



PROGRAMA IBERDROLA INNOVA I+D+i 2021/2022

LÍNEA DE INVESTIGACIÓN

PRESENTE Y FUTURO DE LA MEDICINA ASISTENCIAL EN LA ERA DIGITAL

**Fuentes de datos abiertos en Salud a nivel mundial y el impacto de su
utilización**

Javier Olvera Calderón

ÍNDICE

- **INTRODUCCIÓN**

- Bases del trabajo
- Objetivos

- **INVESTIGACIÓN Y METODOLOGÍA APLICADA**

- Información básica sobre el tema
- Caso práctico
- Ejemplos de páginas de Open Data

- **CONCLUSIONES Y RESULTADOS**

- **REFERENCIAS**

INTRODUCCIÓN

Cada vez, la información está más digitalizada, las cartas se reemplazan por correos electrónicos, los documentos se digitalizan y la información se guarda en bases de datos. Un ejemplo de esto son los hospitales, en los cuales los ficheros de pacientes que contienen su nacimiento, vacunas, operaciones, etc. están digitalizados.

En este sentido, la Open Data consiste en información anónima digitalizada destinada a su divulgación. Esta se usa para realizar estudios teniendo de base una información certera pero encontrar información Open Data puede ser verdaderamente tedioso debido a que esta cuenta con ciertas características para poder ser considerada como tal, las cuales se explicarán más adelante.

Un ejemplo de la aplicación de la OpenData a la medicina sería obtener datos abiertos de información procedente de hospitales que contenga (de forma anónima) información sobre pacientes: edad, peso, enfermedades, grupo sanguíneo. Si se pudiera programar una inteligencia artificial que fuese capaz de leer y analizar los datos de pacientes y en base a ellos sacar conclusiones como por ejemplo: (1) ¿qué rangos de edad tienden más a infectarse por la cepa OMICRON del covid19?, (2) ¿qué peso promedio suelen tener las personas que son obesas? Con esto se podría sacar fácilmente conclusiones sobre la evolución de un virus o enfermedad y por ejemplo sacar conclusiones como qué edades hay que vacunar primero ya que presentan síntomas más severos.

Esto nos lleva a hablar sobre inteligencias artificiales, estas son «la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana» (Petteri Rouhiainen, 1018). Estas toman “decisiones” en base a algoritmos, y debido a que un ordenador es capaz de analizar datos, realizar operaciones, etc. cientos de veces más rápido que los humanos. Así, es más rentable que estos cálculos en base a algoritmos complejos o redes neuronales sean realizados por ordenadores, ya que un código que se encargue de calcular rápidamente porcentajes, medias, etc. de cientos, miles o millones de ficheros de pacientes, podría ser reutilizado cada vez que surja una nueva enfermedad, virus, o síntoma a nivel local, nacional o global.

La pregunta que hay que plantearse ahora es ¿Qué quiere decir todo esto? Que si un programa es capaz de leer y analizar datos de miles de pacientes en segundos, es mucho más efectivo que tener a personas trabajando en esto. El reto aquí sería que el programa o inteligencia artificial sea capaz de acceder por sí solo a la página web de la que se ha

extraído la información (Open Data), ya que si alguien debe meter toda esa información manualmente volvemos al problema de antes, es poco efectivo.

Para alcanzar estos logros es fundamental la integración de todos los datos procedentes de muy diferentes fuentes, así como el desarrollo de nuevas tecnologías que permitan la explotación de dichos datos. No obstante, el verdadero valor de Big Data en salud se conseguirá sólo si los diferentes actores implicados en el proceso (administraciones públicas, empresas privadas, hospitales, médicos, centros de investigación, universidades, ...) se comprometen en este proyecto de forma conjunta para llevar el ámbito sanitario a una nueva era. Esto solo se podrá llevar a cabo en el marco de un ecosistema de Big Data en salud en el que se integre conjuntamente con la tecnología, políticas adecuadas sobre privacidad y confidencialidad, infraestructuras y una cultura de uso compartido de los datos. Todo ello conlleva una serie de retos que hay que afrontar desde diferentes perspectivas y grado de profundidad. (Menasalvas, et al., 2017).

Por todo ello, los principales objetivos del trabajo son:

- 1- Investigar de qué páginas es fiable extraer información Open Data para almacenarlas, creando así una pequeña librería de la que pueda sacar datos abiertos.
- 2- Buscar los lados positivos y negativos de encargar a una Inteligencia Artificial (IA) la tarea de sacar conclusiones médicas que pueda llegar a ser vital para la vida de personas o estudios relacionados con ello.
- 3- Investigar qué personas estarían de acuerdo con esto, ya que como en cualquier otra situación, asignarle a una máquina el trabajo que anteriormente realizaba alguien siempre tiene a personas de acuerdo y en contra.
- 4- Buscar si esto se ha hecho antes (cosa que es bastante posible) y en caso de que sí, cuántas investigaciones actuales, como por ejemplo en el caso del COVID19, han utilizado programas o inteligencias artificiales basadas en datos abiertos para sacar conclusiones.
- 5- En una segunda fase del trabajo, en caso de ser seleccionado, desarrollaremos un programa que mediante cálculos, machine learning o cualquier otro método, saque conclusiones sobre ficheros médicos de Open Data para hacer automáticamente estudios sobre enfermedades en una población.

