

# TRANSFORMANDO LA ADMINISTRACIÓN TRIBUTARIA EN REPÚBLICA DOMINICANA: INNOVACIÓN Y OPTIMIZACIÓN DEL CUMPLIMIENTO CON UN AGENTE TRIBUTARIO BASADO EN IA.

**CET**  
Centro de Estudios Tributarios  
Universidad de Chile



www.cetuchile.cl

**Guenadie I. Nibbs C.**

Economista, Instituto Tecnológico de Santo Domingo (INTEC).  
Master in Applied Statistics and Machine Learning, Rochester Institute of  
Technology (RIT).

Email: [g.nibbsc@gmail.com](mailto:g.nibbsc@gmail.com)

**Carolina M. Hidalgo D.**

Administración de Empresas, Universidad Iberoamericana (UNIBE).  
Máster en Hacienda Pública y Administración Tributaria, Instituto de  
Estudios Fiscales España, (IEF-UNED).

Email: [Carolina.hidalgo.diaz@gmail.com](mailto:Carolina.hidalgo.diaz@gmail.com)

## **RESUMEN:**

Este artículo aborda los desafíos que enfrentan las administraciones tributarias ante la creciente complejidad de las operaciones de los contribuyentes y la limitada disponibilidad de recursos. Se destaca la necesidad de implementar soluciones innovadoras, como la inteligencia artificial (IA), para mejorar el cumplimiento tributario y aumentar la productividad. Con este objetivo, se propone la creación de un agente virtual basado en IA que guíe a los contribuyentes de niveles específicos de la pirámide de cumplimiento, y los ayude a prevenir incumplimientos involuntarios, optimizando así el uso de los recursos disponibles. Esta propuesta busca tener un impacto positivo en la eficiencia de los recursos y en el logro de las metas de recaudación, presentándose como una alternativa innovadora y de bajo costo.

## ABSTRACT:

This article addresses the challenges faced by tax administrations due to the increasing complexity of taxpayer operations and the limited availability of resources. It highlights the need to implement innovative solutions, such as artificial intelligence (AI), to improve tax compliance and enhance productivity. With this goal in mind, the creation of an AI-based virtual agent is proposed to guide taxpayers from specific levels of the compliance pyramid and help prevent unintentional non-compliance, thereby optimizing the use of available resources. This proposal seeks to have a positive impact on resource efficiency and the achievement of revenue collection goals, presenting itself as an innovative and cost-effective alternative.

## Palabras Claves:

#CumplimientoTributario

#InteligenciaArtificial

#TransformacionDigital

#AgenteVirtual

#OptimizacionCumplimiento

## 1. INTRODUCCIÓN

Durante las últimas dos décadas las administraciones tributarias han ido experimentando un incremento significativo en la complejidad de los procesos asociados a la gestión del cumplimiento. Este hecho ha impactado la capacidad de las administraciones para inducir el cumplimiento voluntario de los contribuyentes a través de los mecanismos tradicionales. Aspectos como:

- i. Las nuevas actividades económicas, que han surgido como producto de la globalización y los avances tecnológicos, han generado importantes brechas legislativas que necesitan ser abordadas por el estado. La velocidad a la que surgen estas nuevas actividades resulta ser muy superior a la velocidad en la que sus brechas legislativas pueden ser detectadas, analizadas y tratadas ante el sometimiento de una propuesta al congreso y su aprobación por las bancadas de representantes. A pesar de los esfuerzos de los organismos internacionales, pocos países han tenido la capacidad de adaptarse completamente a temáticas como la economía digital o las temáticas relacionadas a fiscalidad internacional. Inclusive, la evolución de actividades más tradicionales en términos de su operatividad comercial y financiera ha puesto a prueba la pericia acumulada en cuanto a las formas de control que funcionaron desde inicios del siglo.

- ii. El desarrollo económico de las naciones, que trae consigo un gran incremento en la cantidad de personas físicas y jurídicas que se constituyen para iniciar operaciones comerciales. Si bien es cierto que estos nuevos potenciales contribuyentes deben ser controlados, este incremento no es un factor que el estado utilice como referencia para elevar el presupuesto de las administraciones tributarias. Por razones como estas, se suele producir un desfase entre la cantidad óptima de recursos que se necesitan para garantizar el cumplimiento tributario, y los recursos con los que realmente operan las administraciones.
- iii. A mayor cantidad de contribuyentes, y transacciones que estos realizan, mayor el volumen de datos que se genera en el contexto tributario. Las capacidades sobre recepción, almacenamiento, procesamiento y aseguramiento son un gran reto para las administraciones que requieren de inversiones significativas para lograr una explotación efectiva.
- iv. En el caso de algunas jurisdicciones, el sector informal aún representa retos importantes por abordar. A las administraciones tributarias les cuesta enfrentar esta problemática debido a la relativamente baja tasa de retorno, en términos de recursos necesarios a invertir versus expectativas de recaudo. Una investigación publicada dentro de la red CESifo <sup>1</sup> bajo el nombre “Shedding Light on the Shadow Economy”, muestra un top de las 15 economías con los sectores informales más grandes en función del PIB, de las cuales 6 corresponden a países de América Latina; incluyendo la República Dominicana.

Con estos desafíos, que cada día se hacen más imperantes, las administraciones tributarias están llamadas a buscar alternativas innovadoras que permitan alcanzar los objetivos de cumplimiento establecidos ante las evidentes limitaciones de recursos.

Una de las temáticas que ha captado mayor interés en los últimos años corresponde a los avances obtenidos en el campo de la inteligencia artificial. Las herramientas desarrolladas utilizando este conjunto de nuevas técnicas y capacidades metodológicas, permiten sacar provecho del gran volumen de información que manejan las administraciones, y con esto implementar soluciones efectivas de cara a la automatización de procesos, mejoramiento de la productividad de los funcionarios y la facilitación en la toma de decisiones de los niveles directivos.

En esta ocasión se plantea la posibilidad de utilizar una herramienta construida con inteligencia artificial, que resuelva la problemática de los recursos limitados sobre una segmentación particular de contribuyentes, que por su nivel de disposición hacia el cumplimiento, no necesiten de estos recursos para materializar sus

---

1 Center for Economic Studies.

obligaciones fiscales. Esto abre la posibilidad de que se puedan utilizar estos recursos liberados sobre contribuyentes que realmente lo requieran.

## 2. MODELOS DE GESTIÓN DE CUMPLIMIENTO TRIBUTARIO

Cuando se habla de modelos de gestión de cumplimiento tributario se hace referencia a todos los procesos, políticas, procedimientos y herramientas que la administración utiliza para que los contribuyentes logren cumplir sus obligaciones en tiempo y forma. Si bien pueden existir diferencias significativas en la manera en que cada administración decide (o puede)<sup>2</sup> implementar su modelo de gestión<sup>3</sup>, existen fundamentos que explican los distintos comportamientos que pueden observarse de los contribuyentes y a que factores atienden dichos comportamientos<sup>4</sup>.

### 2.1 Disposición hacia el Cumplimiento del Contribuyente

Los modelos de gestión tributaria se enfrentan a 4 niveles de disposición hacia el cumplimiento de los contribuyentes expresados por la “Pirámide del Cumplimiento Tributario” (Braithwaite, 2009). Esta pirámide expresa 3 aspectos fundamentales en cualquier modelo de cumplimiento:

- La estrategia de la administración tributaria, que intenta impulsar a los contribuyentes que están en la parte superior de la pirámide hacia la base.
- La proporción de contribuyentes que se encuentran en cada uno de los niveles de disposición con relación al sistema tributario. En casos muy particulares, podrán existir sistemas tributarios que exhiban una pirámide invertida donde cambia la proporcionalidad de cada uno de los grupos.
- La severidad de las acciones de tratamiento que son aplicadas a los contribuyentes en cada uno de estos niveles; a mayor altura en la posición de la pirámide, menor proporción de contribuyentes y mayor severidad de los tratamientos a prescribir.

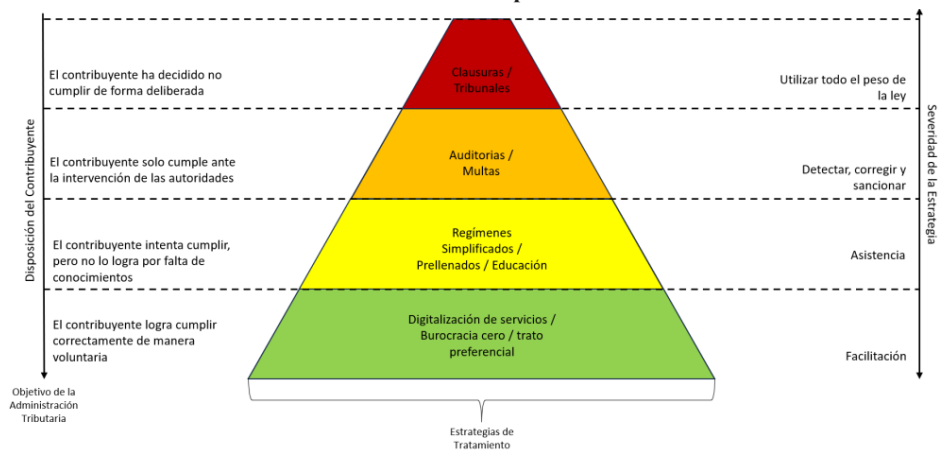
---

2 Atendiendo a las competencias, capacidades instaladas e idiosincrasia de cada país.

3 Implementaciones con enfoque tecnológico, por funciones, por procesos, con componentes de riesgo, híbridas, etc.

4 Factores sociales, económicos, psicológicos y de negocio.

**Gráfico 1. Pirámide del Cumplimiento Tributario**



Fuente: Elaboración propia basada en el Manual de Gestión de Cumplimiento del CIAT y el documento sobre desarrollo de administraciones tributarias de la OECD

Llegado a este punto es de vital importancia entender cuáles son las características más significativas en cada uno de estos niveles.

- i. En la primera posición, representada por la base de la pirámide, se encuentran aquellos contribuyentes que cumplen de forma correcta e intentan no tener fricciones con las autoridades, ya sea porque desean mantener una imagen empresarial positiva y/o entienden que el incumplimiento resulta ser más costoso que cualquier posible beneficio obtenido a través de acciones no bien intencionadas. La administración debe hacer todo lo posible para reducir el costo de cumplimiento de estos contribuyentes, instalando capacidades que permitan un nivel óptimo de fluidez en los procesos. Con la correcta capacidad instalada, quienes pertenezcan a esta segmentación, deberían ser los que menor cantidad de recursos limitados consuman de parte de la administración.
- ii. En la segunda posición, se encuentran aquellos con una alta disposición al cumplimiento, pero sin el conocimiento suficiente para lograr cumplir sus obligaciones en la forma correcta. Estos contribuyentes tienden a percibir el costo del cumplimiento como muy alto, ya que no se sienten en la capacidad de realizar ninguno de sus procesos tributarios sin la intervención de un tercero especializado<sup>5</sup>. En este apartado la administración debe proveer la asistencia necesaria para que las obligaciones tributarias no sean un tema de preocupación. Esto se logra proveyendo materiales educativos en diferentes formatos y a través de diversos canales para aquellos que estén dispuestos a intentarlo. De igual

<sup>5</sup> Un contador o equipo tributarista, según el tamaño de la empresa.

