

## AGENCIAS DE IA Y LA GESTIÓN DE INVERSIONES, UN DESPERTAR INCIPIENTE.

En una columna anterior se discutieron las posibilidades de utilizar modelos de lenguaje a gran escala (o LLMs, “Large Language Models”) y sus eventuales aplicaciones en Finanzas. En esta columna exploramos algunos avances en el desarrollo de IAs enfocándonos en un caso de uso posible: la ejecución de un proceso de inversiones.

Por Rubén Catalán, CFA. | Subgerente de Renta Variable Internacional y Balanceados, BICE Inversiones

### Primer Paso: del *Prompt* al Agente

A estas alturas es muy probable que el lector ya esté familiarizado con lo que es un “*prompt*”. En pocas palabras es un mensaje enviado a una LLM en el contexto de una conversación. Este mensaje puede tener diversos fines: querer entender conceptos complejos, buscar información que el modelo “sabe” al haber sido entrenado con esa información, utilizar su pensamiento lógico para organizar un proyecto o tarea más compleja, o como una interfaz que traduzca entre idiomas, o incluso de inglés a código -como Python, SQL u otros lenguajes de programación-, entre otros.

Asimismo, también tienen limitaciones: son costosos, sufren alucinaciones, reproducen los sesgos presentes en su base de entrenamiento, etc.

Entendiendo que son modelos “de lenguaje”, es evidente que hemos estado pidiéndoles demasiado. Si le pregunto a una persona cuanto es  $2.373 \times 975^7$ , y le pido que responda lo primero que se le viene a la cabeza sin hacer ningún cálculo<sup>1</sup>, el resultado con seguridad estará equivocado. Si le hago una pregunta de un dominio específico, la respuesta estará errada si no se acude a un libro especializado para responderla.

El lenguaje es una herramienta para comunicarnos, expresar ideas y contrastarlas con las de los demás. Los humanos hemos aprendido a usar las herramientas correctas para resolver problemas: una calculadora (o lápiz y papel) para un cálculo, un buscador web o una biblioteca para resolver una duda.

El reconocer esto ha propiciado una nueva vertiente de desarrollo en la IA: los agentes. Un agente sigue siendo un LLM, pero con diferencias clave:

- **Se les dota de una identidad:** por ejemplo, “eres un analista de inversiones del sector *utilities*”.
- **Se les asigna un objetivo explícito:** “Mediante análisis cuantitativos y cualitativos sobre los actores del sector, identificas cambios en la posición competitiva y efectividad de su estrategia, proyectas resultados financieros, etc”.
- Se puede definir una **secuencia de pasos** para desarrollar su tarea,
- **Se les entrega la información relevante para su trabajo**, como el acceso a los EEFF de los últimos 5 años<sup>2</sup>,

<sup>1</sup> Que es como funciona un LLM, sólo elaborando cadenas de texto.

<sup>2</sup> Con técnicas como el RAG (Retrieval Augmented Generation) el modelo busca información en una base de datos o documentos externos y los utiliza para generar respuestas más precisas.

- Usan herramientas, acceden a internet<sup>3</sup> o desarrollan código y modelos matemáticos<sup>4</sup>, etc.

En agosto pasado, cuando Microsoft dio a conocer **AutoGen**<sup>5</sup>, el primer agente autónomo que sigue este modelo, justamente utilizó como ejemplo una tarea en el ámbito financiero: pedirle al agente que graficara la evolución del retorno de AAPL y TSLA.

### El paso siguiente: del Agente a la Agencia

Estas herramientas permiten sobreponerse a buena parte de las limitaciones del LLM básico, aunque no son garantía de éxito. Una de las dificultades es que estos agentes cuando son enfrentados a tareas amplias y complejas tienden a ser menos eficaces: un agente que lo “hace todo” entrega resultados de peor calidad que un agente especializado en una tarea más específica.

Eso nos lleva naturalmente a la especialización de tareas y distribución de las tareas de un proyecto entre múltiples agentes. Quizás la realización de un gráfico es una tarea moderadamente simple: sólo necesitamos buscar los datos en línea, implementar un ETL<sup>6</sup>, y finalmente escribir un código que utilice estos datos para generar el gráfico<sup>7</sup>. Otra cosa es realizar un análisis financiero “hecho y derecho” de una compañía, que contemple tanto su situación financiera, la de sus competidores, el contexto de mercado actual, y múltiples otros análisis que puedan ser utilizados para llegar a la opinión respecto del activo.

Sin embargo, como ya hemos visto, cada una de estas tareas puede ser asignada a un agente. O incluso se puede ser más específico, un agente puede estar especializado en buscar información en internet, otro en generar pdfs desde las distintas fuentes, generar *embeddings*<sup>8</sup> e indexarlos, otro puede asegurarse de chequear la calidad del código que se genere, otro puede estar encargado de la elaboración de los modelos DCF, y así sucesivamente.

Al igual que en las organizaciones reales, esto levanta la necesidad de gestionar los flujos de comunicación entre los distintos agentes, los que se establecen al

organizar la agencia: se establecen roles individuales, estructuras jerárquicas, y checks y controles que se encarguen de asegurar la calidad del output de cada uno de estos agentes.

### Y finalmente a una Agencia Especialista en Inversiones

En marzo de este año ya comenzaron a ser distribuidas en internet aplicaciones algo más sofisticadas, que muestran que efectivamente el establecer este tipo de Agencias es posible. Un ejemplo es el **GPT Investor**<sup>9</sup>, construido sobre Claude3 Opus, que es capaz de especializarse en una industria, recopilar información financiera (desde EDGAR), artículos, noticias, rankings de analistas, e incluso realizar *sentiment analysis* sobre estos materiales. Aunque en su implementación dista mucho de ser exhaustiva, al no disponer de fuentes de información lo suficientemente profundas ni de conocimiento específico respecto a cómo analizar diferentes industrias, sirve para ejemplificar que, a modo de prueba de concepto, el realizar este tipo de tareas ya es posible.