¿De qué manera se va a investigar esto? La intención es organizar los objetivos de este trabajo en 3 sectores.

En primer lugar, se explicarán todas las bases para entender sobre el tema tratado en este trabajo: qué es una inteligencia artificial (IA), su aplicación a la open data, junto con una explicación a fondo sobre esta y sus aplicaciones a la medicina. También hablaremos de bases de datos de las que es fiable extraer información para estudios. Para esto se usará parte de la información obtenida gracias a una entrevista realizada a José Francisco Ávila de Tomás, médico de familia e investigador predoctoral de la práctica asistencial del profesional sanitario en la era digital .

En segundo lugar, nos centraremos en la explicación del funcionamiento de una pequeña simulación de una IA con bases de datos que realicé para mostrarlo como ejemplo.

Finalmente, considero necesario mostrar páginas web de las que sea fiable extraer Open Data para poder utilizar dicha información en nuestra simulación.

INVESTIGACIÓN Y METODOLOGÍA APLICADA

¿Qué es la Open Data?

La Open Data (o datos abiertos), es información anónima destinada al uso público con la finalidad de destinarla principalmente a estudios, trabajos... Para que cierta información pueda ser considerada como Open Data debe cumplir ciertas características:

- 1- Deben ser Datos Abiertos. A través de Google podemos obtenerlos aunque no es siempre fiable del todo, por lo que hay que asegurarse que la fuente es fiable y que no está bajo derechos de autor.
- 2- Deben ser anónimos. Si se desea desarrollar un estudio en base a estos datos, no se debe revelar información privada de las personas involucradas en la Open Data (en caso de que esta incumbe a personas).
- 3- Hay que asegurarse de que el trabajo que se está desarrollando con estos datos cumplen con la ley, esto puede resultar obvio pero en multitud de casos un despiste en olvidarse de acreditar al propietario de una imagen utilizada para decorar un simple trabajo puede dar problemas ya que es ilegal usar material de otros sin acreditarlos.

4- Tener cuidado con mezclar datos abiertos con datos privados. Si se desea publicar los resultados de un estudio basado en estos datos, no se deben revelar los datos privados en los que este se ha basado sin autorización previa.

5- Saber si es ético utilizarlos para tu propósito. Es imposible tener un listado de normas de casos en los que está bien y mal usar datos abiertos para un propósito determinado, entrando en conflicto con la ética.

Una vez explicado esto hablemos del principal problema que podemos encontrar a la hora de querer usarla, y es que hoy en día cualquier persona que requiera información o tenga una duda sobre algún tema, va a buscar en Google y la primera respuesta que obtenga la va a dar por válida. Es cierto que a veces esta respuesta va a ser correcta, pero cuando realizamos estudios o trabajos importantes no podemos arriesgarnos a simplemente confiar en lo que encontramos. Para ello mucha gente contrasta lo que encuentra en varias páginas web a ver si es lo mismo, o investiga si la persona que ha redactado la página web de la que ha sacado la información es experta en ese campo, y aunque esto reduce las posibilidades de que sea falsa información, no las elimina al completo. ¿Qué hacemos entonces? Al buscar en Open Data, debemos centrarnos en las páginas más populares y sobre todo de fuentes fiables, por ejemplo, una página oficial del Gobierno de España como sería datos.gob.es. Otro lugar del que nos podemos fiar es buscando qué fuentes usaron otros investigadores en sus tesis doctorales o trabajos de investigación, los cuales ya han sido corroborados y contrastados.

Así que una vez que ya sabemos que nuestra fuente es fiable, ¿para qué queremos dicha información? Planteemos un caso reciente para todos, el COVID19. Si disponemos de una base de datos abierta con información sobre el número de contagios por regiones, síntomas de cada contagiado... podemos realizar un estudio sobre, por ejemplo, qué síntomas se repiten más en los pacientes y cuáles se repiten menos. De esta manera podemos dar prioridad para, por ejemplo, realizar una PCR a los pacientes cuyos síntomas están más altos en la lista de síntomas más comunes de este virus.

Inteligencia Artificial (IA)

Ahora ya sabemos como obtener datos abiertos y qué aplicaciones tienen, pero ¿qué papel juega aquí la IA? Pongamos de nuevo el ejemplo anterior sobre el COVID19, si deseamos hacer un estudio sobre los síntomas de este, al ser una enfermedad a nivel mundial puede resultar tedioso o imposible ir manualmente apuntando y calculando los porcentajes de repetición de cada síntoma en los contagiados a nivel mundial. Aquí entra la IA, ya que como sabemos, un ordenador es capaz de calcular en menos de un segundo más información de la que somos capaces de imaginar, lo que reduce el tiempo y esfuerzo de los investigadores. Para ello simplemente se puede hacer un simple código que lea y guarde los síntomas de los pacientes y los compare, calculando sus porcentajes de repetición, pudiendo así obtener resultados sobre una población de cientos, miles o millones de personas en cuestión de minutos o segundos.