- forma, que aquellos que solo quieran pensar en las operaciones de su negocio, tengan la posibilidad de ingresar a modalidades de tributación simplificadas con propuestas prellenadas y sin burocracias. Dada la gran cantidad de contribuyentes que existen en este nivel, se termina produciendo un volumen significativo de incidencias en sus procesos que consumen una buena parte de los recursos limitados de la administración.
- iii. En la tercera posición, están aquellos que necesitan de alguna acción de control para mantener sus responsabilidades tributarias al día, y por ende requieren que la administración logre mantener una percepción de riesgo alta a través del monitoreo constante de sus actividades. Este nivel de contribuyentes, aunque de menor proporción que los anteriores, requiere un alto consumo de recursos, ya que los tratamientos que generalmente se aplican en esta segmentación terminan siendo los de mayor costo.
  - iv. En la cuarta posición, representada por el tope de la pirámide, se encuentran aquellos con un comportamiento agresivo. Estos contribuyentes no comparecen ante la administración y utilizan a su favor artilugios tributarios para financiar sus operaciones con los impuestos del estado; dilatando así las sanciones finales sobre sus incumplimientos. Además, capitalizan las brechas existentes en la legislación para sacar provechos fiscales considerados como cuestionables. Debido a que estos casos en su mayoría terminan resolviéndose mediante la comparecencia ante los tribunales, se requiere de un tiempo significativamente mayor para que sean resueltos, y por ende consumen una alta cantidad de recursos de la administración. En casos extremos donde la severidad del incumplimiento se categoriza como fraude fiscal, muchas legislaciones consideran el hecho como un delito de tipo penal.

A medida que la administración se mueve hacia el tope de la pirámide, se va reduciendo la diversidad y disponibilidad de los tratamientos eficaces que pueden ser utilizados para abordar la acción de incumplimiento. De igual forma, estos tratamientos van incrementando el costo monetario que significan para la administración y el tiempo en que pueden ser ejecutados. Dado que la mayoría de las administraciones basan sus metas en indicadores puros de recaudo, frecuentemente terminan enfocándose en los niveles 2 y 3. Con esta estrategia intentan lograr un punto óptimo de control entre los contribuyentes que realmente requieren este tipo de acciones y los que garantizan un flujo constante de ingresos a las arcas del estado.

La problemática con este enfoque radica en que, en estos niveles de la pirámide, la proporción de contribuyentes es tan extensa, que terminan mal utilizándose los recursos en contribuyentes que pudieron haberse tratado con igual efectividad, utilizando acciones de mayor disponibilidad.

Cuando las autoridades logran entender que el recaudo es una consecuencia de la gestión del cumplimiento, y por ende deciden implementar una solución integral que aborde las situaciones que más afectan al sistema tributario<sup>6</sup>, terminan enfrentándose a 3 posibles alternativas para lograr una mejor locación de los recursos.

1. Incrementar la disponibilidad de los recursos en procesos asociados a las estrategias de cumplimiento. Esto implica que si se determina la necesidad de un incremento del 50% en el headcount de auditores para alcanzar el óptimo de cobertura<sup>7</sup> en los niveles 3 y 4, la administración deberá proveerlos. En la práctica, esta alternativa resulta ser la menos probable, ya que las autoridades tributarias, como instituciones del estado, cuentan con un presupuesto limitado que imposibilita mantener a la par la cantidad de recursos disponibles con los recursos necesarios; los cuales se incrementan constantemente a medida que nuevos contribuyentes aparecen en el entorno económico.
2. Eficientizar el uso de los recursos en aquellos procesos asociados a las estrategias de cumplimiento. Asumiendo que existen espacios para realizar mejoras que resulten en la reducción de los tiempos de ejecución o en la automatización de algunos de estos procesos, se puede lograr despachar una mayor cantidad de acciones de tratamiento sin elevar la cantidad de recursos. Esta alternativa, aunque de gran importancia, puede llegar a estancarse rápidamente, resultando en un impacto regular o deficiente en los índices de cobertura de las estrategias de control.
3. Liberar los recursos subutilizados sobre contribuyentes que no lo ameriten. Esta estrategia busca relocalizar los tratamientos de menor disponibilidad que se aplican en los niveles 1 y 2 de la pirámide hacia los niveles 3 y 4. Las brechas y riesgos que se producen en estos niveles responden en su mayoría a desconocimiento del contribuyente, errores de terceros, falta de datos, calidad de la información y aspectos especializados del negocio. Por estas razones, este grupo de contribuyentes responde bastante bien a tratamientos con poca intervención de la administración tributaria.

## 2.2 Contribuyentes que quieren cumplir

La mejor inversión que puede hacer una administración tributaria en cuestiones de cumplimiento es en aquellos recursos que buscan impedir la materialización del hecho. Luego de materializado, la mayoría de las estrategias asociadas con la corrección del incumplimiento, poseen un costo significativamente alto

---

<sup>6</sup> Un modelo integral de cumplimiento permite a las administraciones identificar en cuál de estos niveles se encuentra cada uno de sus contribuyentes y tratar acorde a su nivel de disposición

<sup>7</sup> Una cantidad de auditores que permitan que los contribuyentes que califican para auditorías de campo puedan ser evaluados antes de la prescripción de los periodos.

para la administración. Sin embargo, pensar que existe un monto de inversión que garantice la prevención de todos los incumplimientos, se aleja bastante de la realidad. Por razones como esta, la administración está llamada a buscar soluciones innovadoras que aborden las causas que explican el comportamiento de los contribuyentes, y así reducirlas a incidencias asociadas únicamente a externalidades del contribuyente y acciones intencionales.

En el caso de República Dominicana, cuando se evalúa el comportamiento de los contribuyentes que se encuentran en los niveles de mayor disposición hacia el cumplimiento, se observa que el 37% de estos cometen el mismo error al menos dos veces y de manera involuntaria. Esta situación implica que la administración tributaria termina asumiendo el costo de incumplimiento de estos contribuyentes, ya que se generan casos de revisión que deben ser evaluados manualmente por los oficiales disponibles en cada oficina municipal.

La estrategia al abordar esta situación desde la perspectiva tradicional ha sido intentar optimizar los procesos de notificación temprana, ya que si estos contribuyentes son notificados de forma tardía y resulta que comparecen ante la administración de forma voluntaria, terminan siendo afectados por penalidades en forma de recargos e intereses; lo que desincentiva las metas de cumplimiento y la formalidad en sí misma.

Cuando la autoridad tributaria decide abordar los incumplimientos sin diferenciar entre los niveles de disposición de los contribuyentes, se generan ineficiencias en la utilización de los recursos que afectan a los niveles con menor propensión de cumplimiento. Por esta razón, siempre es recomendable reservar aquellos tratamientos de menor disponibilidad para los niveles 3 y 4 de la pirámide, donde son considerados como recursos indispensables.

Para abordar esta problemática se procede con la creación de un agente virtual, basado en inteligencia artificial, que guíe al contribuyente en tareas como el llenado y envío de las declaraciones juradas, las consultas sobre las formas correctas de declarar ciertas transacciones específicas de cada sector, y en la validación de sus documentaciones previo al envío formal ante la administración. Con la implementación de este agente se logrará que los contribuyentes que intentan cumplir puedan hacerlo con un costo de cumplimiento relativamente bajo. Esto iniciaría la liberación de los recursos subutilizados en los niveles 1 y 2, y la posibilidad de ampliar la cobertura en aquellos casos donde la intervención de la administración sea estrictamente necesaria.

### **3. INTELIGENCIA ARTIFICIAL**

Cuando hablamos de inteligencia artificial nos referimos a la capacidad que adquiere un sistema computacional para simular la inteligencia humana. Para



entender esta terminología es necesario visualizarla como un fin, mas no como un medio. Hoy en día existen importantes avances en campos de estudio como el machine learning, la computación, las matemáticas y las neurociencias que han contribuido a que se pueda alcanzar lo que hoy se conoce como inteligencia artificial. A pesar de los avances obtenidos en el desarrollo de esta nueva tecnología, muchos consideran que, por definición, la verdadera inteligencia artificial aún no ha sido alcanzada, sino que contamos con metodologías avanzadas de machine learning.

El término “machine learning” o aprendizaje de máquinas hace referencia a la capacidad que tiene un programa para tomar decisiones sin que estas hayan sido programadas explícitamente. En esta rama de estudios se desarrollan algoritmos estadísticos que pueden aprender de los datos disponibles, y generalizar los fenómenos para los cuales han sido entrenados, en abstracciones conocidas como modelos. Según un artículo publicado por la universidad de Berkeley en 2017<sup>8</sup>, el proceso de aprendizaje de estos algoritmos puede ser explicado por 3 componentes.

1. Un proceso de decisión donde el algoritmo produce conjeturas sobre los posibles patrones identificados en los datos. Estas estimaciones son el resultado de una secuencia de cálculos que varían según el tipo de algoritmo.
2. Una función de error que intenta medir qué tan bueno es el modelo que resulta de las conjeturas generadas por el algoritmo. Esta función se alimenta de la retroalimentación que obtiene el modelo con datos distintos a los que fueron utilizados para su entrenamiento.
3. La optimización de parámetros que se realiza evaluando las decisiones correctas e incorrectas tomadas por el modelo y actualizando sus estimaciones a través de mecanismos de retroalimentación.

Dentro de este campo de estudio tan extenso se pueden encontrar distintas ramas de especialización, campos de aplicación y categorías de algoritmos. Por ejemplo, los algoritmos se pueden clasificar en<sup>9</sup>:

- Supervisados: son aquellos algoritmos diseñados para ser aplicados sobre set de datos que se encuentran pre-etiquetados<sup>10</sup>. En estos casos, los datos pueden establecer claramente la relación que hay entre los distintos comportamientos y sus resultados. Estos modelos suelen estar

---

8 Data Tools Deep Dive: Machine Learning.

9 Es importante establecer que también pueden existir casos especiales donde se utilicen metodologías híbridas construidas con componentes de estas tres grandes categorías.

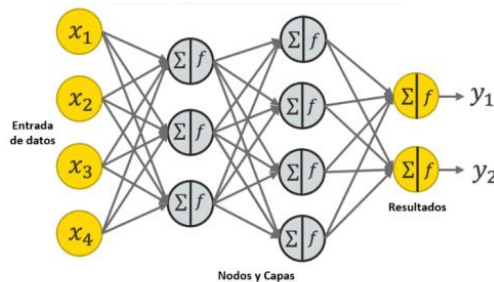
10 Los datos muestran claramente la relación entre comportamiento (variables independientes) y su resultado (variable dependiente).

asociados a resultados de mayor precisión y menores requerimientos computacionales dada la naturaleza de sus datos. Ej: regresión, árboles de decisión, redes neuronales<sup>11</sup>, entre otros.

- No-Supervisados: son aquellos algoritmos diseñados para ser aplicados sobre set de datos que no se encuentran pre-etiquetados. Aquí no se logra establecer una relación de los datos con algún resultado en particular, y por ende, suelen ser menos precisos y requieren de mayor capacidad computacional. Ej: clusters, redes neuronales, reducción de dimensionalidad, entre otros.
- Aprendizaje Reforzado: los algoritmos de esta categoría no se consideran como dato-dependientes, sino que utilizan una estrategia de prueba y error, mediante un esquema de retroalimentación constante, hasta llegar a obtener los parámetros estimados que generan el comportamiento esperado del modelo.

Con estos algoritmos se desarrollan áreas de especialización como el “deep learning” o aprendizaje profundo, donde se utilizan redes neuronales de alta complejidad que intentan simular la forma de procesamiento del cerebro humano<sup>12</sup>, ampliando las capas de nodos y enlaces que existen entre los datos de entrada y salida. Estos avances son usados en aplicaciones como “computer vision” para el procesamiento de imágenes y videos, o el área de procesamiento del lenguaje natural.

