

How I found out about an AI system used by the Spanish govt to asses who is shirking work



INVESTIGATIONS NEWSROOMS ABOUT GOT A STORY? DONATE NOW



Learte / El Confidencial

COPUBLISHED WITH

SURVEILLANCE | SUSPICION MACHINES

Spain's AI Doctor

Unpublished documents and inside sources give rare look into secretive AI system used to decide who is shirking work

Spain has positioned itself as a leader in responsible AI within Europe. Yet promises of transparency have often fallen flat and public watchdogs have [struggled](#) to gain meaningful access to algorithms deployed in justice and welfare systems. This makes Spain part of a growing trend where

PUBLISHED APRIL 17, 2023 BY

Pablo Jiménez Arandia, Marta Ley, Gabriel Geiger, Manuel Ángel Méndez, Justin-Casimir Braun, Rocío Márquez, Eva Constantaras, Daniel Howden, Javier G. Jorrín, Rebeca Fernández, Ángel Villarino

SIMILAR INVESTIGATIONS



The story

- A set of ML algorithms to predict which workers on sick leave are ready to return to work
- The magnitude: millions of sick leave records under scrutiny every year
- The relevance: an AI involved in the provision of a social benefit
- The opacity: no public information on the system or previous reporting

The beginning (April 2022)

Social, siendo la Gerencia de Informática de la Seguridad Social encargada de coordinar y dar soporte a estos de Entidades y Servicios como la Tesorería General de la Seguridad Social, el Instituto Nacional de la Seguridad Social, la Inspección de Trabajo y Seguridad Social y el Servicio Jurídico de la Seguridad Social, entre otros.

Big Data en la Tesorería General de la Seguridad Social

la TGSS apuesta por la innovación y es pionera en la aplicación del Big Data en sus procesos de gestión. Para ello, se ha dotado de las herramientas tecnológicas más avanzadas en **inteligencia artificial** en pro del tratamiento analítico de los datos con el objetivo de la **detección y lucha contra el fraude**, incorporando en sus equipos de trabajo a **científicos de datos**, quienes definen patrones de comportamiento y establecen modelos predictivos.

Con este nuevo ecosistema de datos y tecnología, actualmente se monitorizan los datos de **1,5 millones de empresas y 18,6 millones de trabajadores** en activo. Mensualmente se procesa más de **7.000 millones de registros**, suponiendo un movimiento de información de 3.000Gb. EL Plan integral de lucha contra el fraude permite aprovechar valor de la ingente cantidad de datos acumulados para establecer controles en materia de afiliación, cotización y recaudación del Sistema.

Dentro de la lucha contra el fraude la TGSS decidió priorizar varias líneas de trabajo, las cuales ya están dando sus frutos. La utilización de indicadores y alertas de riesgos facilitan la detección, corrección o prevención de estos casos con una rapidez y eficacia mucho más avanzada ofrecen una valiosa información para atacar el fraude de una manera conjunta. La **detección de empresas ficticias** es la línea de trabajo más madura, la cual persigue tres tipos de fraudes: empresas sin actividad real, empleados ficticios en empresas con actividad y autónomos ficticios. el balance en esta línea de actuación ha sido de **12.170 millones** (diciembre 2017) y **104.347 trabajadores afectados**, mejorando la

Big Data en el Instituto Nacional de la Seguridad Social

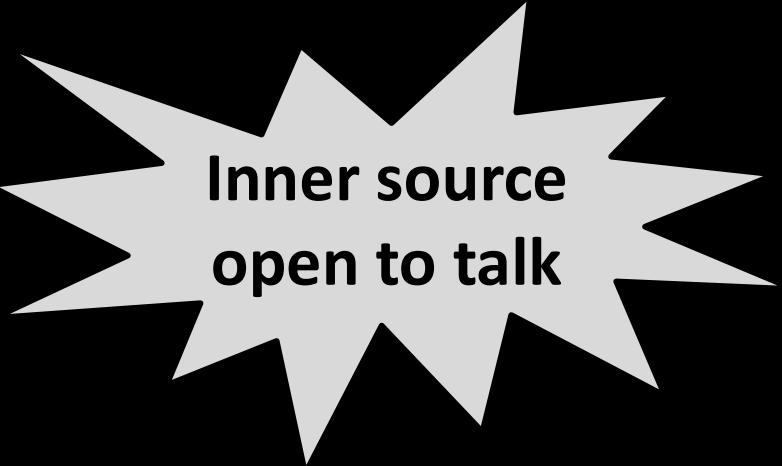
En el INSS no se quedan atrás en la aplicación de las nuevas técnicas de análisis de datos que ofrece Big Data. La Entidad también está haciendo grandes esfuerzos en mejorar los servicios y optimizar resultados con la línea estratégica común de **la lucha contra el fraude**, en este caso, para **mejorar la correcta gestión de las prestaciones económicas** que ofrece la Seguridad Social.