En el ejemplo expuesto, es cierto que lo que se necesita calcular es algo simple, pero una IA es capaz de mucho más.

En una entrevista con José Francisco Ávila de Tomás, un médico de familia experto en el tema de Open Data e inteligencia artificial en la práctica asistencial, hablamos sobre esto. Él denominó a una Inteligencia Artificial como *la capacidad de las máquinas de abstracción de los datos y ser capaces de “pensar”*. Nos habló sobre las 2 primeras capas de una IA, en primer lugar tenemos el sistema experto, el cual es un algoritmo que la IA sigue de forma predeterminada. Esto se podría considerar como una inteligencia artificial básica. Y luego tendríamos la segunda capa, que es el Sistema Programación del Lenguaje Natural (PLN), cuando una IA es capaz de entender el lenguaje humano (Ávila-Tomás et al., 2020a) y (Ávila-Tomás et al., 2020b).

Una vez entendemos las bases de la Open Data y la inteligencia artificial, ¿a qué facetas de la medicina podemos aplicar una inteligencia artificial y bases de datos abiertos? Esto también se lo preguntamos a José Francisco Ávila, quien nos explicó las distintas facetas de la salud. Estas serían:

- Faceta asistencial: recibir, diagnosticar, asistir... pacientes.
- Faceta de investigación: investigar para continuar con el avance de la ciencia.
- Faceta docente: formar a futuras personas en el campo de la medicina.
- Faceta de gestión: saber gestionar el dinero, recursos, tiempo, materiales, personal...

José Francisco también nos habló sobre las aplicaciones de Inteligencia Artificial y Open Data en cada una de estas facetas, comencemos por la asistencial, que es además una de las más interesantes para este trabajo de investigación.

En esta, cada médico o enfermero deberá documentar datos sobre sus pacientes y guardarlos de forma anónima en una base de datos abierta a nivel nacional o internacional. Pongamos un ejemplo, supongamos que tenemos a un paciente que padece de tuberculosis, si todos los síntomas que han padecido personas con esta enfermedad se suben a una base de datos abierta, podemos saber todos los posibles síntomas y cuáles son más o menos comunes, para facilitar los diagnósticos en base a porcentajes, cuantas más variables, este será más preciso. De igual forma con los tratamientos, cuando tengamos que proponerle un tratamiento a un paciente con cáncer, podemos analizar con una IA toda la información de nuestra base de datos abiertos y saber los porcentajes de éxito de dicho tratamiento, al igual que sus posibles efectos secundarios. Esto sería muy útil para las llamadas “enfermedades raras”, las cuales afectan a 1 persona por cada 100.000 habitantes, ya que facilita el diagnóstico o tratamiento de estas sacando información de dicha base de datos. Esto ya se hace en algunos hospitales pero en el futuro sería ideal hacerlo a nivel nacional, internacional o global, con grandes redes de comunicación e investigación. También se aplica a la medicina personalizada, la cual se considera el futuro de la medicina. Esta trata de dar soluciones personalizadas para cada paciente, teniendo en cuenta todas sus circunstancias, la IA analiza y busca coincidencias de pacientes bajo circunstancias similares en una gran base de datos.

Las aplicaciones de la faceta de investigación son similares, ya que igual que antes se subía y recopilaba información de una base de datos abiertas para el diagnóstico o tratamiento, ahora se hace lo mismo pero con fines de investigación y mejora del conocimiento científico en este campo.

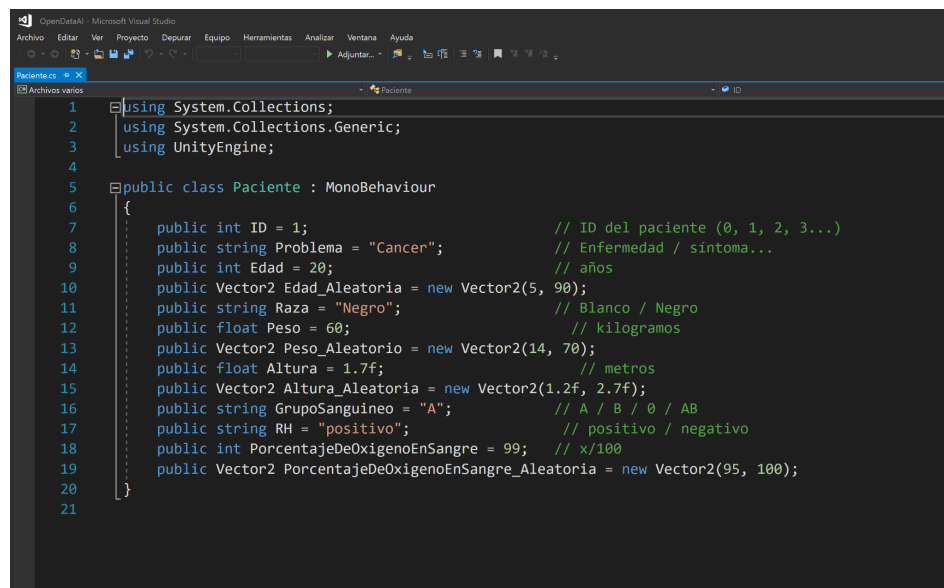
En la faceta docente, uno de los ejemplos que José Francisco nos puso, fue un estudio realizado en una Universidad Francesa, en la cual, se entrenó a una inteligencia artificial para entrevistar a alumnos y deducir su estado de ánimo, salud, capacidad de concentración... y buscar soluciones a sus problemas mejorando así el rendimiento académico. El estudio fue un éxito y aumentó el rendimiento académico de los alumnos.