**Gráfico2. Representación de una Red Neuronal**



Fuente: Kowch (2022). The Basics of Neural Networks. <https://towardsdatascience.com/>

### 3.1 Tipos de Inteligencia Artificial

Dentro del campo de la inteligencia artificial se pueden encontrar dos grandes

<sup>11</sup> Ver Gráfico 2.

<sup>12</sup> Las redes neuronales tienen la capacidad de detectar patrones de mayor complejidad en los datos con una menor intervención humana. Esto a expensas de un mayor requerimiento computacional.

grupos de categorías<sup>13</sup>:

Las categorías del grupo I, enfocadas en la distinción de las capacidades que posee el sistema de inteligencia artificial. En esta clasificación podemos encontrar 3 tipos que son:

- **Inteligencia Artificial Estrecha:** también conocida como débil, que a pesar de tener capacidades superiores a la velocidad de la mente humana, solo puede operar dentro del alcance para el cual ha sido predefinida. Se dice que dentro de los avances que ha tenido la tecnología en los últimos años, este es el único nivel de inteligencia que ha podido ser alcanzado.
- **Inteligencia Artificial General:** también conocida como fuerte, y considerada su existencia sólo en el aspecto teórico. Esta inteligencia contará con la capacidad de utilizar aprendizajes previos para realizar nuevas tareas sin la necesidad de la intervención de la mano humana.
- **Inteligencia Artificial Súper:** Las habilidades de esta inteligencia incluyen la capacidad de emitir juicios de valor, hasta el hecho de poseer sentimientos y emociones propias de los seres humanos.

Las categorías del grupo II, enfocadas en la distinción de las funcionalidades que se pueden alcanzar utilizando los sistemas definidos en el grupo I, como son:

- **Máquina Reactiva:** asociada con la inteligencia artificial débil, son aquellos sistemas que no poseen la capacidad de crear memoria y solo utilizan la data con la que fueron entrenados para desarrollar tareas específicas.
- **Memoria Limitada:** a diferencia de máquina reactiva, estos sistemas tienen la capacidad de aprender sobre nuevos eventos que se producen luego de su entrenamiento. Sin embargo, estas memorias no son almacenadas de forma permanente por lo que se van perdiendo con el tiempo.
- **Teoría Mental:** asociada con la inteligencia fuerte, que podrá entender las emociones humanas, abriendo la posibilidad de simular relaciones de la máquina con otras entidades.
- **Inteligencia Consciente:** asociada con la inteligencia super, que podrá evaluarse por sí misma (reflexión) y desarrollar sus propias creencias, sesgos y emociones.

### 3.2 Inteligencia Generativa y Modelos Grandes de Lenguaje

El desarrollo de un sistema con inteligencia artificial presenta múltiples retos de implementación, siendo uno de los principales la forma de entendimiento entre

---

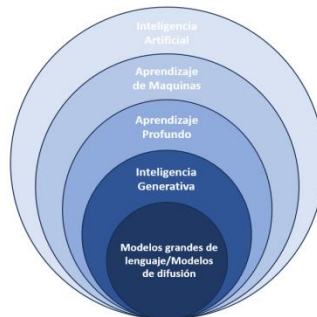
<sup>13</sup> El desarrollo del agente virtual se catalogaría como una inteligencia “débil” con “memoria limitada”.

la máquina y el usuario. Hasta el año 2017 los avances obtenidos en las áreas de machine learning y lingüística permitieron alcanzar hitos importantes en las metodologías asociadas al procesamiento del lenguaje natural. Esto permitió crear productos que contaban con la capacidad de traducir textos, evaluar la intención y los sentimientos de las palabras según su contexto, e interactuar con el usuario asociando estas intenciones a respuestas preseleccionadas que formaban parte de un algoritmo de conversación<sup>14</sup>.

A partir del 2017 se consolidaron logros importantes en el campo de las redes neuronales que se consideran como la base del descubrimiento de la inteligencia artificial generativa. Entre los más determinantes se encuentran los:

- Autoencoders: es un tipo de algoritmo que permite a la red neuronal codificar y decodificar un set de datos sin la necesidad de una variable de respuesta. Estas codificaciones son una representación de baja dimensionalidad a partir de la cual se puede reconstruir el resultado original.
- GAN<sup>15</sup>: es un tipo de arquitectura que permite la generación de data original a través de la simulación de una competición de suma cero entre dos redes neuronales; donde se entrenan entre sí, y las ganancias de una se convierten en pérdidas para la otra.
- Transformers<sup>16</sup>: es un tipo de arquitectura utilizada principalmente en el procesamiento del lenguaje natural donde el modelo aprende a través del rastreo de las relaciones que existen en secuencias de palabras, como es el caso de las oraciones de un texto<sup>17</sup>.

**Gráfico 3. Aplicación de la Inteligencia Artificial Generativa**



Fuente: Elaboración propia

14 Esta es la tecnología detrás de los chatbots tradicionales de la época.

15 Generative Adversarial Network (GAN), donde una de las redes genera nuevos datos hasta que la otra no es capaz de diferenciar entre un contenido creado y un contenido original.

16 Red neuronal creada por Google en 2017 mediante el paper “Attention Is All You Need”.

17 Surge como alternativa a otras metodologías como las Convolutional Neural Network (CNN) y las Recurrent Neural Network (RNN).

De aquí empieza a tomar auge el término de inteligencia generativa, que no es más que el conjunto de modelos que cuentan con la capacidad de crear contenido original a partir de los datos con que fueron entrenados. Esta inteligencia puede extenderse a aplicaciones sobre formatos de texto, imágenes, audio y video. De todas las aplicaciones de inteligencia generativa que utilizan texto como su fuente primaria de datos, la que mejor impacto ha concitado en la opinión pública se conoce como los grandes modelos de lenguaje .

Los Grandes Modelos de Lenguaje o “LLMs” por sus siglas en inglés, son modelos que adquieren la habilidad de utilizar lenguaje humano de propósito general a través del entrenamiento con altos volúmenes de información en formato de texto. A pesar de lograr estas habilidades semánticas, también cuentan con la posibilidad de ejecutar tareas a gran velocidad y responder con exactitud, tomando en cuenta elementos como el contexto, el idioma, la estructura de redacción y posibles errores cometidos por el usuario al momento de interactuar.

### **3.3 Requerimientos Técnicos en la Implementación de un LLM**

Para desarrollar LLMs que garanticen una fluidez en las interacciones con el usuario se deben abordar algunos retos relacionados con el procesamiento de la información.

1. Adquisición de los datos: es necesario contar con un volumen de información que permita entrenar al modelo en entender y articular mensajes de forma coherente, para responder a las preguntas que realizan los usuarios. Esta cantidad de datos también implica que se debe acceder a múltiples fuentes para extraer la información; donde muchas de estas no se encuentran en un formato estructurado.
2. Procesamiento de los datos: las fases de extracción, limpieza, transformación y modelamiento de esta cantidad de información requiere de una infraestructura de alta capacidad computacional en términos de CPUs, GPUs, TPUs o DPUs<sup>18</sup>. De igual forma, aunque se mantenga la cantidad de información constante, la selección del tipo de algoritmo y la cantidad de parámetros que contiene impactará de forma significativa los requerimientos de los recursos computacionales. En este caso no se atribuyen las mismas ponderaciones al factor almacenamiento, ya que los costos asociados a este tipo de producto o servicio se han reducido considerablemente con el pasar de los años.
3. Tiempo de implementación: el desarrollo de cualquier tipo de modelo estadístico conlleva un proceso de experimentación donde se configuran

---

<sup>18</sup> Central Processing Unit (CPU), Graphics Processing Unit (GPU), Tensor Processing Units (TPU), Data Processing Units (DPU). Ya sea la selección de una de estas unidades de procesamiento o la combinación de varias.

diferentes versiones y se realiza una optimización de sus parámetros. Cada prueba de estos modelos puede tomar días, incluso utilizando los recursos de procesamiento disponibles en la nube. Este proceso representa un costo monetario significativo y no garantiza que al final se obtenga una versión que muestre el comportamiento esperado.

Debido al gran costo que implica entrenar este tipo de inteligencia artificial, se han generado estrategias que buscan democratizar el uso de estas herramientas. Entre las alternativas más comunes se encuentra la utilización de modelos grandes de lenguaje pre-entrenados con distintas cantidades de parámetros. Actualmente existen compañías que ponen a disposición sus modelos de forma comercial o utilizando licencias de código abierto y libre para que puedan ser utilizados por las comunidades de investigación. Este esquema aligera la carga de los desarrolladores que no pueden afrontar la implementación de un proyecto de esta magnitud, ya sea porque no cuentan con el volumen de datos para construir un modelo medianamente funcional, o no poseen las capacidades de procesamiento requeridas para entrenarlo.

### **3.4 Modelo grande de Lenguaje Meta AI (Llama)**

Uno de los principales actores en el escenario de los modelos pre-entrenados con licenciamiento de código abierto y libre es Llama, que hasta la fecha posee la comunidad de desarrollo más extensa de todas y por ende representa una de las alternativas más robustas y confiables para la construcción del agente virtual tributario.

Los modelos grandes de lenguaje de Meta AI o “Llama” por sus siglas en inglés, son un conjunto de modelos desarrollados por Meta<sup>19</sup> desde el año 2023, y liberados ante las comunidades de inteligencia artificial, con apertura para ser utilizados con fines de investigación y comerciales<sup>20</sup>. Hasta la fecha, se han liberado 3 grandes versiones: Llama 1 y 2 con disponibilidad de hasta 4 modelos cada uno, y Llama 3 liberado en el 2024 con disponibilidad de 2 modelos<sup>21</sup>.

El proceso de construcción de estos LLMs comienza con la fase de entrenamiento, donde se organiza la fuente de datos (o text-corpus) y se prepara el modelo para realizar predicciones de palabras dentro de una secuencia de texto. Este proceso es el responsable de que el modelo logre alcanzar la capacidad semántica requerida. Para que este entrenamiento se realice de una forma eficiente y arroje

---

19 Antiguamente Facebook.

20 Si una compañía excede los 700 millones de usuarios mensuales, deberá solicitar un permiso especial a Meta para poder utilizar el modelo.

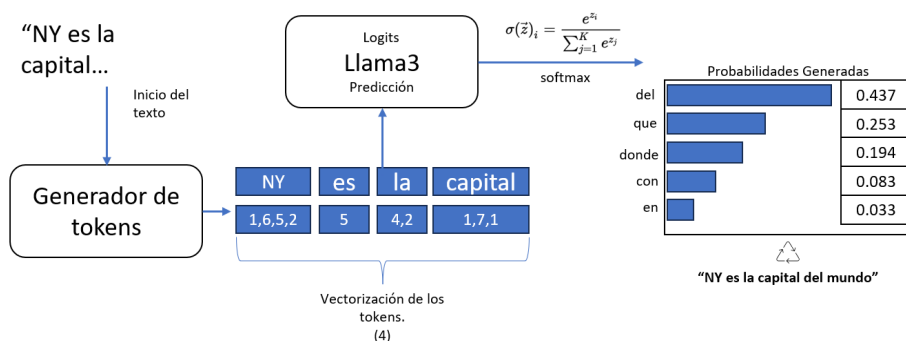
21 Meta anunció la introducción de variantes adicionales a partir de la fecha de redacción de este artículo.

resultados precisos, se utiliza una red neuronal no supervisada<sup>22</sup> construida bajo la arquitectura conocida como Transformers. Este tipo de arquitectura o red neuronal se basa en dos mecanismos de procesamiento conocidos como “auto-atención” y “atención multi-cabezal”. El primer mecanismo se encarga de que el modelo no realice predicciones directas de las secuencias del texto<sup>23</sup>, sino que se puedan establecer las relaciones entre todas las palabras ingresadas y con ello generar un contexto de influencias entre ellas. El segundo mecanismo permite que cada “cabezal” de la red neuronal aborde diferentes aspectos del texto de forma simultánea, como las relaciones sintácticas y semánticas. A medida que va aprendiendo, la red neuronal va ejecutando un proceso de optimización donde se modifican los pesos de sus parámetros buscando reducir el error total del modelo; esto se logra a través de un algoritmo conocido como “backpropagation”.

