¿QUÉ TENDRÍA QUE SER VERDAD PARA DEJAR ACTUAR A UN AGENTE DE IA?

**"La demo prueba la respuesta.
Producción prueba el sistema."
— Diana Paredes**

Por Charles Spencer Buchanan, CPA, CFA.

Basta abrir cualquier red social para toparse con la misma promesa: reemplaza todo con agentes, automatiza tu empresa de punta a punta, toma este curso y aprende a crear agentes en ocho semanas. La publicidad es seductora y el lenguaje, ambicioso. Pero detrás del entusiasmo aparecen dos preguntas que rara vez se responden con seriedad. La primera es casi de diccionario: en la práctica, ¿qué es un agente? La segunda es la que de verdad quita el sueño a quien tiene que firmar la decisión: una vez que creamos uno y lo dejamos actuar solo, ¿qué impide que se convierta en un robot estilo Terminator empeñado en destruir el universo, o, versión menos cinematográfica pero más probable, en un departamento de compras automatizado que ordena 5.000 rollos de papel higiénico mientras nadie mira?

El segundo escenario no es una caricatura. La propia Anthropic, creadora de Claude, dejó a un agente de IA a cargo de una pequeña tienda en sus oficinas durante un mes, el experimento que llamaron Project Vend, con una instrucción simple: generar ganancias. El agente, apodado "Claudius", hacía cosas razonables hasta que un empleado, en broma, le pidió un cubo de tungsteno. A Claudius le encantó la idea y se obsesionó: empezó a comprar "artículos de metal especializados" y llenó el refrigerador de snacks con cubos de metal, para luego rematarlos a pérdida. Terminó perdiendo dinero, inventando una cuenta bancaria inexistente para que le pagaran y hasta sufriendo una "crisis de identidad" en la que insistía en que era una persona de carne y hueso vestida con un blazer azul. Anthropic lo resumió con honestidad: si hoy tuvieran que contratar a alguien para administrar esa máquina, no contratarían a "Claudius".

La anécdota es graciosa, pero el punto es serio y va al corazón de lo que importa para quienes administramos capital, gestionamos riesgos o nos sentamos en un directorio: la distancia entre un agente que impresiona en una demostración y un sistema en el que de verdad se puede confiar para que actúe. Justamente esa distancia fue el centro de mi conversación con Diana Paredes.

Y no podía haber mejor interlocutora para abordarla. Diana es Head of Data Governance & Strategy en Parque Arauco y trae consigo una trayectoria poco habitual: fue Senior IT Architect en BCG Platinion, donde lideró iniciativas de transformación digital y estrategia de datos e IA para clientes globales, y antes pasó varios años en BHP como Principal Lead Data Engineer, diseñando plataformas de datos de extremo a extremo y estableciendo marcos de calidad y gobierno de datos para decisiones ejecutivas. Esa combinación, arquitectura, ingeniería de datos, gobierno y estrategia, le da una mirada inusualmente completa: conoce tanto el "corazón" técnico de un modelo como las reglas de negocio, los controles y las personas que deben rodearlo para que funcione en el mundo real.

Diana planteó en el Data Saturday Santiago en junio 2026 una pregunta que se convirtió en el eje de nuestra conversación: ¿qué tendría que ser verdad para dejar actuar a un agente? Es una pregunta engañosamente simple. Detrás de ella se esconde casi todo lo que hoy separa una demostración impresionante de un sistema en el que una organización puede, de verdad, delegar decisiones. Lo que sigue es un recorrido de mi conversación con Diana.

¿Qué es realmente un "agente"?

Conviene empezar por el vocabulario, porque la palabra "agente" se usa hoy para casi todo. Diana parte de la intuición con que el negocio suele recibirla: la de un compañero o copiloto que ayuda a hacer una tarea más rápido, de forma más eficiente o a mayor escala. Bajo esa idea, alguien que trabaja en finanzas o en inversiones podría tener un agente que lo acompañe en su ámbito, igual que otro lo acompañaría en una estrategia de ventas.