Finalmente, en la faceta de gestión, no hay tantas aplicaciones a la medicina aunque sí que se pueden usar IA para gestionar presupuestos o personal de empresas.

Caso práctico: simulación de IA y Open Data creada para el trabajo

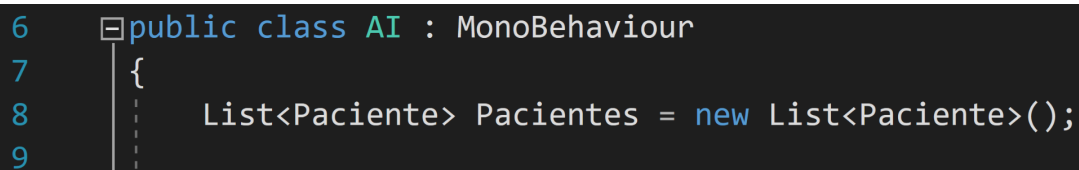
Una vez hablado de la Open Data, Inteligencias Artificiales y sus aplicaciones en la medicina, veamos un ejemplo que realicé para este trabajo. Consiste en una simulación de un base de datos aleatoria con información anónima de pacientes y una IA que analiza estos datos y saca porcentajes probabilísticos de enfermedades en base a la repetición de características o síntomas. Esto se ha realizado en el motor Unity y se ha programado en el lenguaje C#.

Comenzamos creando una código denominado *Paciente* que contenga como variables todas las características que va a almacenar mi simulación de una OpenData. Las características que vamos a usar son un ejemplo para crear una base de datos con ella.



```
1 using System.Collections;
2 using System.Collections.Generic;
3 using UnityEngine;
4
5 public class Paciente : MonoBehaviour
6 {
7     public int ID = 1; // ID del paciente (0, 1, 2, 3...)
8     public string Problema = "Cancer"; // Enfermedad / sintoma...
9     public int Edad = 20; // años
10    public Vector2 Edad_Aleatoria = new Vector2(5, 90);
11    public string Raza = "Negro"; // Blanco / Negro
12    public float Peso = 60; // kilogramos
13    public Vector2 Peso_Aleatorio = new Vector2(14, 70);
14    public float Altura = 1.7f; // metros
15    public Vector2 Altura_Aleatoria = new Vector2(1.2f, 2.7f);
16    public string GrupoSanguineo = "A"; // A / B / 0 / AB
17    public string RH = "positivo"; // positivo / negativo
18    public int PorcentajeDeOxigenoEnSangre = 99; // x/100
19    public Vector2 PorcentajeDeOxigenoEnSangre_Aleatoria = new Vector2(95, 100);
20 }
21
```

A continuación, en otro código llamado *AI* vamos a generar una lista con una cantidad de pacientes predefinida y vamos a aleatorizar los datos que contengan.



```
6 public class AI : MonoBehaviour
7 {
8     List<Paciente> Pacientes = new List<Paciente>();
9 }
```

También crearemos otras listas para los valores aleatorios no numéricos de nuestros pacientes. Por ejemplo, una lista con nombres de enfermedades para asignarle una de forma aleatoria a cada paciente de nuestra simulación. También crearemos varias variables para realizar más adelante el estudio sobre nuestra simulación de una OpenData.

```
[Header ("Variables:")]
public int NumeroDePacientes = 5;
public List<string> Problemas;
public List<string> Razas;
public List<string> GruposSanguineos;
public List<string> RH;

[Header("Estudio")]
public string ProblemaInvestigando;
List<Paciente> PacientesEnfermos = new List<Paciente>();
public float PorcentajeEnfermo;
public Vector2 RangoEdades= new Vector2(5, 90);
public List<int> Edades;
public List<int> RepeticionEdades;
public List<int> EdadesEnfermasMasRepetidas;
public List<int> EdadesMasRepetidas;
public List<string> StringEdadesMasRepetidas;
public Text FinalResultText;
```

Además, se van a necesitar otras variables que en este caso, al usar el motor Unity, vamos a tener que crear. Estas están relacionadas con componentes de los elementos, botones y textos que deberemos mostrar en pantalla, junto con los resultados de la Inteligencia Artificial. Pero son aspectos más orientados al motor gráfico y no a la Open Data o IA, por lo que los omitiré en este trabajo, pero cabe destacar que están ahí y son vitales para que todo funcione.