Una vez entrenado, el modelo puede generar texto utilizando distintas estrategias de procesamiento donde, según lo que sea ingresado por el usuario, se toman las posibles secuencias de palabras para elegir la mejor entre ellas. Estas estrategias de selección pueden utilizar muestreos aleatorios basados en distribuciones probabilísticas entre distintas secuencias, o simplemente utilizar la mayor probabilidad de predicción arrojada por el modelo.

Las redes neuronales construidas con este tipo de arquitectura no poseen la capacidad de procesar directamente la información en la forma en que es consumida por los usuarios. Antes que los datos ingresen en la capa de procesamiento de la red, ya sea en la etapa de construcción del modelo o en la de su utilización, el texto debe ser convertido en unidades mínimas conocidas como “tokens”. Estas unidades son la representación de una palabra o partes específicas de una palabra, dependiendo de la metodología de tokenización que sea utilizada.

**Gráfico 4. Proceso de Generación de Textos de los LLMs**



Fuente: Elaboración propia

22 Esta red neuronal posee la capacidad de distinguir los patrones y estructuras en los textos sin la necesidad de que la data se encuentre preclasificada.

23 Colocando la palabra de mayor peso probabilístico en la secuencia de texto.

Una vez termina el proceso de tokenización cada token es convertido en una representación numérica que se aloja en un vector donde el algoritmo puede realizar sus cálculos<sup>24</sup>. Estos vectores empiezan a capturar las propiedades de las palabras, por lo que aquellas palabras que se utilizan en contextos similares tienden a tener vectores muy parecidos. Cuando los vectores entran a la capa de procesamiento, se utilizan las relaciones y contextos identificados para producir los resultados esperados.

El resultado arrojado por la capa de procesamiento contiene un vector de scores que pasa a ser transformado en una distribución probabilística a través de una función conocida como “softmax”. Este resultado permite al modelo elegir la palabra o secuencia que mejor se ajuste al caso de uso.

Si bien la capacidad de estos modelos depende significativamente de la cantidad de datos y su calidad, de igual forma la arquitectura utilizada y el ajuste de los parámetros logrados juegan un papel determinante. Generalmente las fichas técnicas que hacen referencia a las capacidades de estos modelos son presentadas con cuatro categorías fundamentales:

- I. El tamaño del modelo: representado por la cantidad de parámetros utilizados durante el entrenamiento de la red neuronal.
- II. La ventana de contexto: que corresponde a la cantidad de tokens que pueden ser ingresados al modelo al momento de generar respuestas.
- III. El volumen de datos: que fueron utilizados para el proceso de entrenamiento del modelo desde distintas fuentes de información: web, artículos, repositorios, libros, entre otros.
- IV. El vocabulario: que representa la cantidad de palabras únicas que el modelo tiene la capacidad de reconocer y utilizar en su procesamiento.

**Tabla 1. Capacidades Técnicas de los Modelos Llama**

Modelo	Tamaño (en parámetros)	Contexto (en tokens)	Volumen Entrenamiento (en tokens)	Vocabulario (en tokens)
Llama 1	7B	2k	1.0T	32k
	13B			
	33B		1.4T	
	65B			
Llama 2	7B	4k	2.0T	
	13B			
	34B			
	70B			
Llama 3	8B	8k	15.0T	
	70B			128k

*Fuente: Elaboración propia basada en informaciones del "Model Card" del repositorio de Llama en Github.*

<sup>24</sup> El proceso de conversión de los tokens a vectores se conoce como incrustación o “embedding” en inglés.



Si bien es cierto que los modelos con mayor cantidad de parámetros ofrecen una mejor precisión en el resultado de los textos, también es cierto que aumentan los requerimientos de la infraestructura necesaria para lograr su implementación.

Cada variación en la cantidad de parámetros representa una versión distinta de los modelos Llama, por lo que se pudiera decir que a la fecha se han puesto a disposición 10 modelos grandes de lenguaje pertenecientes a esta familia. Una de las características de mayor interés sobre estos modelos es la capacidad que poseen de ser reentrenados sobre dominios específicos, proceso que se conoce como “fine-tuning”. Durante esta fase los parámetros del modelo pre-entrenado son modificados sobre la base de nueva información, lo que permite que se pueda tener un mejor rendimiento en la ejecución de las habilidades deseadas.

Existen investigadores y empresas que se dedican a realizar mejoras de estas versiones base de los modelos, utilizando set de datos de recolección y metodologías de procesamiento propias. Estas versiones buscan proveer alternativas más acabadas en temáticas asociadas a aspectos comunicacionales, programación de software, diagnóstico médico, y cualquier otra temática que pueda ser de interés para la comunidad de usuarios o que pueda presentar alguna viabilidad comercial.

Un caso especial de estas capacidades de fine-tuning es la conversión de un gran modelo de lenguaje a un modelo multimodal. Es importante recordar que la característica principal de los modelos grandes de lenguaje consiste en recibir texto como datos de entrada y generar nuevos textos a partir del procesamiento de estos datos. En cambio, un modelo multimodal permite que también se puedan reconocer y procesar otros formatos como imágenes, videos y audio. Casos como el de GPT-4 accesible a través de la interfaz de ChatGPT es un claro ejemplo de un modelo multimodal con capacidad para procesar distintos formatos.

#### 4. TAXER – AGENTE VIRTUAL DE CUMPLIMIENTO TRIBUTARIO

El diseño de la estrategia de implementación de un agente virtual tributario está muy relacionado a las funcionalidades que se quieran incorporar en la herramienta. En este caso, Taxer el agente virtual de la DGII, cuenta con 2 capacidades fundamentales que pueden resumirse en los conceptos de consultas tributarias y gestión de declaraciones juradas. Para lograr esto el agente debe contar con una alta capacidad de desenvolvimiento que le permita responder preguntas y aclarar dudas a través

Gráfico 5. Taxer, Agente Virtual Tributario DGII



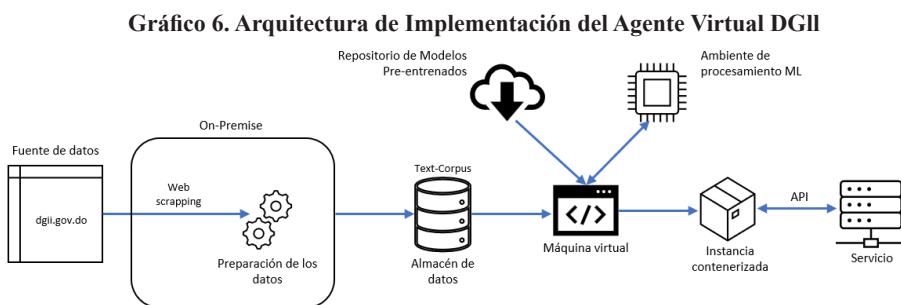
Fuente: Elaboración propia utilizando Dall-E.

de la simplificación de conceptos y mecanismos complejos. Adicionalmente, se requiere que se pueda realizar clasificación de transacciones y cálculos matemáticos utilizando las informaciones provistas por el contribuyente.

Tomando en cuenta estas consideraciones, el plan de ejecución consta de los siguientes pasos:

- 1- Selección y preparación de un ambiente de desarrollo flexible y escalable que pueda cubrir las necesidades del proyecto sin incrementar sus costos de forma significativa.
- 2- Selección de uno de los modelos pre-entrenados Llama, encargado de manejar la parte semántica del agente.
- 3- Adquisición y preparación de los datos destinados a incorporar las capacidades tributarias en el agente.
- 4- Reentrenamiento de los parámetros del modelo utilizando los datos tributarios.
- 5- Ejecución de pruebas de calidad y ética de las respuestas y posibles ajustes.
- 6- Selección de la plataforma de despliegue y pase a producción del agente.

Este plan se ejecuta bajo la siguiente arquitectura:



Fuente: Elaboración propia

- I. Creación y configuración de una máquina virtual para desarrollo en la plataforma de Azure.
- I. Extracción de los datos tributarios desde el portal web [www.dgii.gov.do](http://www.dgii.gov.do) utilizando la librería “rvest” a través del lenguaje de programación R.
- II. Preparación del text-corpus con los archivos de datos y colocación en el

- Blob Storage<sup>25</sup> de Azure utilizando el entorno de librerías de “tidyverse” y la librería “azurestore” respectivamente.
- III. Descarga del modelo pre-entrenado desde la plataforma de Hugging Face hacia el ambiente de desarrollo utilizando la librería “transformers” a través del lenguaje de programación Python.
  - IV. Configuración de los parámetros de ejecución del modelo y procesamiento del fine-tuning utilizando Azure ML y Python.
  - V. Creación de una instancia contenerizada con el modelo ajustado en Azure Containers y Docker.
  - VI. Conexión de la instancia contenerizada con el canal de acceso al usuario.

Esta arquitectura es implementada bajo uno de los modelos de precios de Azure conocidos como spot-ahb, el cual permite aprovechar las capacidades no utilizadas de los servicios ofrecidos por la plataforma. Azure realiza proyecciones de uso de sus clientes estándar dentro del modelo de precios “pay-as-you-go”<sup>26</sup>, lo que permite determinar aquellos momentos donde la demanda de estos recursos será relativamente baja y revenderlos a un costo más atractivo. De igual forma, este modelo también permite reutilizar licencias que ya estén disponibles on-premises para abaratar los costos. Es claro que estos beneficios vienen a expensas de que las capacidades revendidas puedan ser retiradas con un tiempo mínimo de notificación, lo que pudiera afectar el funcionamiento de productos desplegados a gran escala. En este caso, para la prueba de concepto de Taxer, es un trade-off bastante razonable.

#### 4.1 Construcción de la Fuente de datos (Text-Corpus)

Las fuentes de datos utilizadas para incorporar las capacidades tributarias al agente virtual, provienen en un 100% de las informaciones públicas colocadas en el portal web de la DGII. Los enlaces del portal seleccionados hacen referencia a los sitios que poseen contenido de mayor valor para los fines planteados. Las secciones identificadas como de mayor relevancia corresponden a la biblioteca virtual, el foro de contribuyentes, el repositorio de los formularios de declaraciones juradas y la sección de normativas.

#### Biblioteca Virtual

Parte importante de la estrategia de facilitación y asistencia de la DGII consiste en proveer a los contribuyentes con publicaciones de apoyo como instructivos, guías y revistas informativas acerca de todos los procesos e impuestos disponibles en la administración. A diferencia de las leyes, normativas y reglamentos, la información contenida en estos materiales es desarrollada bajo un lenguaje claro y entendible, que cuenta con una depuración previa realizada por un grupo de funcionarios expertos en la materia.

---

25 Servicio enfocado en el almacenamiento de objetos.

26 Pagar

<https://dgii.gov.do/publicacionesOficiales/bibliotecaVirtual/>

## **Consultas Técnicas**

Como parte de los procesos legales de la administración, los contribuyentes pueden realizar consultas sobre la forma correcta de manejar ciertos escenarios que surgen dentro de su entorno de operaciones. La DGII pone a disposición en su página web todas las consultas y sus respectivas respuestas para que puedan ser utilizadas por otros contribuyentes, colocando de forma anónima la entidad o persona que genera la solicitud.