Las líneas de trabajo desplegadas por el INSS en la aplicación de las tecnologías más avanzadas de **análítica avanzada** han consistido en **definir modelos para optimizar el control médico del INSS ante las incapacidades temporales**, a través de la herramienta SAS, a modo de programa piloto en seis direcciones provinciales, permitiendo no citar a un reconocimiento médico a los casos que están correctamente justificados dentro del periodo de los primeros **365 días**, en estos casos, los modelos predictivos nos indican de manera mucho más ágil **cuales son los casos susceptibles de la actuación de los inspectores médicos**.

En relación al correcto reconocimiento y duración de las prestaciones, el Big Data permite la creación de **cuadros de mando**, que gracias al Sistema MBI, destinado a conocer el comportamiento de las prestaciones, ayuda a la toma de decisiones y la implantación de nuevos controles. Los cuadros de mando entrelazados permiten obtener patrones de comportamiento personalizados al tipo de ámbito que pueda interesar, tanto geográfica, como profesional o económica y con ellos la posibilidad de realizar estudios comparativos de las distintas prestaciones(actualmente de la incapacidad temporal).

Como futura línea de trabajo se pretende identificar los factores que condicionan la **prolongación de procesos de baja** más allá de los 365 días, o incluso 545 sin que sean calificadas como incapacidades permanentes. Otro modelo en desarrollo pretende identificar **desviaciones** en el comportamiento de determinados procesos de incapacidad temporal con respecto a otros procesos similares en función de parámetros como el diagnóstico o la actividad económica o laboral.

Respecto del análisis de la **incapacidad temporal**, el Big Data ofrece la realización de estudios según patologías y duración estándar, en función de las distintas variables, ofreciendo optimizar las **tablas de duración estándar** de los procesos de incapacidad temporal y la adecuada selección de los **procesos susceptibles de reconocimiento médico**, permitiendo detectar las motivaciones de los excesos de duración de las incapacidades temporales y favoreciendo una actuación más correcta sobre ellos.

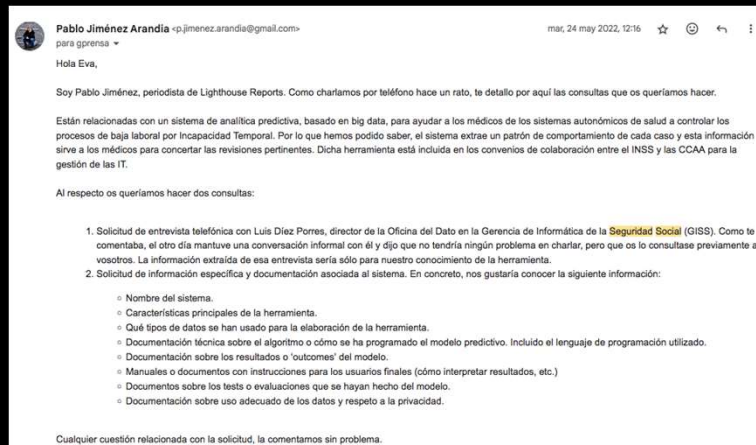


Inner source
open to talk

Problems, FOIA, alternatives

Press request (May 2022)

The answer: 4 lines



- FOIA time (September 2022)
- But thinking on alternatives: inner source, LinkedIn searches for doctors, empty calls...

Valuable information

- Medical inspectors who receive the algorithm-generated scores
- We have a FOIA response! (December 2022)