Pero Diana introduce de inmediato un matiz importante. Cuando le pregunté cuántos agentes usa en su propio trabajo, respondió con honestidad que le costaba decirlo, porque distingue entre herramientas y agentes. Ella usa de forma intensiva asistentes basados en lenguaje natural, les llama "gemas" porque trabaja con Gemini, con propósitos distintos: uno la ayuda a revisar presentaciones, otro con temas legales, otro con la estrategia que está armando, otro con sus correos y hasta uno especializado en escribir buenos prompts. Las utiliza para procesar, estructurar y revisar información de forma más eficiente, pero la interpretación final, la validación y la decisión siguen siendo suyas. Por eso, se resiste a llamarlos agentes.

"No les digo que son agentes porque no deciden. Son mis "helpers"."

La distinción es deliberada y tiene consecuencias prácticas: para Diana, lo que convierte a una herramienta en agente no es que procese lenguaje ni que suene convincente, sino que decida y actúe. Mientras solo interpreta, redacta o sugiere bajo supervisión humana, sigue siendo un colaborador. El salto cualitativo, y el problema interesante, aparece cuando le pedimos que cruce esa frontera y opere por su cuenta.

La demo que engaña: capacidad frente a sistema confiable

Una buena parte de la responsabilidad actual de Diana consiste en sentarse frente a demostraciones que otros equipos preparan y ayudar a definir los próximos pasos. Desde ese asiento ha aprendido a desconfiar del efecto que produce una buena demo. La demo, dice, se ve poderosa en su momento: responde rápido, entrega un número, sugiere una decisión de negocio aparente y genera mucha impresión. Justamente ahí empieza lo que ella llama, con cuidado, el engaño.

El riesgo, explica, vive en dos lugares al mismo tiempo: en el modelo de lenguaje y en las personas. Cuando alguien le hace una pregunta al modelo, tiene en su cabeza un contexto que rara vez traslada completo a la pregunta; pregunta asumiendo que el otro entiende lo que quiso decir. El modelo, por su parte, interpreta ese enunciado y responde. No hay garantía de que ambos hayan entendido lo mismo, y por eso una respuesta puede ser "correcta" para la interpretación del modelo y, a la vez, equivocada para la intención de quien preguntó.

Diana lo ilustra con un ejemplo deliberadamente simple. Si uno pregunta "2 por 2" y el modelo responde "cuatro", parece correcto; pero si en realidad lo que hizo fue "2 + 2", el resultado coincide por casualidad y, al cambiar las variables, el sistema entregará algo distinto de lo que se esperaba. Es como preguntar la hora sin dar el contexto del huso horario: el número es correcto, pero la decisión que tomamos con él puede ser errónea. El dato exacto no garantiza la interpretación correcta.

De ahí nace la distinción que, a mi juicio, es la columna vertebral de toda la conversación. Diana separa con claridad la capacidad de un sistema confiable. Una capacidad es poder responder: te entrego un número, un cálculo, una recomendación. Un sistema confiable, en cambio, responde con justificación: te entrego el cálculo junto con una fórmula clara y transparente, con la métrica definida no solo por mí sino acordada por el negocio, con los datos que usé y con trazabilidad y registro de todo el proceso detrás de la respuesta.

"Una capacidad es poder responderte. Un sistema confiable te responde con justificación clara, trazabilidad y registro por detrás."

Para un público financiero la analogía es inmediata. Es la diferencia entre un número y un número auditable. Ningún comité de inversiones aceptaría una recomendación sin entender los supuestos, las fuentes y la metodología; Diana sostiene que con los agentes de IA debe regir el mismo estándar. Un modelo tradicional puede recomendar "por recomendar", porque estadísticamente puede sugerir muchas cosas; un sistema confiable debe recomendar a partir de datos

[1] Meta Grid es un concepto desarrollado por Ole Olesen-Bagneux para describir una arquitectura descentralizada de metadatos, en la que distintos repositorios y sistemas permanecen conectados como parte de una red de contexto empresarial.

actuales, con un contexto y un objetivo actualizados, y dejando registro de bajo qué reglas de negocio hizo esa recomendación.