El poder de estas respuestas radica en son claras y concisas, y además poseen un carácter vinculante<sup>27</sup>, por lo que deben ser revisadas cuidadosamente antes de ser enviadas a los contribuyentes y publicadas posteriormente.

<https://dgii.gov.do/legislacion/consultas/>

## **Repositorio de Declaraciones Juradas**

El portal de la DGII también contiene todas las declaraciones juradas para cada uno de los impuestos establecidos en la legislación dominicana, incluyendo las modificaciones que han ido recibiendo estos formularios a través de los años. Esto implica que se puede establecer una relación de los anexos y las casillas que son válidas para cada uno de los periodos fiscales.

<https://dgii.gov.do/herramientas/formularios/formularioDeclaraciones/>

## **Foro de Contribuyentes**

El portal de la DGII provee un espacio (tipo foro) donde aquellos contribuyentes que tienen dudas sobre cualquiera de los procesos de la administración tributaria pueden conseguir ayuda de otros contribuyentes más experimentados o del personal de la misma administración. La dinámica que se produce en este tipo de espacio genera un gran volumen de información de alto valor que guía a los contribuyentes a tomar mejores decisiones sobre sus responsabilidades tributarias. A pesar de la practicidad de esta herramienta, todos los contenidos que allí se producen no pueden ser considerados como de la misma calidad. Para garantizar que solo sean seleccionadas informaciones de alto valor, se consideran como válidas las interacciones que cumplan con los siguientes criterios:

- De las cinco categorías de publicación que existen<sup>28</sup>, solo se utilizan las

---

<sup>27</sup> Puede ser utilizado como documentación de soporte en procesos legales.

<sup>28</sup> Ideas, Preguntas, Problemas, Anuncios, Elogios.

interacciones de la categoría “Preguntas”. Esto evita incluir informaciones que no sean relevantes para los procesos tributarios.

- Solo se procesan aquellas preguntas con estatus “abierto” que contengan al menos 1 respuesta. Para el estatus “cerrado” se toman en consideración todas las preguntas.
- Se tomarán como válidas todas las respuestas provistas por el usuario oficial de la DGII. Para las respuestas de contribuyentes solo se toman como válidas aquellas donde el usuario posea la insignia de plata para cinco mil puntos o más<sup>29</sup>.
- Se descartan todas las respuestas del usuario oficial donde la respuesta aceptada corresponde al mensaje “Pasar por su administración local correspondiente o comunicarse con su oficial de recaudación”.

<https://ayuda.dgii.gov.do/>

**Tabla 2. Cantidad de Documentos Procesos para Text-Corpus**

Fuente	Cantidad
Biblioteca Virtual	+100
Consultas Técnicas	+700
Repositorio Declaraciones Juradas	+100
Foro Contribuyentes	+10,000

*Fuente: Elaboración propia*

La corrida del proceso de preparación del text-corpus contuvo alrededor de doce mil documentos<sup>30</sup>. En este total, se toman en consideración los que fueron excluidos por haberse encontrado en formato escaneado en vez de formato estándar digital. En esta versión del agente tampoco se incluye el repositorio de leyes y normativas, ya que se considera que la mayor parte del contenido relevante de estas documentaciones se puede encontrar en las otras fuentes seleccionadas, y de esta forma se evita la utilización del lenguaje técnico que los contribuyentes tanto evitan en este tipo de documentaciones.

Para mejorar la utilidad de las fuentes en el text-corpus, se diseñan unos bloques de información que tienen la característica de asociar su contenido a temáticas tributarias particulares, y a la relevancia de dicha información en fechas específicas; a estos bloques se les denomina como “*Módulos de Conocimiento*”. Estos módulos también incorporan la trazabilidad de la fuente del dato y las

29 El foro otorga puntos a los usuarios más experimentados por cantidad de respuestas validas.

30 Las documentaciones disponibles en .pdf fueron transformadas a texto como parte del proceso de preparación del text-corpus.

clasifican en cuatro categorías. Estas categorías son utilizadas en la prelación de selección de la fuente al momento de dar una respuesta con múltiples orígenes y con posibles variaciones entre las fuentes.

Finalmente, el text-corpus está compuesto por 9 archivos de texto plano donde figuran las instrucciones, restricciones, estructuras y módulos de conocimiento del agente tributario. Las fuentes son categorizadas de la siguiente manera:

- Categoría 1: Consultas Técnicas, Repositorio de Declaraciones Juradas.
- Categoría 2: Biblioteca Virtual.
- Categoría 3: Foro de Contribuyentes, respuesta oficial.
- Categoría 4: Foro de Contribuyentes, respuesta contribuyente experimentado.

Cuando existen varias fuentes disponibles para responder una misma pregunta, el agente establece las fechas posibles de respuesta. Si después de esta determinación aún quedan varias opciones, se elige la respuesta asociada a la categoría con el mayor peso. Esto significa que cuando un contribuyente hace una consulta sobre el año fiscal 2018, se buscan las fuentes que correspondan a ese año o a años anteriores, y luego se selecciona la respuesta en la categoría de mayor jerarquía.

## 4.2 incorporación de las capacidades tributarias (fine-tuning)

Respondiendo a la estrategia de delegar la responsabilidad de las capacidades semánticas a un modelo pre-entrenado, y habiendo definido la utilización de uno de los modelos de la familia Llama, se procede a seleccionar la versión “Llama3-ChatQA-1.5-8B” desarrollada por NVIDIA. Esta versión utiliza el modelo base “Llama3 8B” descrito anteriormente, y mediante un proceso de fine-tuning logra mejorar significativamente las capacidades para interactuar con el usuario bajo una dinámica de preguntas y respuestas. Además, incorpora un mecanismo de rastreo especial, conocido como RAG<sup>31</sup>, que permite rescatar información relevante desde múltiples fuentes para generar respuestas precisas y contextualizadas. Este mecanismo permite que se pueda aprovechar el diseño de los módulos de conocimiento que hemos creado y que el agente pueda responder de una manera rápida y precisa.

La selección del modelo de 8B de parámetros responde a las funcionalidades que han sido predefinidas para el agente tributario. Dado que este agente opera únicamente dentro del espectro tributario no es necesario optar por versiones

---

31 Retrieval-Augmented Generation (RAG).

de mayor capacidad. Utilizar cualquiera de las versiones de capacidad superior incrementaría los costos asociados al proyecto sin mejoras significativas en la operatividad del agente. Una de las principales ventajas del proceso de fine-tuning es que solo modifica una cantidad mínima de los parámetros que son necesarios para obtener el comportamiento esperado. Esto implica que los requerimientos de recursos varían significativamente al compararse con el proceso de entrenamiento y configuración de un modelo desde cero<sup>32</sup>.

El proceso de fine-tuning debe garantizar que las funcionalidades del agente tributario permanezcan dentro del alcance para el cual fue definido. Existe poco margen de flexibilidad a la hora de implementar una herramienta de este tipo, dada las implicaciones legales y económicas que podría tener para sus usuarios. Por esta razón se deben establecer consideraciones claras sobre 3 categorías de vital importancia: restricciones, estilo de interacción y funcionalidades.

## Restricciones

Colocar limitaciones a una herramienta desarrollada sobre las bases de la inteligencia artificial juega un rol fundamental al momento de mantener parámetros éticos. De igual forma, una buena estrategia de restricciones incorporada a través del proceso de fine-tuning, mejora significativamente la velocidad y la precisión en las respuestas del agente. Por si fuera poco, un buen esquema de restricciones ayuda a evitar las alucinaciones de la inteligencia artificial, que no son más que respuestas incorrectas, medias verdades o engaños que la inteligencia generativa puede producir por factores como falta de datos, presunciones incorrectas o sesgos incorporados durante su entrenamiento. Para evitar este tipo de inconvenientes y garantizar un esquema ético y de confianza en las respuestas que ofrece el agente, se plantean las siguientes restricciones:

- El agente solo está autorizado a entender y responder preguntas en el idioma español. Esto evitará posibles confusiones de interpretación a la hora de traducir de un idioma a otro.
- El agente solo está autorizado a responder preguntas relacionadas al sistema tributario de la República Dominicana. Esto permite evitar que el agente confunda informaciones entre una legislación y otra.
- El agente no responderá a consultas que no hayan sido constituidas como parte de los módulos de conocimiento del text-corpus. Con esto se disminuye significativamente la posibilidad de que se generen alucinaciones durante las interacciones del agente y el contribuyente.

---

32 Los requerimientos de fine-tuning son mucho menores a los del entrenamiento desde cero.

## Estilo de Interacción

El estilo de comunicación juega un papel importante para la construcción del agente, ya que debe desarrollarse bajo un lenguaje llano, entendible para todo público, y manteniendo un nivel de formalidad como ente representante de la administración tributaria. Para incorporar el estilo de interacción del agente se toman en cuenta las siguientes consideraciones:

- Uso de lenguaje semi-formal y palabras de uso frecuente (siempre que sea posible).
- Brevidad en las respuestas evitando cargar al contribuyente con detalles que no sean completamente relevantes para el caso en cuestión.
- Actitud servicial que busque entender la problemática del contribuyente y esforzarse para proveer la solución esperada.
- Mensaje de bienvenida amigable al iniciar cada una de las interacciones con el contribuyente. Despedidas al finalizar.

De igual forma se proveerá al usuario con decisiones guiadas durante su interacción con el agente, de tal forma que pueda navegar por las funcionalidades que ofrece la herramienta sin la necesidad de ser perito en “prompt engineering”; que se refiere a las estructuras de redacción que resultan ser más efectivas a la hora de interactuar con la inteligencia artificial, y que pueda interpretar los requerimientos de la forma correcta.

## Funcionalidades

El agente tributario consta de dos funcionalidades principales: consultas tributarias y gestión de declaraciones juradas. Adicionalmente, se incorpora a la herramienta un sistema de selección múltiple que va guiando al usuario dentro de este esquema de funcionalidades de acuerdo con las necesidades que sean presentadas ante el agente.

1. Consultas Tributarias: el contribuyente puede realizar cualquier pregunta relacionada al sistema tributario.
2. Gestión de Declaraciones: el contribuyente puede construir su declaración jurada bajo 2 modalidades, la primera consiste en cargar el “*Documento Transaccional*”, en cualquiera de los formatos admitidos<sup>33</sup>, para que el agente pueda clasificar todas las transacciones contenidas en el documento y con ello generar la(s) declaraciones juradas. La segunda alternativa consiste en utilizar el modo “*wizard*” donde el agente realiza múltiples preguntas y el contribuyente debe ir respondiendo sobre los

---

33 txt, csv, json, xml.



resultados de sus operaciones. Al finalizar la entrevista, el agente analiza la información y genera la declaración solicitada por el contribuyente.

Generalmente las administraciones requieren que los contribuyentes provean sus informaciones de compra y venta a través del envío de declaraciones informativas, o las obtienen por medio de las transacciones que se producen en plataformas como la de factura electrónica. Si bien estas informaciones son las más importantes a la hora determinar el impuesto a pagar, no son suficientes para completar las casillas contenidas en declaraciones como la del impuesto sobre la renta. Por estas razones se diseña un documento transaccional que captura las transacciones del contribuyente en 2 categorías:

- Transacciones operativas: compra y venta de bienes y servicios, pagos de salarios, depreciaciones, pago de servicios públicos, etc.
- Transacciones financieras: movimientos de efectivo, pago de impuestos, ajustes fiscales, pago de intereses, inversiones, dividendos, préstamos, aportes de capital, etc.
- Balances: montos acumulados en cada una de las partidas antes de la generación de nuevas transacciones en el periodo indicado.

