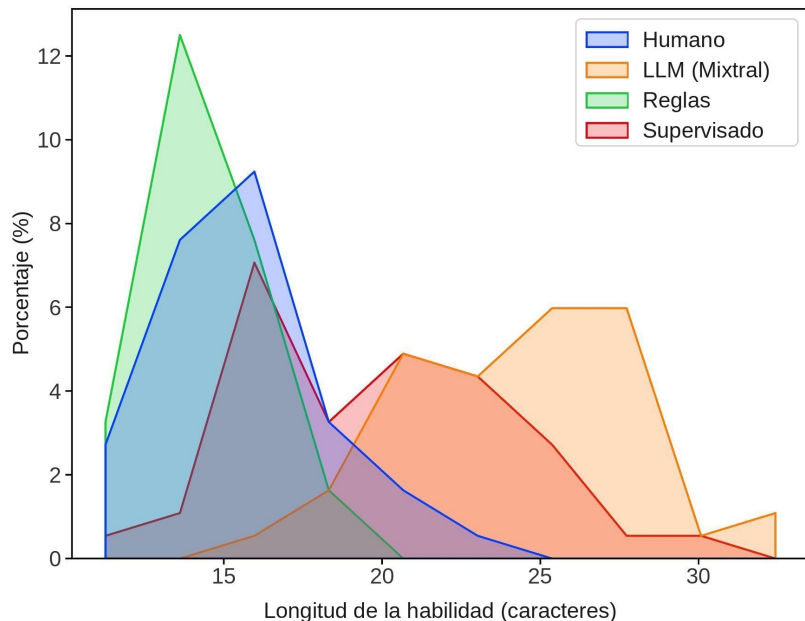
Las **reglas** encuentran casi **la mitad**.

El modelo **supervisado** es el más cercano en cantidad.

Tabla 2 del artículo.

¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Longitud: del término a la frase¹



LONGITUD EN CARACTERES

Humanos: la mayoría de las habilidades tienen **~15 caracteres**.

Reglas: más cortas, cerca de los humanos. Supervisado: **más largas**.

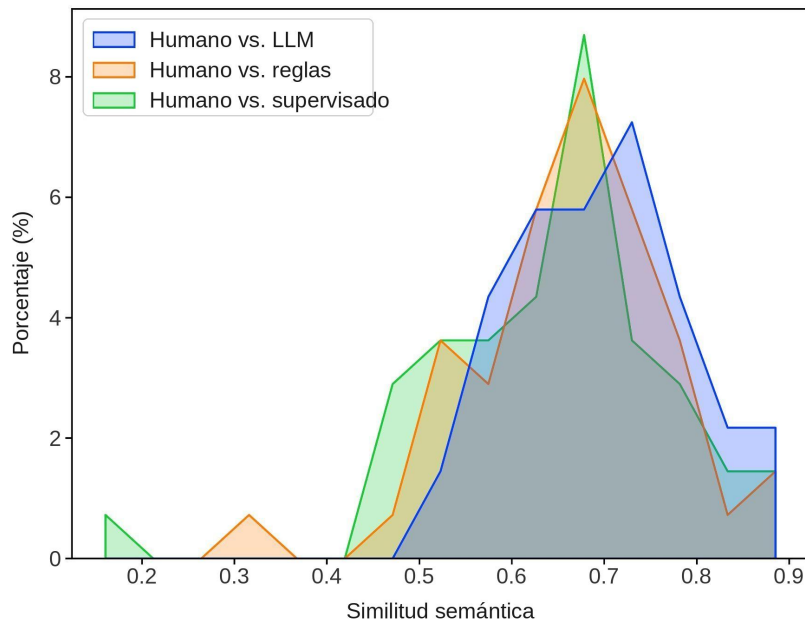
LLM: las más largas, entre **15 y 35**.

El LLM extrae **frases**; las reglas, **términos**.

Figura 3a del artículo.

¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Significado: el LLM se acerca más^{1,2}



SIMILITUD SEMÁNTICA CON HUMANOS

Medimos la similitud con **SBERT** (Sentence-BERT).

La mayoría de los pares está **entre 0.5 y 0.8**.

Ningún par llega a 0.9: no hay coincidencias exactas.

El **LLM** es el más cercano al significado humano.

Figura 3c del artículo.

¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

² Reimers y Gurevych (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. EMNLP 2019.

Diez anotaciones al azar de cada fuente¹

Fuente	Anotaciones (10 al azar por fuente)
Humano	negotiation, market research, project planning, language skill-English, Microsoft 365, training skills, tlb management, marketing and strategy, culinary certificate, interview experience
Reglas	hotel led, contractor, management graduated, mathematics, sports, prepared materials, digital strategy, business schools, sales consultant, operational processes provided support
Supervisado	maintained consistent communication with customers, nexus, overseeing, led the role of a project manager, my vision, degree, and maintained real estate project tracking and portfolio, systems, with market research, the restaurants daily operations, sustainability
LLM conversacional	Forbes standards, contract negotiation, coaching, multitasking, brand image and communication strategy, menu planning, budget & financial management, carrying out feasibility studies of new projects/refurbishment projects, design, managing large groups

Tabla 3 del artículo: diez anotaciones elegidas al azar de cada fuente, texto original en inglés.

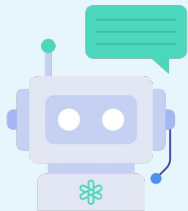
¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Significado vs. trazabilidad^{1,2}

01 · SIGNIFICADO

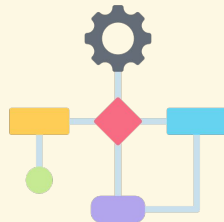
Los **LLM** se acercan más al significado humano.

Pero son **muy verbosos** y no respetan los **límites exactos** del texto.



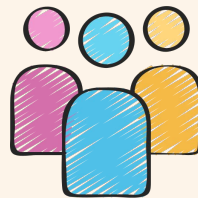
02 · TRAZABILIDAD

Reglas y métodos semánticos vinculan cada habilidad a un **fragmento exacto** del texto. Así se puede **auditar** cada decisión.



03 · CANTIDAD

El **LLM** extrae 1.7 veces más que los humanos. Pero más habilidades no es mejor: también aparecen fragmentos que **no son habilidades** («hotel led», «my vision»).



¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

² Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

Las decisiones

¿Qué significa esto para la regulación y para RR. HH.?



01

El CV

02

Los datos

03

Los sistemas

04

Personas vs. sistemas

05

Las decisiones

06

Tu postulación

Contratar con IA es de alto riesgo^{1,2,3}

UE

LEY DE IA DE LA UNIÓN EUROPEA (AI ACT)

- Las obligaciones recaen en las empresas que desarrollan y usan el sistema. El candidato debe ser informado y puede pedir una explicación (arts. 26 y 86).

The 4 risk levels under the EU AI Act

The EU AI Act's risk-based approach classifies AI systems into categories based on their potential harm to ensure appropriate regulatory measures are applied.

Unacceptable Risk

Definition: AI systems that pose a clear threat to safety, livelihoods, and rights, leading to their prohibition. Ex. Real-time Biometric Surveillance

High Risk

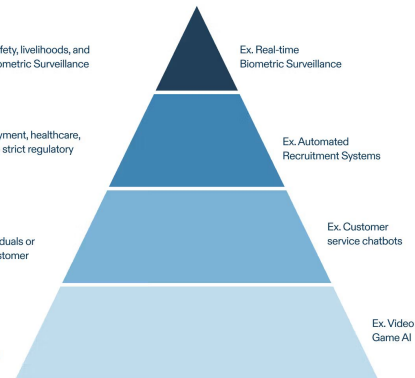
Definition: AI systems in critical sectors like employment, healthcare, and financial services (particularly lending), require strict regulatory compliance. Ex. Automated Recruitment Systems

Limited Risk

Definition: AI systems with limited impact on individuals or society, requiring transparency obligations. Ex. Customer service chatbots

Minimal Risk

Definition: AI systems with minimal or no risk, subject to voluntary codes of conduct and minimal regulatory oversight. Ex. Video Game AI



Nuestra posición: construir sistemas que **asisten** a RR. HH., no sistemas que deciden.