Por qué la IA parece brillante en unos terrenos y tropieza en otros

Un punto que conversamos a partir de su experiencia en BHP ayuda a entender por qué la IA avanza a velocidades tan distintas según el dominio. En el mundo minero, observa Diana, la parte de los datos se resuelve de manera relativamente más sencilla, porque suele tratarse de sensores y series de tiempo: información semiestructurada proveniente de un camión autónomo, de un movimiento, de una medición. Pero advierte que una cosa es tener el dato del camión y otra muy distinta es capturar todo el ecosistema que ocurre a su alrededor; por eso, incluso con flotas autónomas, muchas veces detrás hay operadores remotos que intervienen cuando se requiere una acción distinta.

Aquí surgió una de las observaciones más reveladoras. ¿Por qué estos sistemas parecen tan inteligentes en ciertos contextos y, sin embargo, hay áreas en las que todavía no nos atrevemos a dejarlos actuar?

Le planteé a Diana una observación cotidiana: esos test que nos piden "seleccionar todas las imágenes que contienen un camión" para demostrar que no somos robots. Lo que para una persona resulta trivial, para una máquina puede ser sorprendentemente complejo. A partir de ello, conversamos sobre esa asimetría, lo fácil para nosotros que resulta difícil para el sistema, y viceversa, precisamente la zona donde aún estamos aprendiendo cuánto podemos confiar en la inteligencia artificial.

La explicación técnica es importante y Diana la traduce sin tecnicismos. Los modelos de lenguaje, dice, generan respuestas a partir de relaciones probabilísticas aprendidas de grandes volúmenes de datos. Pueden interpretar el lenguaje y producir resultados muy convincentes, pero no comparten automáticamente el contexto de la misma manera en que lo hace una persona. Por eso advierte sobre una creencia extendida: que cada nuevo modelo, al aparecer más arriba en los rankings de desempeño, necesitaría menos datos o menos contexto. No es así. El contexto hay que seguir entregándolo igual; un mejor modelo no exime de hacer el trabajo de alimentárselo bien.

El contexto como agua: datos, metadatos y el "Meta Grid"

Si el contexto es tan decisivo, la pregunta obvia es si basta con dárselo al modelo. La respuesta de Diana es matizada: depende de cómo se lo entreguemos. Subir un PDF con un procedimiento, por ejemplo, no garantiza nada por sí solo. El modelo terminará convirtiendo ese documento en números, de modo que el verdadero desafío es preparar la información, con la metadata necesaria, para que quede "bien digerida". Mientras mejor sea ese trabajo previo, menor es la probabilidad de que el sistema alucine. Es lo que en la jerga se llama retrieval-augmented generation (RAG), pero Diana insiste en que la pregunta interesante no es la sigla, sino qué se recupera y con qué estructura.

Para explicar por qué un agente no puede simplemente "lanzarse y listo", Diana recurre a una metáfora que se quedó conmigo. Un agente en producción es como una planta: si la dejo ahí sin cuidarla, se muere; si le sigo echando agua, crece. El agua es el contexto, y ese contexto debe nutrirse de manera constante. El reto no es técnico solamente: es cómo lograr que el contexto se actualice de forma sostenible, sin depender de que alguien lo cargue manualmente cada vez que algo cambia. ¡Lamentablemente, eso contradice la narrativa de que basta con ocho semanas para crear un ejército de agentes de IA que hagan todo el trabajo mientras uno disfruta una margarita frente al mar!

Diana conecta este desafío con el concepto de Meta Grid[1]: una arquitectura que permite articular contexto y metadatos distribuidos entre distintos repositorios y sistemas de una organización. La idea es que los metadatos no vivan en un único lugar, lo que obligaría a una sola herramienta gigante, frágil y difícil de mantener, sino en cada área de la organización, en su propia herramienta. Un agente útil suele cruzar al menos tres áreas, pongamos finanzas, marketing y legal, y cada una debe comprometerse a mantener actualizado su propio contexto: los metadatos de SAP en finanzas, los del sistema de CRM en marketing, las políticas vigentes en legal. Más que herramientas, lo que se requiere es responsabilidad de cada área sobre su parte.

[1] Meta Grid es un concepto desarrollado por Ole Olesen-Bagneux para describir una arquitectura descentralizada de metadatos, en la que distintos repositorios y sistemas permanecen conectados como parte de una red de contexto empresarial.