El análisis de emisores es sólo uno de varios tipos de análisis dentro de la disciplina financiera. Este esquema de funcionamiento puede extenderse hacia análisis macroeconómicos, el uso de modelos *quant* para la valorización de oportunidades de mercado, sistemas de monitoreo constante de información en redes sociales, medios de comunicación, o incluso medios oficiales (pensando en hechos esenciales o reportes financieros). Pueden ser combinados con algoritmos de operación en mercados, con modelos de riesgo, evaluación de escenarios, y tantos otros.

Al igual que en la práctica tradicional, la forma en cómo se estructuran los procesos de inversiones y se implemente cada una de sus etapas con seguridad generará visiones distintas respecto a los activos. En algunas instituciones priman visiones *top-down*, mientras que en otras el enfoque es *bottom-up*. En algunas, la presencia de un CIO permea las decisiones que se toman dentro de cada fondo y estrategia, en otras se fomentan las visiones distintas y no se

<sup>3</sup> Ya sea navegación normal y acceso a portales de búsqueda, o vía el acceso a APIs con información especializada.

<sup>4</sup> Con el acceso a una consola de comandos y la posibilidad de generar código, desarrollan procedimientos para realizar modelos financieros o cálculos matemáticos.

<sup>5</sup> “AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation”, Wu et al, <https://arxiv.org/abs/2308.08155>

<sup>6</sup> Acrónimo que significa Extract - Transform - Load, y que representa el proceso de **extraer** datos de múltiples fuentes, su

**transformación** y preparación para finalmente **cargarlos** en un sistema que permita su uso posterior en otros procesos y análisis.

<sup>7</sup> Vía Plotly o similares.

<sup>8</sup> El proceso de transformar un contenido (pdf, texto, etc) en vectores para ser utilizados con RAG.

<sup>9</sup> Ver [https://www.youtube.com/watch?v=\\_JHjCPal-Zc](https://www.youtube.com/watch?v=_JHjCPal-Zc) y <https://github.com/mshumer/gpt-investor>

depende de un único “house view” para la toma de decisiones. Estas distintas maneras de tratar la disciplina, cada una con sus beneficios y costos, pueden ser desmenuzadas y traspasadas a Agencias Especialistas que pueden apoyar y complementar el análisis de forma sustantiva. El objetivo no es eliminar al analista o al PM del *loop*, sino que liberarlo para permitirle tomar decisiones más informadas y, presumiblemente, que generen mejores resultados.

¿Qué termina rondando en la cabeza al enfrentarse a esta nueva realidad? Varias cosas:

**Los datos son críticos.** Los datos son la piedra angular sobre la que esta nueva tecnología se levanta. Son necesarios millones de *datapoints* para entrenar estos modelos y permitir que “emerjan” características como el pensamiento lógico o la comprensión del lenguaje por parte de los LLMs. Así como el texto es la fuente primaria de LLMs como ChatGPT, otro tipo de datos pueden ser usado para entrenar otro tipo de modelos. Tesla utiliza los datos recopilados por toda su flota global para enseñarle a los autos a manejarse solos. Nvidia usa datos simulados de como funcionan las propiedades físicas en la realidad para enseñarle a robots a caminar, abrir puertas o tomar cosas. Midjourney utiliza millones de imágenes para enseñar a la IA a hacer ilustraciones, fotos o pinturas. Sora utiliza horas de grabaciones para poder elaborar videos a partir de texto. Entonces el desafío para el mundo financiero es: ¿estamos guardando los datos respecto a nuestros procesos de inversiones de forma sistemática?; asumiendo que los tengamos, ¿qué le podemos enseñar a una IA a partir de ella?

**El foco del humano a cargo del proceso pasa de estar en el “hacer” a estar en el “pensar”.** El sistema por sí mismo será capaz de automatizar las tareas tediosas y repetitivas, idealmente llegando hasta a proponerle un curso de acción. En ese punto el criterio humano es el encargado de evaluar si se llegó a una conclusión correcta o no, corrigiendo los errores que pueda haber cometido el agente. ¿Qué hacemos con todo el tiempo que nos tomaba el “hacer”? Mi sospecha es que aparecerá una nueva ventana de oportunidad: un espacio más amplio para evaluar el proceso de inversiones y sus etapas, para sofisticarlo, testearlo, rediseñarlo y mejorarlo constantemente. Ya no estaremos 100% inmersos en la gestión, sino que -

manteniendo un pie dentro y otro fuera-, podremos hacer un meta-análisis y mejorar la forma como se hacen las cosas de forma continua.

Estos dos elementos combinados nos llevan a un tercero. Si los datos son lo más crítico, y estaremos más preocupados del proceso, los que más capacidad tengan de agregar valor en el armado de estas agencias serán **aquellos que más información recopilen respecto a cómo se han hecho las cosas**, y cuáles fueron sus resultados. Esto será la materia prima para entrenar un nuevo modelo, y seguir optimizando el proceso hacia adelante.

El último: si ya nos extendimos y (por lo menos como prueba de concepto) somos capaces de crear una agencia de análisis de inversiones, ¿por qué parar ahí? ¿Qué espacios abre esta nueva tecnología para las labores de implementación y transaccionalidad en las carteras?, ¿para el control de políticas de inversión, reglamentos, mallas de límites, etc.?, ¿para comunicar a banqueros, clientes, reguladores, lo que se ha estado haciendo con los portafolios? **Es muy probable que las posibilidades que vislumbramos hoy incluso se queden cortas frente a lo que los modelos del mañana serán capaces de realizar.**

