isnos ai

CÓMO OPERAMOS · MODELOS MENTALES

Los modelos con los que decidimos qué construir, cuándo matar una idea y cómo sabemos que algo funcionó

Una guía corta de cómo pensamos y manejamos el trabajo con datos e IA. No son teorías: son los filtros que usamos todas las semanas con los equipos y los líderes de negocio con los que trabajamos.

Esfuerzo → Entregable → Resultado → **Impacto**

LA PREGUNTA DETRÁS DE TODO

Si esto funciona, ¿qué cambia de verdad?

isnos ai
IA aplicada · diagnóstico de 14 días

isnosai.com · hola@isnosai.com
Noviembre de 2026

PUNTO DE PARTIDA

El problema que estos modelos resuelven

La mayoría de iniciativas de IA no fallan porque la tecnología no funcione. Fallan porque el equipo optimiza el desempeño del modelo y no el resultado del negocio. La herramienta funciona y nadie la usa.

Construir algo impresionante ya es barato. Lo caro es saber qué pedir, saber si la respuesta está bien y lograr que la gente de verdad lo use. Por eso trabajamos con un conjunto pequeño de modelos mentales y los usamos siempre igual. Nos ayudan a parar antes, a no optimizar lo que no importa y a poner el esfuerzo donde devuelve valor.

Dónde entra cada modelo

Cada etapa tiene una pregunta que debe quedar resuelta antes de pasar a la siguiente.

ETAPA 1 Elegir el problema	ETAPA 2 Probar el valor	ETAPA 3 Construir	ETAPA 4 Lograr que se use	ETAPA 5 Medir y decidir
¿Qué cambia si esto funciona?	*¿Importa, o solo es posible?*	*¿Qué es lo mínimo que sirve?*	*¿La gente cambió cómo trabaja?*	*¿Escalar, ajustar o cerrar?*
00 Producto de datos	02 Proof of Value, no Proof of Concept	05 El carpintero	09 Curva J de adopción	01 Esfuerzo → Impacto
04 Primeros principios	03 Semana 2: validación o se mata	11 Escalera de autonomía	07 El silencio de los usuarios	01 Gradiente de medición
06 Dolor del jefe ≠ dolor del usuario		12 Carril verde por riesgo		08 Optimizar o reimaginar
10 Matriz de tareas para IA				15 Ofensivo o defensivo

Contenido

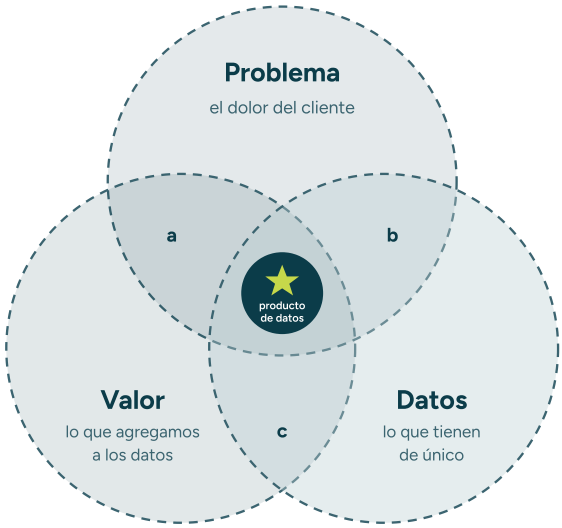
00 Productos de datos 03
01 Esfuerzo → Entregable → Resultado → Impacto 04
02 Proof of Value, no Proof of Concept 05
03 Semana 2: validación o se mata 05
04 Primeros principios 06
05 El carpintero 06
06 Dolor del jefe ≠ dolor del usuario 07
07 El silencio de los usuarios 07
08 Optimizar o reimaginar 08
09 Curva J de adopción 08
10 Matriz de tareas para IA 09
11 Escalera de autonomía 10
12 Carril verde por niveles de riesgo 10
13 Cómo trabajamos con tu equipo 11
14 Dependencias que no controlamos 11
15 Proyectos ofensivos y defensivos 12
·· El flujo completo de decisión 13
·· Tres casos: uno que funcionó, dos que no 14
·· Cómo trabajamos y las preguntas que siempre hacemos 15

00 · QUÉ CONSTRUIMOS

Productos de datos: donde se cruzan problema, datos y valor

DATA PRODUCTS

Lo que construimos son productos de datos. Un producto de datos solo existe donde se cruzan tres cosas: un dolor real del cliente, algo único en los datos y un valor que les agregamos. Si falta una, no es un producto. Es un reporte, una herramienta que cualquiera copia o una solución buscando problema.



Si falta un círculo

a **Problema y valor, sin datos únicos.** Funciona, pero cualquiera lo copia mañana con la misma herramienta. No hay ventaja que defender.

b **Problema y datos, sin valor agregado.** Un tablero o un reporte. Muestra el dolor, pero no cambia ninguna decisión.

c **Datos y valor, sin problema.** El modelo brillante que nadie pidió. Una solución buscando problema.