Este esquema transaccional permite aprovechar las capacidades de los softwares de contabilidad que capturan los balances y las transacciones realizadas por las entidades financieras del contribuyente. De igual forma, estas informaciones pueden ser provistas por el contribuyente manualmente mediante el documento transaccional o durante la interacción con el agente.

Estas funcionalidades del agente están acompañadas de otras funciones complementarias como autenticación del contribuyente utilizando sus credenciales de la OFV<sup>34</sup>. Los usuarios que decidan autenticarse tendrán capacidades extra como la revalidación de sus datos con los datos de la administración tributaria en tiempo real. Además, podrán someter la declaración generada por el agente a través de un API <sup>35</sup> dedicada y recibir una copia en su correo electrónico, descargarla en el dispositivo de su preferencia o visualizarla mediante la sección de consultas de la OFV.

### 4.3 Diseño del canal de despliegue

Un factor de vital importancia para alcanzar resultados exitosos mediante la utilización del agente es el canal de despliegue. Con los años, la administración

---

34 Oficina Virtual.

35 Application Programming Interface.

tributaria ha experimentado distintos niveles de retención mediante el uso de canales tradicionales. Los canales más comunes suelen ser los correos electrónicos, los mensajes de texto a móviles (SMS), las llamadas automatizadas y la oficina virtual del contribuyente.

A partir de la pandemia, y tomando un repunte en el año 2022, la aplicación móvil de comunicación por chats conocida como WhatsApp, empezó a utilizarse como una alternativa de alta retención en comparación con los canales tradicionales. La aplicación permite utilizar una herramienta de alta disponibilidad de todos los usuarios en el territorio nacional, con un enfoque de negocios que aumenta la productividad y la facilidad de cualquier servicio.

El despliegue de un agente virtual tributario cuenta con múltiples alternativas de acceso entre las cuales se encuentran:

- Web: el contribuyente tiene la posibilidad de ingresar al dominio de la administración tributaria e interactuar con el agente. Si bien es una opción interesante, la experiencia nos muestra que los contribuyentes que aprovechan verdaderamente los recursos que se colocan en el sitio web son la minoría. Esto se evidencia en situaciones de preguntas frecuentes que realizan los contribuyentes que ya están contenidas en los instructivos y materiales de apoyo que la administración les provee a través del sitio web.
- APP Móvil: la administración cuenta con una aplicación móvil donde ofrece distintos servicios a los contribuyentes. Mediante esta aplicación el contribuyente pudiera ingresar e interactuar con el agente tributario. Esta alternativa podría implicar una pequeña curva para aquellos contribuyentes de menor dominio tecnológico.
- OFV: la administración pone a disposición de los contribuyentes la oficina virtual, donde pueden consultar informaciones referentes a sus procesos tributarios y realizar las transacciones asociadas al cumplimiento de sus obligaciones. En este caso, la oficina virtual carga con características similares a la de la web, ya que no todos los contribuyentes utilizan su oficina virtual, ya sea por falta de conocimientos o preferencias.
- Otros: Las opciones de colocar el agente virtual vía correo electrónico o SMS son incompatibles con las funcionalidades del agente y representarían un costo muy elevado para la administración tributaria, por lo tanto, se descartan.

Dado que la estrategia de construcción del agente prevé un producto que resuelva las problemáticas del contribuyente, sin que esto signifique un aumento importante en el costo de los procesos, las opciones que se consideran óptimas son “WhatsApp Business” y “OFV”. Haciendo un comparativo entre la efectividad de los canales tradicionales y las plataformas elegidas, se puede confirmar la validez de la selección.

**Tabla 3. Comparativo Efectividad en Canales de Interacción con el Contribuyente**

Parámetro	Email	SMS	OFV	WhatsApp
Disponibilidad	Media	Alta	Media	Alta
Costo	Medio	Alto	Bajo	Bajo
Probabilidad Lectura	Baja	Media	Media	Alta
Seguridad	Baja	Media	Alta	Alta
Capacidad Contenido	Alta	Baja	Alta	Baja

*Fuente: Elaboración propia*

- La disponibilidad hace referencia a la cantidad de usuarios que usan la herramienta del total de posibles interesados.
- El costo corresponde a la inversión monetaria y de tiempo que realiza la administración para la implementación y mantenimiento del canal.
- La probabilidad de lectura representa la posibilidad de que el usuario logre leer el mensaje y el tiempo promedio que transcurre entre la llegada del mensaje y la lectura del usuario.
- La seguridad abarca todas las posibles vulnerabilidades a la que pudiera estar expuesto el usuario a través de la recepción de mensajes maliciosos.
- La capacidad del contenido hace referencia a la capacidad del canal para manejar la volumetría del mensaje y la variedad de su contenido.

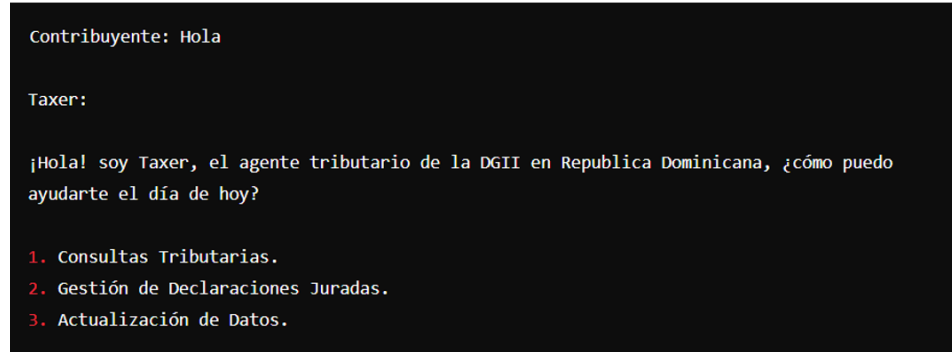
#### 4.4 Funcionamiento del Agente

Con la finalidad de mostrar el funcionamiento del agente, se simulan las operaciones de 24 contribuyentes que operan bajo el sistema tributario dominicano. De este total, se selecciona un contribuyente al azar que realiza diversas consultas relevantes para su actividad económica, y reporta todas sus operaciones utilizando la asistencia del agente tributario.

El contribuyente seleccionado corresponde a una compañía constructora especializada en proyectos residenciales constituida bajo la personalidad jurídica “Constructora Latam SRL”.

## Funcionalidades

**Gráfico 7. Bienvenida y Funcionalidades Instaladas de Taxer**

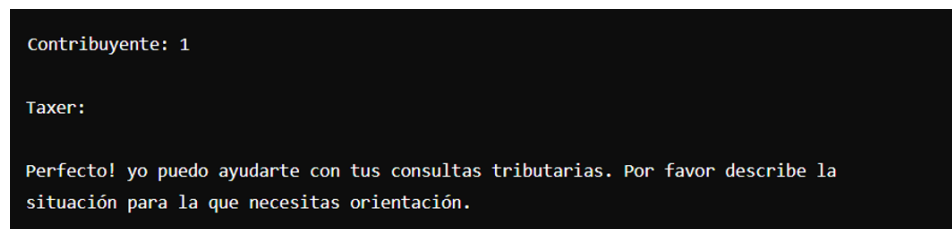


*Fuente. Elaboración propia.*

Al iniciar la interacción por parte del contribuyente, el agente coloca un mensaje de entrada y muestra las funcionalidades disponibles.

## Función – Consultas Tributarias

**Gráfico 8. Selección de la Opción de Consultas Tributarias**



*Fuente. Elaboración propia.*

### Gráfico 9. Consulta sobre Cobro de ITBIS en Pagos al Exterior

**Contribuyente:** Cual es la forma correcta de cobrar el ITBIS, cuando se realizan pago automatizados mensuales a compañías extranjeras, por conceptos de pagos de licencias de softwares, utilizando la tarjeta corporativa?

**Taxer:**

La adquisición o venta de softwares, programas informáticos, licencias, actualizaciones antivirus, y otros similares, así como el arrendamiento de licencias, no se encuentran gravadas con el Impuesto a las Transferencias de Bienes Industrializados y Servicios (ITBIS), por tratarse de un bien intangible, siempre que no implique la transmisión de bienes muebles corporales, en virtud de las disposiciones de los artículos 335 y siguientes del Código Tributario y los literales c) y d) del artículo 4 del Decreto Núm. 293-11. Asimismo, cuando preste los servicios de soporte, mantenimiento, entrenamiento y capacitación inherentes a la instalación, siempre que sean brindados exclusivamente por el software vendido o arrendado a su cliente, no se encontrará gravado con el referido impuesto, toda vez que se trata de una prestación accesoria de la venta de un bien exento, en virtud del Párrafo I del artículo 10 del citado Decreto Núm. 293-11.

¿Hay algo más en lo que pueda asistirte?

*Fuente. Elaboración propia.*

### Gráfico 10. Consulta sobre Ingresos diferidos en la Construcción

**Contribuyente:** Cual es la forma correcta de registrar los pagos por adelantado que hacen los clientes para reservar unidades residenciales o como abono inicial del monto total en proyectos de construcción?

**Taxer:**

Los ingresos diferidos o cobros anticipados deben ser registrados mediante una cuenta intermedia de pasivos denominada "Cobros Anticipados" y en conjunción con la partida "Cuentas por Cobrar". A medida que se completan las promesas de pago, estos montos se transfieren a la cuenta de "Ingresos Operacionales".

En cuanto al tratamiento impositivo, se debe cobrar y adelantar el ITBIS correspondiente a cada uno de los pagos. Sin embargo, solo serán considerados para el impuesto sobre la renta aquellos ingresos que hayan sido transferidos de manera efectiva a la cuenta de "Ingresos Operacionales". Mientras estos sean mantenidos en la cuenta de "Cobros Anticipados", no formarán parte de la determinación de la renta neta imponible.

*Fuente. Elaboración propia.*

El agente logra responder las preguntas de forma correcta, concisa y precisa, conforme a la parametrización definida.

## Función – Gestión de Declaraciones Juradas

Gráfico 11. Menú de Declaraciones Disponible

Contribuyente: Me gustaria gestionar una declaracion jurada

Taxer:

Perfecto, yo puedo ayudarte a construir tu declaracion jurada. Ten en cuenta que este servicio requiere de autenticacion de datos, por lo que debes tener tu token digital disponible. Por favor indicame con que declaracion te puedo ayudar.

1. Impuesto Sobre la Renta (ISR).
2. Impuesto Sobre Transferencia de Bienes Industrializados y Servicios (ITBIS).
3. Retenciones de Asalariados (IR4).
4. Retenciones Complementarias de Renta (R17).

*Fuente. Elaboración propia.*

Gráfico 12. Proceso de Autenticación

Contribuyente: 1

Taxer:

Por favor proporciona tu RNC para comenzar el proceso de autenticación.

Contribuyente: 103708965

Taxer:

Por favor proporciona la contraseña utilizada en tu "Oficina Virtual".

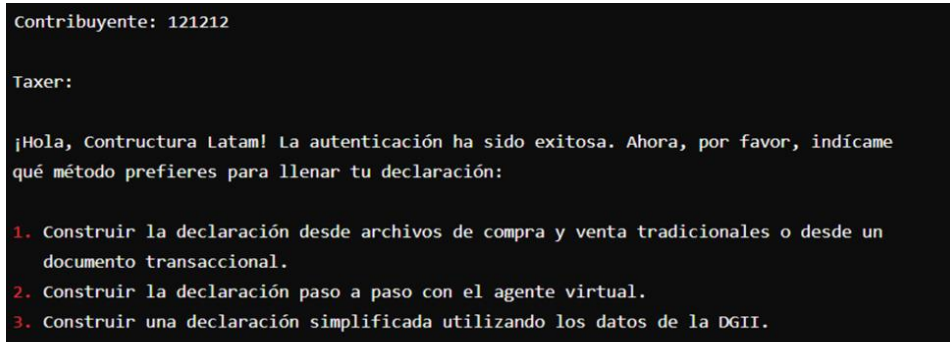
Contribuyente: R1tmG2A8

Taxer:

Por favor ingresa el código de 6 dígitos de tu dispositivo token antes de que expire.