A continuación, crearemos una función llamada *CrearSimulación* que se encargará de generar la lista de los pacientes que van a formar parte de nuestra Open Data y aleatorizarán sus características.

```
public void CrearSimulacion()
{
    //Resetea la Lista para no interferir con los datos de la simulación anterior
    Pacientes.Clear();

    //Designa el número de Pacientes que habrá en ella
    for (int i = 0; i < NumeroDePacientes; i++)
    {
        Pacientes.Add(new Paciente());
    }

    //Aleatoriza sus rasgos
    for (int i = 0; i < Pacientes.Count; i++)
    {
        Pacientes[i].ID = i;
        Pacientes[i].Problema = Problemas[Random.Range(0, Problemas.Count)];
        Pacientes[i].Edad = (int)Random.Range(Pacientes[i].Edad_Aleatoria.x, Pacientes[i].Edad_Aleatoria.y);
        Pacientes[i].Raza = Razas[Random.Range(0, Razas.Count)];

        //Peso *Comprobar
        if (Pacientes[i].Edad >= 5 && Pacientes[i].Edad <= 10)
        {
            Pacientes[i].Peso = Random.Range(35, 40);
        }
        if (Pacientes[i].Edad >= 11 && Pacientes[i].Edad <= 20)
        {
            Pacientes[i].Peso = Random.Range(45, 70);
        }
        if (Pacientes[i].Edad >= 21 && Pacientes[i].Edad <= 50)
        {
            Pacientes[i].Peso = Random.Range(60, 90);
        }

        //Altura *Comprobar
        if (Pacientes[i].Edad >= 5 && Pacientes[i].Edad <= 10)
        {
            Pacientes[i].Altura = Random.Range(1, 1.4f);
        }
        if (Pacientes[i].Edad >= 11 && Pacientes[i].Edad <= 20)
        {
            Pacientes[i].Altura = Random.Range(1.5f, 1.8f);
        }
        if (Pacientes[i].Edad >= 21 && Pacientes[i].Edad <= 50)
        {
            Pacientes[i].Altura = Random.Range(1.7f, 2.1f);
        }

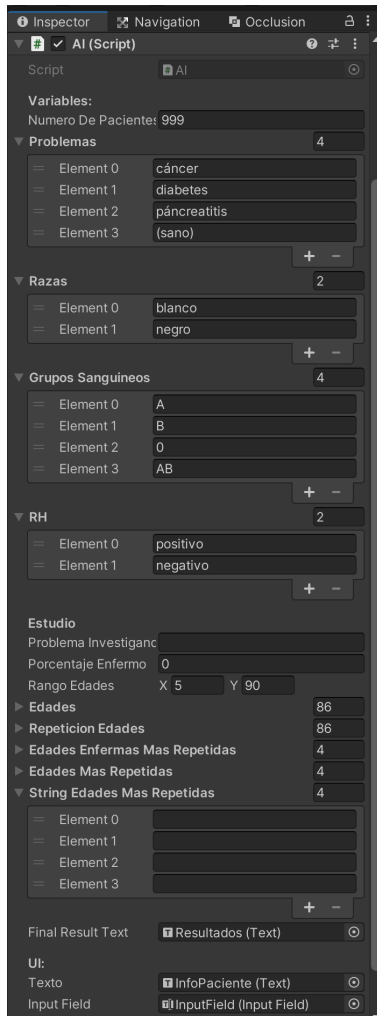
        Pacientes[i].GrupoSanguíneo = GruposSanguíneos[Random.Range(0, GruposSanguíneos.Count)];
        Pacientes[i].Rh = Rh[Random.Range(0, Rh.Count)];
        Pacientes[i].PorcentajeDeOxígenoEnSangre = (int)Random.Range(Pacientes[i].PorcentajeDeOxígenoEnSangre_Aleatoria.x, Pacientes[i].PorcentajeDeOxígenoEnSangre_Aleatoria.y);
    }
}
```

Como se puede ver en la imagen anterior, primero asignamos la ID del paciente (para poder buscarlos en la lista, ya que al ser una OpenData no podemos poner información privada como por ejemplo un nombre o apellidos), después le asignamos una enfermedad de nuestra lista de enfermedades y también su raza.

Después, comprobamos qué edad tiene nuestro paciente, y dependiendo de esta aleatorizaremos su peso y altura entre distintos valores, esto es para no obtener como resultado de nuestro paciente aleatorio, por ejemplo, a un niño de 5 años que mide 1,70 y pesa 30kg.

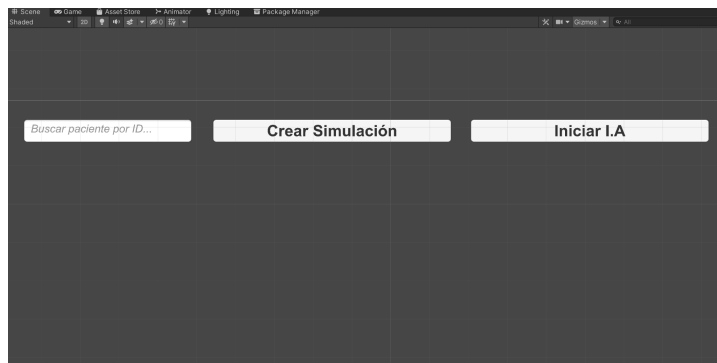
Después generamos también su grupo sanguíneo, Rh, y porcentaje de oxígeno en sangre.

Ahora volvamos al motor Unity, donde si asignamos nuestro código *AI* a un objeto, veremos que se despliegan todas las variables que hemos creado para poder modificarlas directamente desde el programa y no tener que estar editando el código constantemente.