MINISTERIO
DE INCLUSIÓN, SEGURIDAD SOCIAL
Y MIGRACIONES

SECRETARÍA DE ESTADO
DE LA SEGURIDAD SOCIAL
Y PENSIONES

GERENCIA DE INFORMÁTICA
DE LA SEGURIDAD SOCIAL

Nº Expediente: 001-071994
Solicitante: PABLO JIMENEZ ARANDIA
NIF: 44473097K

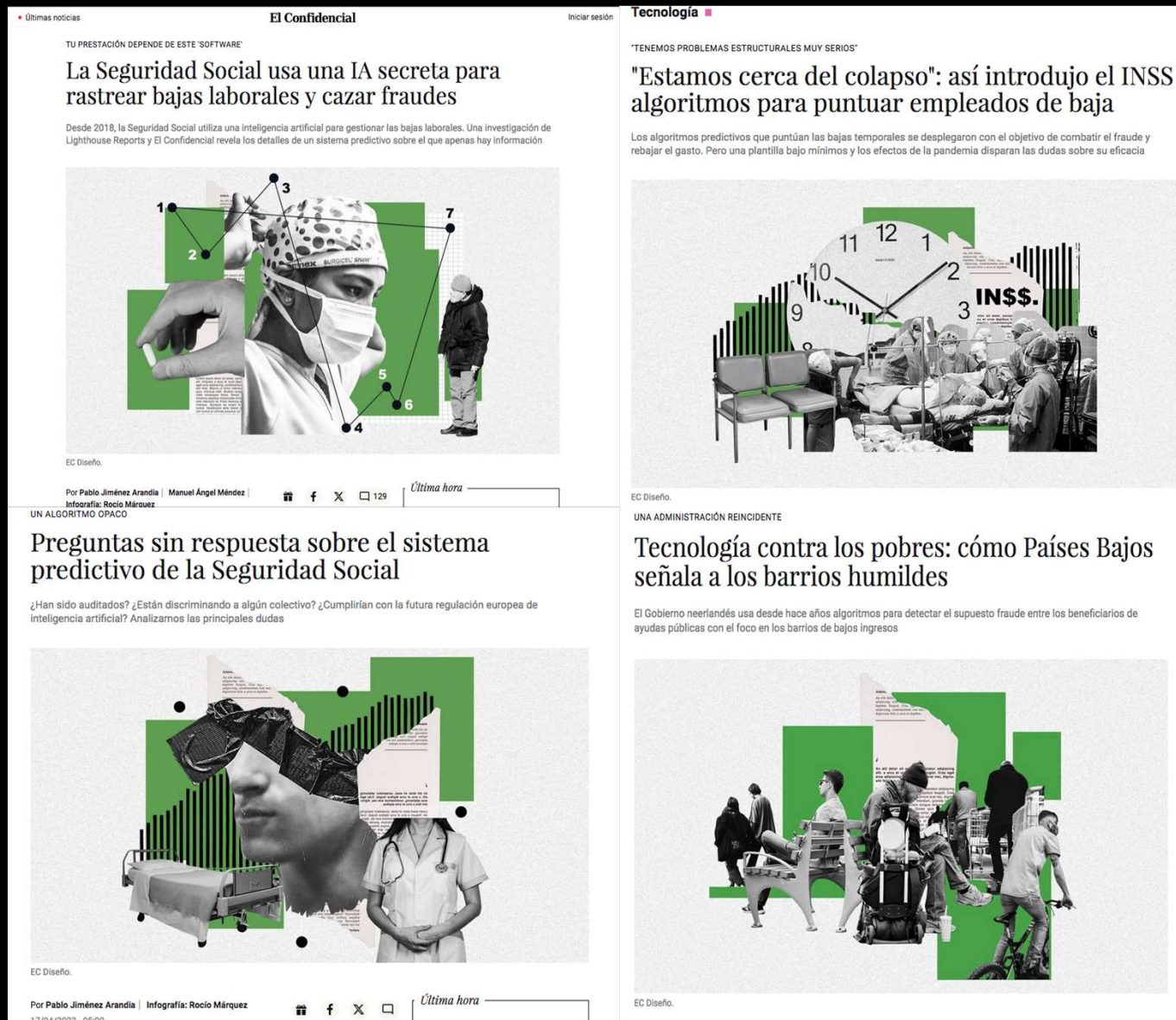
Asunto: Resolución en materia de acceso a la información presentada al amparo del artículo 12 de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (en adelante también denominada "LTAIBG").

En respuesta a la solicitud presentada por D. PABLO JIMENEZ ARANDIA recibida en esta unidad el 7 de octubre de 2022 ante el portal de transparencia de la Tesorería General de la Seguridad Social, y en la que solicita información sobre "*Solicitud de información sobre el modelo predictivo usado en la gestión de las bajas laborales por incapacidad temporal.*" la Gerencia de Informática de la Seguridad Social, como órgano competente, adopta la siguiente **RESOLUCIÓN**:

Una vez analizada la solicitud y en referencia a esta Gerencia de Informática de la Seguridad Social se resuelve admitir la solicitud **reconociendo el derecho del interesado a acceder a la información solicitada**, comunicándole que, la Gerencia de Informática de la Seguridad Social le facilita, a través del Portal de Transparencia, los siguientes documentos en contestación a las preguntas formuladas en la solicitud de la siguiente manera:

- 1-algoritmo.pptx: punto 1 de la solicitud, referida a "*Documentación técnica relacionada con el algoritmo, el código y el modelo predictivo usados por el INSS para la gestión de las bajas laborales, incluidos los manuales, los powerpoints técnicos y los memorandos explicativos.*"

The outcome: a rare look into a secretive AI system



Links and contact

- Summary and methodology
 - [Spain's AI Doctor - Lighthouse Reports](#)
- Publications:
 - [La Seguridad Social usa una IA secreta para rastrear bajas laborales y cazar fraudes](#)
 - ["Estamos cerca del colapso": así introdujo el INSS algoritmos para puntuar empleados de baja](#)
 - [Preguntas sin respuesta sobre el sistema predictivo de la Seguridad Social](#)
 - [Tecnología contra los pobres: cómo Países Bajos señala a los barrios humildes](#)

pjimenezarandia@proton.me
@pjarandia (TW)