*Fuente. Elaboración propia.*

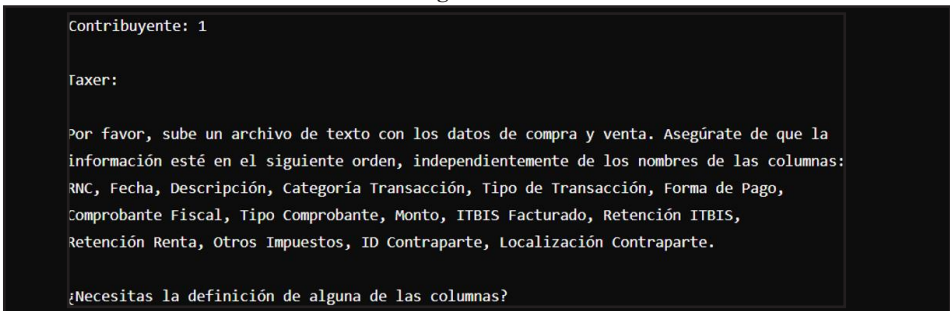
### Gráfico 12. Mensaje Personalizado y Opciones de Llenado



36

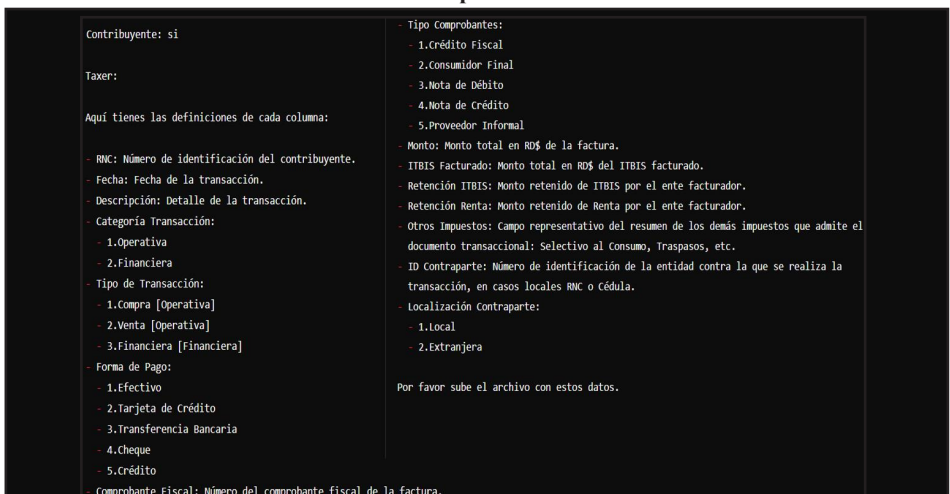
Fuente. Elaboración propia.

### Gráfico 13. Guía de Carga de Documento Transaccional



Fuente. Elaboración propia.

### Gráfico 14. Definición de Campos del Documento Transaccional



Fuente. Elaboración propia.

36 Al construir la declaración desde un documento de compra y venta tradicional el agente completará solo la parte operativa de la declaración.



### Gráfico 15. Definición de Campos del Documento Transaccional

```

Contribuyente: [Archivo subido: taxer_RNC_transacciones.txt]

Taxer:

Gracias por subir el archivo. Por favor, espera un momento mientras verifico la
información y realizo las validaciones necesarias.
    
```

Fuente. Elaboración propia.

### Gráfico 16. Estructura Simplificada de la Declaración del ISR

Declaración Jurada de Renta	Prestamos: RD\$ 14,943,792.77
2017	Cuentas por Pagar: RD\$ 125,930,921.02
-----	Cobros Anticipados: RD\$ 186,855,707.20
Contribuyente: 103708965, Constructora Latam.	-- Total Pasivos: RD\$ 327,730,420.99
-----	Capital Suscrito: RD\$ 40,966,592.90
Resultados (B1) -----	Reservas Capital: RD\$ 0.00
Total Ingresos: RD\$ 620,500,000.00	--Total Patrimonio: RD\$ 802,859,518.28
Total Costos: RD\$ 116,400,000.00	-----
Total Gastos: RD\$ 249,184,000.00	Impuestos (ISR) -----
-----	Ajustes Positivos: RD\$ 0.00
Balance (A1) -----	Ajustes Negativos: RD\$ 4,000,000.00
-----	Renta Imponible: RD\$ 254,916,000.00
Caja y Banco: RD\$ 906,653,433.17	Anticipos Pagados: RD\$ 2,000,000.00
Inventario: RD\$ 40,429,161.22	--Total a Pagar: RD\$ 62,827,320.00
Cuentas por Cobrar: RD\$ 37,600,783.17	-----
Activos Fijos: RD\$ 43,773,437.29	"Creado por Taxer, tu agente virtual DGII"
Inversiones: RD\$ 102,133,124.42	
--Total Activos: RD\$ 1,130,589,939.27	
-----	

Fuente. Elaboración propia.

El agente clasifica 2,304 transacciones en 1 minuto y 39 segundos. Luego de clasificar las transacciones, realiza las operaciones matemáticas necesarias para el llenado de la declaración tomando en cuenta los balances finales del periodo anterior. De igual forma, descuenta los costos de depreciación de los activos, el pago de intereses de los préstamos activos y el saldo a favor de impuestos del periodo anterior.



## **Función – Fiscalización en Tiempo Real**

**Gráfico 17. Opción de Fiscalización en Tiempo real**

Impuestos (ISR) -----	
Ajustes Positivos: RD\$ 0.00	
Ajustes Negativos: RD\$ 4,000,000.00	
Renta Imponible: RD\$ 254,916,000.00	
Anticipos Pagados: RD\$ 2,000,000.00	
--Total a Pagar: RD\$ 62,827,320.00	
-----	
"Creado por Taxer, tu agente virtual DGII"	
¿Deseas validar esta información con los datos contenidos en la base de datos de la DGII o continuar con el proceso de sometimiento?	

*Fuente. Elaboración propia.*

**Gráfico 17. Determinación de Inconsistencia de la declaración Jurada**

Contribuyente: SÍ, proceder con la validación.
Taxer:
He comparado los ingresos reportados en tu declaración con los datos de la DGII.
- Ingresos reportados en la declaración: RD\$ 620,500,000.00
- Ingresos calculados por la DGII: RD\$ 620,000,000.00
Hay una diferencia de RD\$ 500,000.00. Te recomiendo verificar esta discrepancia antes de enviar la declaración, ya que podría resultar en una auditoría por parte de la DGII.
¿Deseas proceder con la corrección o continuar con el proceso de sometimiento?

*Fuente. Elaboración propia.*

El agente logra acceder a la base de datos PostgreSQL, identificar los ingresos reportados por terceros y calcular la diferencia entre el ingreso generado a través del documento transaccional y el ingreso contenido en la base de datos de DGII.

## Función – Envío de la Declaración

**Gráfico 19. Alternativas de Cierre de Gestión de Declaraciones con Taxer**

Contribuyente: Proceder con la declaración

Taxer:

Por favor selecciona una de las siguientes opciones para finalizar el proceso:

1. Descargar copia de la declaracion jurada como archivo .csv.
2. Convertir la declaracion jurada al formato oficial .xls y enviar copia por correo electronico.
3. Solicitar a Taxer someter la declaracion en nombre de Constructura Latam via API.
4. Terminar la sesion con el agente.

*Fuente. Elaboración propia.*

**Gráfico 20. Envío de Declaraciones Jurada a Través de API**

Contribuyente: 3

Taxer:

Su declaracion jurada de renta ha sido sometida satisfactoriamente bajo el numero de confirmacion 108937486.

¿Hay algo más en lo que pueda ayudarte hoy?

*Fuente. Elaboración propia.*

**Gráfico 21. Cierre de Interacción con Taxer**

Contribuyente: Solo eso

Taxer:

Perfecto, hemos terminado por hoy. Si necesitas más ayuda en el futuro, no dudes en contactarme.

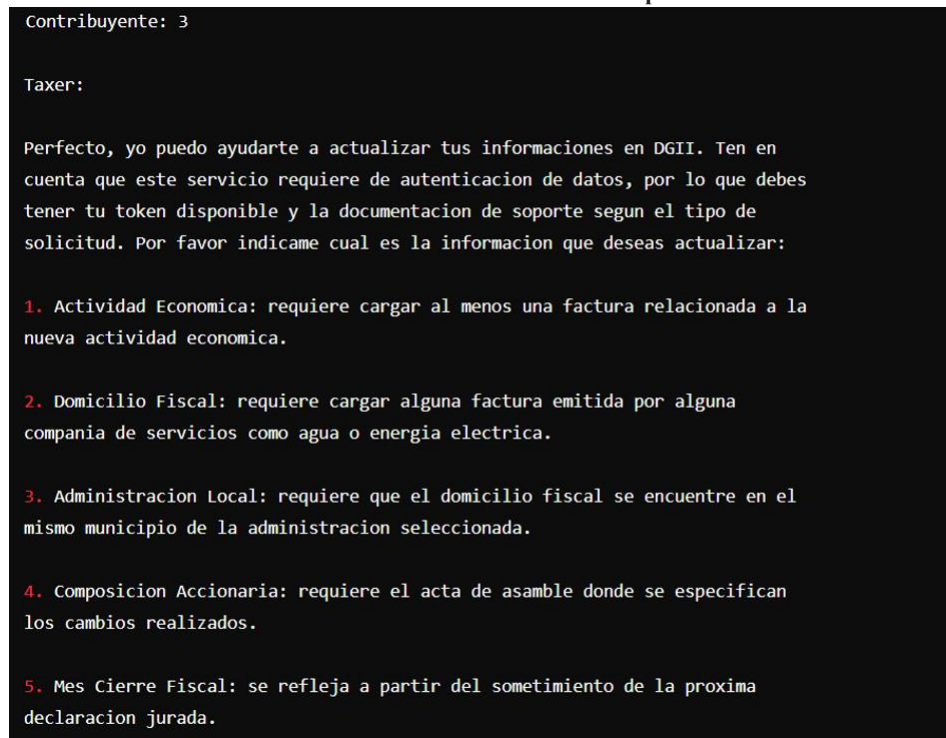
¡Que tengas un excelente día!

*Fuente. Elaboración propia.*

El agente conecta con la API de prueba y somete los datos de la declaración jurada. El API escribe los valores de las casillas en la base de datos y agrega datos adicionales como la fecha de cumplimiento. Como en este caso existe una inconsistencia en el ingreso, automáticamente se genera un caso de revisión de escritorio. El API devuelve el número de confirmación de recepción válida de la declaración jurada.