He asignado unos valores predeterminados para la simulación. Estos se pueden observar en la imagen de la izquierda.

Lo próximo que debemos hacer es crear una interfaz básica para controlar la ejecución de funciones como la creada anteriormente (*CrearSimulación*) y otras que programaremos más adelante. También crearemos un buscador de texto para introducir el ID de un paciente y mostrar así su información aleatoria. También crearemos un texto vacío donde se mostrarán los datos de este paciente. Finalmente crearemos un botón para iniciar la simulación y calcular los porcentajes de repetición de las enfermedades o síntomas, lo cual programaremos más adelante. Este sería el resultado:



Volvamos al código *AI*. Ahora crearemos una función llamada *BuscarPaciente* y otra llamada *MostrarFichero*. La primera función la ejecutaremos al introducir un número en el buscador creado anteriormente, y comprobaremos si existe algún paciente con un ID igual al número introducido. De ser así ejecutaremos la segunda función. La cual escribirá en pantalla todos los datos de nuestro paciente.

```
public void BuscarPaciente()
{
    int id = int.Parse(InputField.text);

    if (id < Pacientes.Count)
    {
        MostrarFichero(id);
    }
    else
    {
        Texto.text = "*No se ha encontrado al paciente";
    }
}
```

```
//Mostrar fichero de paciente
public void MostrarFichero(int PacienteMostrar)
{
    Texto.text = "Paciente: " + Pacientes[PacienteMostrar].ID.ToString() + "\n \n" +
        "Problema: " + Pacientes[PacienteMostrar].Problema + "\n" +
        "Edad: " + Pacientes[PacienteMostrar].Edad.ToString() + " años" + "\n" +
        "Raza: " + Pacientes[PacienteMostrar].Raza + "\n" +
        "Peso: " + Pacientes[PacienteMostrar].Peso.ToString() + "kg" + "\n" +
        "Altura: " + Pacientes[PacienteMostrar].Altura.ToString() + "m" + "\n" +
        "Grupo Sanguíneo: " + Pacientes[PacienteMostrar].GrupoSanguíneo + "\n" +
        "RH: " + Pacientes[PacienteMostrar].RH + "\n" +
        "Porcentaje de oxígeno en sangre: " + Pacientes[PacienteMostrar].Porcentaje
```

Hasta ahora hemos creado una base de datos con pacientes ficticios aleatorios (tratando de cumplir las características de una OpenData) y una interfaz que nos muestra esos datos. A continuación llega la parte complicada, y es crear la Inteligencia Artificial que debe analizar todos los datos de esta OpenData ficticia que hemos desarrollado y hacer que obtenga porcentajes de repetición de características en personas para relacionarlo con enfermedades. Una de las características que creamos fue el peso, pues la idea es que nuestra IA compare el peso de, por ejemplo, todas las personas con cáncer, para que nos diga qué pesos se repiten más en personas que padezcan de esto y sacar como conclusión que si pesas entre “x” e “y” kilogramos tienes más probabilidad de padecer cáncer. Obviamente este no es un ejemplo coherente si lo llevamos a la realidad, ya que estamos generando datos aleatorios, pero si reemplazamos esta base de datos creada por nosotros por una real, las conclusiones que nuestra IA saque serán también reales.

Para empezar, haremos que nuestra inteligencia artificial analice los datos de la edad, para ello crearemos una nueva función llamada *GetResult* en la que primero almacenaremos a los pacientes que deseamos estudiar (en este caso los que hayan enfermado con la enfermedad especificada en la variable *problema*).

```
//Obtengo los pacientes con el problema que quiero estudiar
for (int i = 0; i < Pacientes.Count; i++)
{
    if(Pacientes[i].Problema == problema)
    {
        PacientesEnfermos.Add(Pacientes[i]);
    }
}
```

Después calcularemos el porcentaje de enfermos que hay entre todos los pacientes y almacenaremos en una lista llamada *RepeticiónEdades* cuantas veces se repiten las edades más repetidas de nuestro estudio.

```
PorcentajeEnfermo = (PacientesEnfermos.Count-1) * 100 / (Pacientes.Count-1);

for (int i = 0; i < PacientesEnfermos.Count; i++)
{
    for (int u = 0; u < RepeticionEdades.Count; u++)
    {
        if(PacientesEnfermos[i].Edad == Edades[u])
        {
            RepeticionEdades[u]++;
        }
    }
}
```

Finalmente escribiremos un texto diciendo: “De personas de \underline{x} , \underline{y} ... años han enfermado \underline{z} personas de cada edad”.

```
//Primer puesto
StringEdadesMasRepetidas[0] = "De las personas de ";
for (int i = 0; i < RepeticionEdades.Count; i++)
{
    if(RepeticionEdades[i] > EdadesEnfermasMasRepetidas[0])
    {
        StringEdadesMasRepetidas[0] = "De las personas de ";
        EdadesEnfermasMasRepetidas[0] = RepeticionEdades[i];
        EdadesMasRepetidas[0] = Edades[i];
        StringEdadesMasRepetidas[0] += Edades[i].ToString() + " años";
    }
    else if (RepeticionEdades[i] == EdadesEnfermasMasRepetidas[0])
    {
        StringEdadesMasRepetidas[0] += ", " + Edades[i].ToString() + " años";
    }
}
StringEdadesMasRepetidas[0] += " han enfermado " + EdadesEnfermasMasRepetidas[0].ToString() +
" personas de cada edad.";
```