¹ Reglamento (UE) 2024/1689, Ley de IA de la Unión Europea, Anexo III.

² Reglamento (UE) 2026/1744, Ómnibus Digital sobre IA.

³ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

<https://optro.ai/blog/what-is-the-eu-ai-act>

Recomendaciones para la regulación¹

- 01 Guías claras:** definir qué es una habilidad, junto con profesionales de RR. HH.
- 02 Representatividad:** aclarar qué cuenta como representativo. ¿Sirve para CVs un sistema entrenado con ofertas?
- 03 Precisión vs. interpretabilidad:** aceptar algo menos de precisión a cambio de sistemas explicables.
- 04 Distinguir tipos de sistemas:** la ley usa una sola etiqueta, «alto riesgo», para sistemas muy distintos (reglas, supervisados, LLM).
- 05 Transparencia para candidatos:** derecho a entender cómo se tomó la decisión.



¹ Vásquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

Tu postulación

¿Qué puede hacer hoy una persona que busca empleo?



01

El CV

02

Los datos

03

Los sistemas

04

Personas vs. sistemas

05

Las decisiones

06

Tu postulación

Cómo postular hoy¹

01 · UN CV LEGIBLE POR MÁQUINAS

Texto seleccionable, una columna y títulos estándar (Experiencia, Educación, Habilidades).

02 · ADAPTA CADA POSTULACIÓN

Usa los términos de la oferta cuando describen lo que sí haces. **No inflas habilidades:** las etapas siguientes lo revelan.

03 · LEGIBILIDAD ANTES QUE DISEÑO

Frases cortas y términos claros. Evita tablas, íconos y barras de nivel en el CV que envías.

04 · EL CV BONITO, PARA CONFERENCIAS

Diseño, foto y gráficos para eventos y networking, donde lo lee una persona.

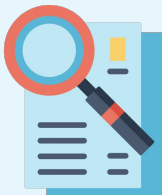
El ATS es solo el primer filtro. El objetivo es una relación laboral a largo plazo.

¹ Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

En resumen^{1,2}

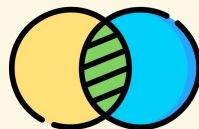
01 · LOS CVs SON DIFÍCILES

De ofertas a CVs, las habilidades bien identificadas caen casi a la mitad (**45.1%** → **26.6%** en inglés). Los CVs no siguen un formato.



02 · CADA SISTEMA LEE DISTINTO

Con el mismo CV, humanos, reglas, modelos supervisados y LLM extraen habilidades distintas en **cantidad**, **longitud** y **significado**.



03 · LA TRAZABILIDAD IMPORTA

En contratación hay que poder explicar de dónde sale cada habilidad: sistemas **auditables**, **supervisión humana** y **transparencia** para quien postula.



¹ Vázquez-Rodríguez et al. (2026). Skill Extraction from Resumes and Job Offers across Six Languages. SwissText 2026.

² Vázquez-Rodríguez et al. (2026). Comparing AI-based Approaches to Candidate Screening: Recommendations for Policy and Regulation. Data & Policy, 8, e35.

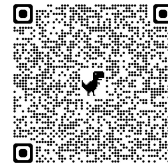
¡Muchas Gracias!

¿Preguntas?

Laura Vásquez-Rodríguez · Imvasque.github.io



SwissText 2026



Data & Policy 2026



Portfolio

PROYECTO Y SOCIOS



Créditos de la plantilla: Idiap.ch 2026 · Íconos de flaticon.com · Diapositivas preparadas con el apoyo de Claude