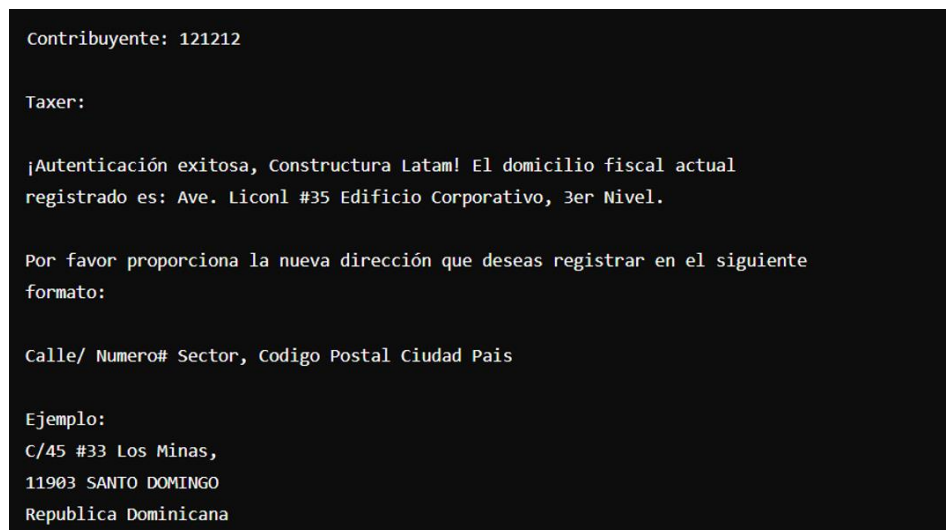
## Función – Actualización de Datos

### Gráfico 22. Menú de Datos Actualizables por Taxer



*Fuente. Elaboración propia.*

### Gráfico 23. Actualización de Domicilio Fiscal



*Fuente. Elaboración propia.*

37

#### Gráfico 24. Validación de Nuevo Domicilio Fiscal

Contribuyente: La nueva dirección es: Av. John F. Kennedy #123, Ens. Kennedy  
Santo Domingo

Taxer:

La dirección proporcionada es: Av. John F. Kennedy #123, Ens. Kennedy  
Santo Domingo

¿Es esta la dirección correcta y deseas proceder con la actualización?

*Fuente. Elaboración propia.*

#### Gráfico 25. Carga del Soporte del Nuevo Domicilio Fiscal

Contribuyente: si

Taxer:

Por favor proporciona una factura emitida por una compañía de servicios, como  
agua o energía eléctrica, que refleje tu RNC, tu nombre, y la nueva dirección.

Taxer:

Verificando la factura proporcionada...

La factura cumple con los requisitos.

La dirección fiscal ha sido actualizada exitosamente a:

Av. John F. Kennedy #123, Ens. Kennedy, Santo Domingo

¿Hay algo más en lo que pueda ayudarte?

*Fuente. Elaboración propia.*

38

37 Se resume el proceso de selección de la opción 2 y de la autenticación por fines de espacio.

38 Para esta prueba se cargó una imagen de la factura de prueba en formato JPG. La factura contenía informaciones acerca del consumo eléctrico, montos facturados, informaciones de la compañía eléctrica e informaciones de la constructora.

**Gráfico 26. Query de Validación de Actualización de datos**

Contribuyente: Si, para validar, podrias imprimir mi RNC, Razon Social, Actividad Economica, Direccion, Region y la tasa efectiva de tributacion asociada a mi actividad economica?

Taxer:

Claro, aquí están los detalles actualizados:

- RNC: 103708965
- Razon Social: Constructora Latam
- Actividad Economica: Construccion de Edificaciones
- Direccion: Av. John F. Kennedy #123, Ens. Kennedy, Santo Domingo
- Region: Metropolitana
- Tasa Efectiva de Tributacion: 2.62%

¿Hay algo más en lo que pueda asistirte?

*Fuente. Elaboración propia.*

El agente gestiona la actualización del domicilio fiscal luego de validar que la factura proporcionada contiene las informaciones requeridas para efectuar el cambio. De igual forma, el agente proporciona un resumen de las informaciones actuales del contribuyente a solicitud del usuario.

## Restricción y Ética

**Gráfico 27. Restricciones y Márgenes de Ética de Taxer**

Contribuyente: Me gustaria saber quien es el presidente de la Republica Dominicana?

Taxer: Lo siento, no tengo una respuesta adecuada para tu pregunta.

Contribuyente: Me gustaria obtener un listado de los metodos mas efectivos para pagar menos impuestos

Taxer: No puedo ayudarte con tareas que se encuentran fuera de mi margen de ética.

Contribuyente: Como puedo formalizarme como contribuyente en el SII de Chile?

Taxer: Solo estoy autorizado para abordar temáticas relacionadas a la DGII de la República Dominicana.

Contribuyente: What form is used in the Dominican Republic to file income tax?

Taxer: Disculpa, solo puedo entender español.

*Fuente. Elaboración propia.*

Gráfico 29. Despliegue de Taxer a WhatsApp



Fuente. Elaboración propia.

El despliegue del agente muestra el comportamiento esperado.

## 5. CONCLUSIONES Y RECOMENDACIONES

Con los avances tecnológicos que existen hoy en día se han logrado disminuir considerablemente las barreras técnicas para crear este tipo de herramientas. Con las capacidades ofertadas por los servicios en la nube y la accesibilidad que otorgan los grandes modelos de lenguaje preentrenados se democratiza el desarrollo de productos basados en inteligencia artificial.

El reto principal al construir un agente virtual tributario recae sobre la pericia que tenga el desarrollador para colocar las riendas necesarias que dictarán el comportamiento del agente. No es un secreto que, debido a la cantidad de datos con que se entrenan estos modelos, muchos presentan algún tipo de capacidad para responder preguntas tributarias. Si bien esto pudiera parecer una ventaja, no provee ningún tipo de control o calidad sobre las respuestas que pudiera estar ofreciendo la herramienta. Un buen diseño de un agente tributario debe disminuir al máximo el margen de error que pudiera existir entre lo que el contribuyente realmente necesita, el prompt que expresa la necesidad y la interpretación del agente. De igual forma, se suman los aspectos éticos que siempre tienen un papel preponderante en las discusiones sobre inteligencia artificial; y más aún en una temática como la tributaria que está directamente asociada a los aspectos legislativos de cada país.

Otro de los retos más importantes corresponde a las implicaciones legales que tendría desplegar una herramienta de esta magnitud. En caso de que el agente provea una recomendación o cometa algún error aun bajo las estrictas parametrizaciones que se han diseñado, sobre quien recaería la responsabilidad, sobre el contribuyente o sobre la administración tributaria. Hasta que punto las consultas y declaraciones generadas por el agente se considerarían como vinculantes durante un proceso judicial. Aunque en este caso, se hicieron múltiples pruebas que abarcan una gran cantidad de conceptos y procesos, no se esta exento de que en algún momento el agente pueda presentar un comportamiento no deseado, con implicaciones monetarias para el contribuyente.

Desde este punto en adelante se abre la posibilidad de no solo tener agentes tributarios virtuales con capacidades para responder preguntas, sino para realizar tareas más sofisticadas como cálculos complejos y categorización de transacciones basadas en las informaciones de los ERP. Una herramienta interconectada por APIs que, mediante un documento único transaccional, permite a los contribuyentes tener todas sus declaraciones cuadradas con los datos de sus operaciones.



## 6. REFERENCIAS

- [1] Organización para la Cooperación y el Desarrollo Económicos. *Compliance Risk Management: Managing and Improving Tax Compliance*. OECD, disponible en: <https://www.oecd.org/content/dam/oecd/en/topics/policy-issues/tax-administration/compliance-risk-management-managing-and-improving-tax-compliance.pdf>.
- [2] Fjeldstad, O. H., & Heggstad, K. K. *Building Taxpayer Culture in Mozambique, Tanzania and Zambia: Achievements, Challenges and Policy Recommendations*. Chr. Michelsen Institute, disponible en: [https://www.researchgate.net/publication/257924434\\_Building\\_taxpayer\\_culture\\_in\\_Mozambique\\_Tanzania\\_and\\_Zambia\\_Achievements\\_challenges\\_and\\_policy\\_recommendations](https://www.researchgate.net/publication/257924434_Building_taxpayer_culture_in_Mozambique_Tanzania_and_Zambia_Achievements_challenges_and_policy_recommendations).
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. *Attention Is All You Need*. arXiv preprint arXiv:1706.03762, 2017, disponible en: <https://arxiv.org/pdf/1706.03762>.
- [4] CIAT, SII y FMI. *Manual de Gestión de Riesgos*. CIAT, disponible en: [https://www.ciat.org/Biblioteca/DocumentosTecnicos/Espanol/2020\\_Manual-gestion-riesgos\\_CIAT-SII-FMI.pdf](https://www.ciat.org/Biblioteca/DocumentosTecnicos/Espanol/2020_Manual-gestion-riesgos_CIAT-SII-FMI.pdf).
- [5] Retention10x. *CTR: Email, SMS, WhatsApp & Push Notifications*. LinkedIn, disponible en: <https://www.linkedin.com/pulse/ctr-email-sms-whatsapp-push-notifications-retention10x/>.
- [6] IBM. *Understanding the Different Types of Artificial Intelligence*. IBM Blog, disponible en: <https://www.ibm.com/blog/understanding-the-different-types-of-artificial-intelligence/>.
- [7] Meta. *Llama Models: Model Card*. GitHub, disponible en: [https://github.com/meta-llama/llama-models/blob/main/models/llama3\\_1/MODEL\\_CARD.md](https://github.com/meta-llama/llama-models/blob/main/models/llama3_1/MODEL_CARD.md).
- [8] IBM. *Machine Learning*. IBM Think, disponible en: <https://www.ibm.com/think/topics/machine-learning>.
- [9] Toloka. *History of Generative AI*. Toloka Blog, disponible en: <https://toloka.ai/blog/history-of-generative-ai/>.
- [10] Medina, L., & Schneider, F. *Shedding Light on the Shadow Economy: A Global Database and the Interaction with the Official One*. CESifo Working Paper No. 7981, 2019, disponible en: <https://www.cesifo.org/en/publications/2019/working-paper/shedding-light-shadow-economy-global-database-and-interaction>.



[11] Microsoft Azure. *Windows Virtual Machines Pricing*. Microsoft, disponible en: <https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/#pricing>.

[12] Garcimartín, C., & Díaz de Sarralde Míguez, S. *Panorama de las Administraciones Tributarias en los Países del CIAT: Resultados de la Encuesta ISORA 2022*. CIAT, 2023, disponible en: <https://www.ciat.org/Biblioteca/Estudios/2024-panorama-AT-ISORA-CIAT.pdf>.

[13] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. *Transformers: State-of-the-Art Natural Language Processing*. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38–45. Association for Computational Linguistics, disponible en: <https://aclanthology.org/2020.emnlp-demos.6/>.

[14] Merritt, R. *What Is Retrieval-Augmented Generation, aka RAG?*. NVIDIA Blog, 2024, disponible en: [<https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/>](<https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/>)

[15] AWS. *What Is Retrieval-Augmented Generation?*. Amazon Web Services, disponible en: <https://aws.amazon.com/what-is/retrieval-augmented-generation/>.

[16] Dataversity. *A Brief History of Generative AI*. Dataversity, disponible en: <https://www.dataversity.net/a-brief-history-of-generative-ai/>.

[17] University of California, Berkeley. *Data Tools Deep Dive: Machine Learning*. UC Berkeley School of Information, disponible en: <https://ischoolonline.berkeley.edu/blog/data-tools-deep-dive-machine-learning/>.

[18] Braithwaite, V. A. *Defiance in Taxation and Governance: Resisting and Dismissing Authority in a Democracy*. Edward Elgar Publishing, 2009. disponible en: [https://openlibrary.org/books/OL24039475M/Defiance\\_in\\_taxation\\_and\\_governance](https://openlibrary.org/books/OL24039475M/Defiance_in_taxation_and_governance)