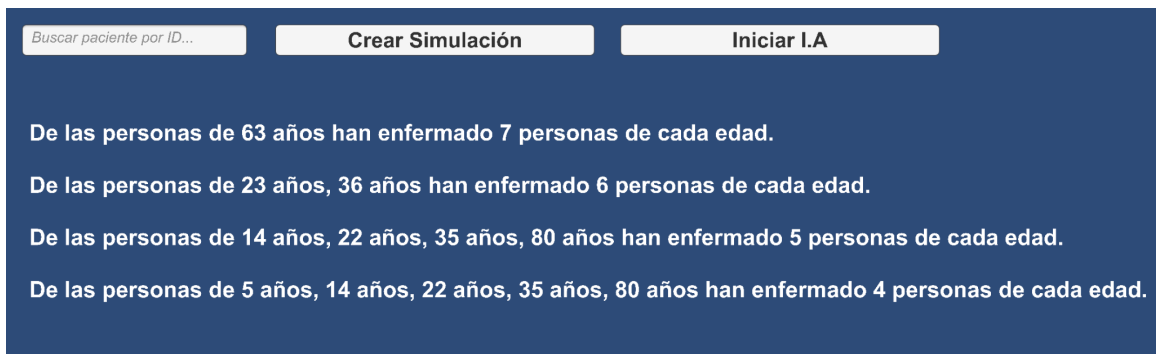
Esto lo repetiremos 4 veces, creando así una lista con los 4 grupos de edades en los que más se ha padecido la enfermedad que se está investigando.

El resultado final sería el siguiente; si vamos de nuevo al motor de Unity y ejecutamos el proyecto, al pulsar el botón de *Crear Simulación*, generaremos la base de datos aleatoria, en la cual podemos buscar pacientes según su ID introduciendo este en el campo de búsqueda.

Paciente: 5

Problema: diabetes
Edad: 88 años
Raza: blanco
Peso: 60kg
Altura: 1,7m
Grupo Sanaguíneo: B
RH: positivo
Porcentaje de oxígeno en sangre: 99%

Si después pulsamos el botón *Iniciar Inteligencia Artificial*, nos aparecerán en pantalla los resultados del estudio de la enfermedad especificada en la variable llamada *ProblemaInvestigando* (en este caso yo escogí cáncer y realicé el estudio sobre 999 personas).



The screenshot shows a dark blue interface with three buttons at the top: 'Buscar paciente por ID...', 'Crear Simulación', and 'Iniciar I.A.'. Below the buttons, there are four lines of white text on a dark blue background, representing simulation results for cancer cases by age group.

Age Group	Number of Cases
63 años	7
23 años, 36 años	6
14 años, 22 años, 35 años, 80 años	5
5 años, 14 años, 22 años, 35 años, 80 años	4

Como resultado podemos observar que las personas de 63 años son las que tienen más probabilidad de padecer cáncer, seguidas de las de 23 y 36 años. De nuevo aclaro que esto está basado todo en una base de datos aleatoria, no es aplicable a la realidad, para ello debemos cambiar nuestra base de datos por una real. También aclaro que hemos realizado el estudio únicamente sobre la edad de los pacientes con cáncer, faltaría hacerlo sobre todas las características que le añadimos a nuestro fichero de paciente predeterminado.

Ejemplos de fuentes de datos abiertos reales

Como ya hemos visto, se puede crear una simulación de bases de datos, pero como hemos recalcado anteriormente, esta es generada de forma aleatoria, por lo que para poder darle algún uso, habría que reemplazarla por una real. Entonces nos surge la pregunta ¿de dónde sacamos esta información? Pues o de páginas web oficiales o también de las referencias bibliográficas de tesis de otros investigadores. Aquí hay algunos ejemplos:

<https://www.ine.es/>
<https://data.europa.eu/es>
<https://ec.europa.eu/eurostat/>
<https://datos.bancomundial.org/>
<https://datos.madrid.es/portal/site/egob>

<https://pubmed.ncbi.nlm.nih.gov/>
<https://www.cochrane.org/es/evidence>
<https://ourworldindata.org/>
<https://datahub.io/>
<https://www.kaggle.com>

Para pasar algunas de estos datos a nuestra simulación, simplemente habría que descargarlas y rellenar las variables de nuestro código *Paciente* con los datos extraídos de la Open Data. Para ello se descarga en formato csv, tsv, json, o txt y crearíamos un código que lea esta información y la rellene por sí sola nuestro código paciente.

En este caso he realizado este ejemplo en el motor gráfico Unity para ahorrar tiempo en la interfaz de usuario, pero podría hacerse usando otros lenguajes de programación, motores, o incluso sin motor. Un ejemplo de esto sería usar el lenguaje de programación python para leer los archivos descargados en los formatos nombrados anteriormente y seguir el mismo procedimiento explicado en mi simulación en C#.