Diana lo conecta con la forma en que decidimos las personas. Cuando enfrentamos una decisión importante, vamos a marketing a pedir contexto, luego a legal, luego a tecnología, y solo entonces decidimos. Un agente robusto necesita algo equivalente: escuchar a varias fuentes de metadatos, actualizarse, recomendar y, si corresponde, escalar la decisión a un humano en lugar de ejecutarla por su cuenta. La capacidad de buscar información en distintos contextos, y no en una sola caja cerrada, es parte de lo que hace fiable a un sistema.

El ingeniero de IA y la arquitectura que no se ve

Hablamos también de las personas que construyen estos sistemas, en parte a propósito de las cifras millonarias con que algunas empresas tecnológicas fichan a ingenieros de IA estrella. ¿Qué hace, en concreto, este perfil? Diana responde que depende mucho de la madurez de la organización, pero que los obstáculos aparecen en un orden bastante predecible: primero el dato, después la arquitectura. Aunque el dato esté en orden, queda por resolver dónde se ejecuta y se aloja el modelo, qué plataforma se usa, cómo se registran los agentes, dónde y cómo se almacena la base de conocimiento.

Una observación suya, surgida de entrevistar a estos profesionales, resulta contraintuitiva. El auge de los modelos de lenguaje generó tal demanda que muchos ingenieros se volcaron a la parte "más sexy", el procesamiento de lenguaje natural, sin profundizar en los modelos estadísticos tradicionales. Como resultado, hoy existen perfiles muy especializados en IA generativa que no siempre cuentan con la misma profundidad en modelamiento estadístico, ingeniería de datos o conocimiento de industria. Los científicos de datos con formación más tradicional aportaban una virtud que sigue siendo crítica: se acercaban al negocio, entendían la industria y sabían traducir las reglas de negocio al modelo. Los sistemas productivos requieren integrar esas capacidades, no reemplazar unas por otras.

El punto de fondo es que la complejidad se desplazó hacia la arquitectura. El ingeniero de IA debe articular con gobierno de datos, con ciberseguridad, con negocio y con la arquitectura tecnológica tradicional para decidir, por ejemplo, si todo corre en la nube o localmente. Y debe sumar componentes que rara vez se ven en una demo pero que sostienen la confianza: el registro de agentes, la base de conocimiento vectorizada, el almacenamiento de cada pregunta y respuesta con su trazabilidad, los guardrails que acotan el comportamiento, las reglas de decisión y los "loops" de aprendizaje que re-inyectan contexto para ir afinando el sistema. Nada de eso brilla en una presentación; todo eso es lo que distingue a un experimento de un producto.

Tres preguntas antes de soltar al agente

Le planteé a Diana un ejercicio: si pudiera hacer solo tres preguntas antes de poner un agente a operar, ¿cuáles serían? Le costó acotarlo a tres, pero el resultado es una excelente lista de verificación para cualquier directivo.

La primera apunta a la madurez del dato: ¿qué tan listo está el dato en la organización para este caso de uso concreto? Es la base sobre la que se sostiene todo lo demás. La segunda se desplaza de los datos a las personas: ¿saben los equipos de negocio y de tecnología qué es la IA y qué es un uso ético de la IA? Sin ese entendimiento compartido, el modelo operativo que debe sostener al agente se vuelve frágil. La tercera es la dimensión legal: ¿tiene la empresa una política interna sobre el uso de la IA, o no?

Esta tercera pregunta dio paso a uno de los momentos más francos de la conversación. Diana contó una historia de una empresa donde llegó a usarse un algoritmo para calcular el bono de las personas, y reconoció que aquello no le gustó: genera demasiada duda. Sin una política interna, advierte, podríamos terminar construyendo algo que para la organización no está bien, un "recomendador" que decida despidos o que defina bonos. Y aun cuando la pregunta legal no tenga respuesta, hacerla sirve para estimar el esfuerzo regulatorio y de cumplimiento que habrá que asumir.

Para Diana, además, no basta con que el sistema exista: las áreas afectadas deben entenderlo. Si se introduce un cálculo de bonos asistido por IA, Recursos Humanos debe tener al menos la capacitación básica para responder cuando un empleado pregunte por qué se está usando IA. En una organización grande, un banco con miles de empleados, por ejemplo, eso exige un grupo especializado que comprenda el caso antes

[1] Meta Grid es un concepto desarrollado por Ole Olesen-Bagneux para describir una arquitectura descentralizada de metadatos, en la que distintos repositorios y sistemas permanecen conectados como parte de una red de contexto empresarial.

de implementarlo. “Tomamos un modelo open source y lo ponemos, ya está” no es una estrategia aceptable.

“Qué tendría que ser verdad para dejar actuar a un agente de IA?”

Cerré la entrevista volviendo a su pregunta original: después de preparar la charla y conversar con más personas, ¿había cambiado su respuesta sobre qué tendría que ser verdad para dejar actuar a un agente? Diana reconoció que su respuesta evolucionó, “el contexto cambia”, dijo, fiel a su propio argumento, y que con más información se volvió, paradójicamente, más simple. Antes pensaba en términos de arquitectura: necesitamos un Meta Grid, metadatos que fluyan, controles. Hoy la respuesta, para ella, está en el riesgo.

“¿Qué tendría que ser verdad? Para mí se convirtió en otra pregunta: ¿qué riesgo quiere asumir la compañía?”

El giro es sutil pero profundo. Si para lanzar un agente nos harían falta las condiciones A, B y C, pero solo logramos A y B, ¿se puede lanzar igual? La respuesta, dice Diana, depende del riesgo que la empresa esté dispuesta a aceptar. Puede que sí, incluso porque a la organización le resulte atractivo hacerlo; pero entonces debe asumir ese riesgo de manera consciente y preparar su “bolsa de riesgo”. Lo que no es opcional es articular ese riesgo con claridad hacia quienes deciden. La condición no es la perfección técnica, sino que los decisores entiendan en detalle el riesgo que están aceptando. Después de cumplir los mínimos no negociables en materia legal, ética, privacidad, seguridad y gobierno, la organización debe decidir qué nivel de riesgo residual está dispuesta a aceptar. La condición no es alcanzar una perfección técnica imposible, sino asegurar que los decisores comprendan claramente el riesgo que autorizan y los controles existentes.

De ahí se desprende una recomendación de gobierno corporativo que los miembros de CFA Society Chile reconocerán de inmediato. Así como hoy existen comités de datos, Diana sostiene que debería existir un comité de IA que apruebe los casos de uso, con presencia de legal y ética, de negocio y de las áreas relevantes, capaz de decidir cuándo vale la pena asumir el riesgo. Incluso plantea que en los directorios podría hacer falta una mirada especializada en IA, no solo el CTO (Director de Tecnología), que sepa traducir el nivel de riesgo y lo que podría salir mal. Su advertencia es directa: si quien decide no entiende el riesgo, no debería autorizar la compra ni el despliegue.

Y el conocimiento, subraya, es lo que habilita esa decisión. Si el comité va a decidir pero quien presenta el caso no expone todas las aristas, o si sus integrantes no son capaces de pedir más información y cuestionar, la presentación se convierte en una caja negra y la decisión pierde sentido.

“Por eso insiste en que los gerentes se capaciten en inteligencia artificial: no para volverse ingenieros, sino para comprender los riesgos, la ética y lo que está realmente en juego.”

Más allá de la inteligencia

La reflexión final de Diana ofrece una brújula honesta para este dilema. Confiar en un sistema autónomo no es un acto de fe en la tecnología, sino una decisión de riesgo informada, tomada por personas que entienden lo que están autorizando y que responden por sus consecuencias. La autonomía no se concede de una vez; se gana gradualmente, sosteniendo el contexto como quien cuida una planta, y se limita deliberadamente allí donde lo que está en juego son las personas.

Con esta conversación inauguramos la serie. En las próximas ediciones seguiremos dialogando con profesionales, investigadores, ingenieros y líderes de negocio que están construyendo, regulando y cuestionando estas tecnologías.

Si algo nos deja Diana Paredes como punto de partida, es que la pregunta verdaderamente importante no es si la inteligencia artificial puede generar una respuesta, sino en qué condiciones podemos confiar en que actúe.

[1] Meta Grid es un concepto desarrollado por Ole Olesen-Bagneux para describir una arquitectura descentralizada de metadatos, en la que distintos repositorios y sistemas permanecen conectados como parte de una red de contexto empresarial.