CONCLUSIONES

Después de todo lo expuesto anteriormente, podemos concluir afirmando que la tecnología está evolucionando hacia el campo de la medicina, el impacto de la inteligencia artificial en la asistencia sanitaria ya es un hecho y su futuro apunta a una medicina que cada vez estará más digitalizada con bases de datos abiertos descomunales, las cuales servirán para realizar diagnósticos e investigaciones. Así, las inteligencias artificiales analizarán estas bases de datos para sacar una serie de porcentajes de posibles enfermedades que podría sufrir el paciente y servirán como un importante punto de partida en la toma de decisiones a la hora de proponer un tratamiento al mismo.

Llegados a este punto, debemos mencionar que las IA deben ser seguras y tener una clara transparencia ante posibles fallos del sistema, aunque estas antes de salir al uso público pasan por diversas comprobaciones para asegurar su correcto funcionamiento. Además, los constructores de estos sistemas (IA) deben hacerlas compatibles con los ideales de dignidad humana, los derechos y las libertades de la humanidad independiente de su cultura, raza, condición económica, etc para poder permitir el acceso a estas herramientas del ámbito de la salud a todas las personas posibles. Además, siempre deberá prevalecer la privacidad personal del paciente, tal y como dicen las características principales de la Open Data. Es una tecnología que debe ser compartida por el mayor número de personas ya que hablamos de prosperidad humana, por tanto, debe ser utilizada con los fines para las que ha sido creada, dándole un buen uso para el bien común.

Para sacarle la máxima eficiencia al uso de IA, se tienen que comprometer las investigaciones en este campo junto con los sectores implicados (hospitales, centros de salud, personal sanitario, seguridad, constructores-programadores, matemáticos, etc.). Un factor clave será subir a la base de datos los últimos avances sobre enfermedades para ampliar y tener lo más actualizada posible las bases de datos abiertos. En este sentido es necesario una buena relación ciencia-política y una buena financiación a la investigación en este campo. Así y tal y como se ha expuesto anteriormente, cuantos más datos y más información actualizada se tenga, más precisos serán los resultados y, por tanto, mejores serán los diagnósticos de las Inteligencias Artificiales que los analicen.

Cuanto más nos acerquemos a ese objetivo, más rápido evolucionará la medicina, ya que se dispondrá de todos los datos que se requieran en una sola base de datos abiertos. Incluso las personas podrían contar en un futuro con avatares digitales capaces de responder a preguntas específicas de los pacientes e incluso hacer diagnósticos rápidos usando programas que comparen sus síntomas con los de dicha base de datos. Actualmente existen los “chatbots” que son programas que participan en conversaciones con humanos, y que se utilizan para la toma de decisiones en línea. Se podría incluso realizar una jerarquía de las enfermedades (priorizando las que puedan ser diagnósticos graves de mayor probabilidad). Así se podría ahorrar en tiempos de espera y agilizar trámites en el caso de pacientes con altos porcentajes de padecer una enfermedad grave (los cuales se priorizarían para ser tratados cuanto antes, aumentando su probabilidad de éxito incluso de supervivencia al empezar un tratamiento a tiempo). Sin duda alguna, la ayuda de este tipo de sistemas podría ser una gran revolución en la medicina asistencial para salvar muchas vidas.

AGRADECIMIENTOS

Me gustaría agradecer a mis compañeros de este trabajo de Iberdrola por su apoyo y colaboración en este proyecto que tanta ilusión nos hacía a todos. De igual manera agradezco a otros participantes que se han retirado del proyecto por motivos varios, que también aportaron ideas.

En especial, gracias a mi profesora Carmen Corada Fernández, y a José Francisco Ávila de Tomás, quienes han ayudado enormemente y este trabajo no sería posible sin ellos.

También menciono a otros profesores del instituto como Antonio Varela y Sergio Villalobos, por sus consejos en el apartado de programación, manejo de datos e inteligencia artificial.

REFERENCIAS BIBLIOGRÁFICAS

Información sobre inteligencias artificiales (Petteri Rouhiainen, 2018):

https://static0planetadelibroscom.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf

Artículo *Big Data en salud: retos y oportunidades* (Menasalvas, et al., 2017):

<https://dialnet.unirioja.es/servlet/articulo?codigo=6207516>

Características de la Open Data:

https://www.programainvestiga.org/pdf/guias2021-22/GuiaIntrodutoriaal%20tema_FuentesdedatosabiertosenSaludanivelmundialyelimpactodesuutilizacion.pdf

Entrevista a José Francisco Ávila de Tomás:

Ávila-Tomás, J.F., Mayer-Pujadas, M.A., Quesada-Varela, V.J. La inteligencia artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA y la robótica. Aten. Primaria. 2020a; 52 (10): 778-784.

Ávila-Tomás, J.F., Mayer-Pujadas, M.A., Quesada-Varela, V.J. La inteligencia artificial y sus aplicaciones en medicina II: importancia actual y aplicaciones prácticas. Aten. Primaria. 2020b; 53 (1): 81-88

Motor Gráfico con el que se realizó la simulación:

<https://unity.com/es>

Programa con el que se programó la simulación:

<https://visualstudio.microsoft.com/es/>