> Este modelo define **qué** construir. La cadena Esfuerzo → Impacto, en la página siguiente, mide **si funcionó**.

PROBLEMA

¿Qué dolor resuelve, y de quién?

Un usuario con nombre y un dolor que se puede ver en su día de trabajo. No "mejores insights".

DATOS

¿Qué tienen de único?

Datos propios, históricos o de comportamiento que la competencia no tiene. Ahí está la ventaja difícil de copiar.

VALOR

¿Qué les agregamos?

Una predicción, una prioridad, una alerta o una respuesta, puesta dentro del flujo de quien decide.

Un ejemplo: un score de riesgo de churn por cuenta

- **Problema:** el equipo de cuentas se enteraba tarde de que un cliente se iba, cuando ya no había nada que hacer.
- **Datos:** facturación, actividad en llamadas e historial de servicio de cada cliente. Solo la empresa los tiene juntos.
- **Valor:** un score semanal con alertas que dice a quién llamar primero, y por qué.

LA PREGUNTA **¿Están los tres círculos, o solo dos?**

01 · EL MODELO CENTRAL

Esfuerzo → Entregable → Resultado → Impacto

EFFORT → OUTPUT → OUTCOME → IMPACT

Todo lo que hacemos pasa por esta cadena. El error más común es dejar de contar antes de tiempo: celebrar el esfuerzo o el entregable y nunca rendir cuentas por el impacto.

EFFORT
Esfuerzo
Lo que hacemos. Planear, limpiar datos, programar, entrenar.

→

OUTPUT
Entregable
Lo que lanzamos. Un modelo, un bot, un tablero, el código.

✕

OUTCOME
Resultado
Lo que el usuario cambia. Una decisión, un hábito, un proceso.

→

IMPACT
Impacto
Lo que gana el negocio. Ingresos, horas, retención, menos churn.

AQUÍ SE ROMPE CASI SIEMPRE: LANZAR NO ES ADOPTAR

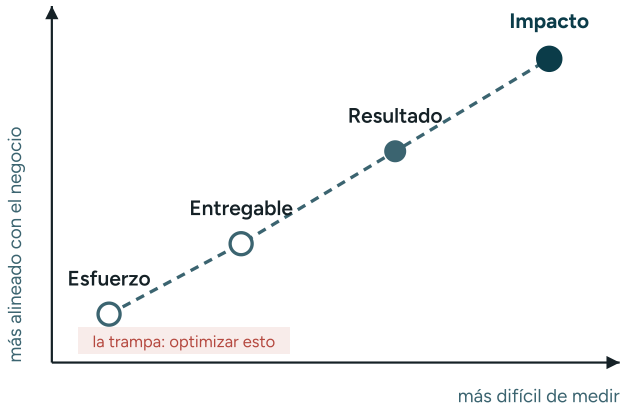
Lo que el modelo deja ver

- **Desperdicio.** Mucho esfuerzo con poco impacto es un fracaso, aunque el modelo sea excelente. Poco esfuerzo con mucho impacto es excelencia.
- **Secuencia.** No se puede optimizar el impacto sin validar el resultado, y no se puede suponer el resultado si los usuarios no cambiaron su comportamiento.
- **Dónde estamos.** Cada iniciativa tiene un punto en la cadena. Lo decimos en voz alta en cada actualización: "estamos en entregable, el resultado aún no está probado".

Cómo lo aplicamos

- Definimos cada iniciativa en términos de resultado o impacto, nunca de entregable. "Un modelo de churn" no es una meta. "Que el equipo de cuentas cambie a quién llama primero" sí.
- Nuestros reportes semanales abren con adopción y resultado, no con lo que se terminó.
- Lo combinamos con RICE para priorizar: buscamos lo de poco esfuerzo y mucho impacto.

El gradiente de medición



Mientras más a la derecha, más alineada está la métrica con el negocio y más difícil es medirla, sobre todo al comienzo. Por eso los equipos optimizan lo fácil: precisión, latencia, costo por consulta. Son métricas de calidad del entregable, no de valor.

> **Para medir un equipo** usamos las cuatro capas a la vez. Solo esfuerzo y entregable premia la cantidad. Solo impacto premia los atajos y rompe la colaboración. El balance es medir las dos puntas.

LA PREGUNTA **¿En qué punto de la cadena estamos midiendo el éxito, y por qué no más a la derecha?**

02 · PROBAR EL VALOR

Proof of Value, no Proof of Concept

Para un buen equipo técnico, que algo sea factible casi nunca es el riesgo real. El riesgo es construir algo que funciona y no le importa a nadie.

	Proof of Concept	Proof of Value
Pregunta	¿Es técnicamente posible?	**Si funciona, ¿importa?**
Foco	Construir la capacidad	Cambiar una decisión o un proceso
Éxito	Precisión del modelo, demo que impresiona	Adopción, confianza, cambio de comportamiento
Cuándo se mata	Cuando no se puede construir	Cuando no cambia nada, aunque se pueda construir

Cómo probamos el valor

- Miramos a los usuarios sufrir con el proceso actual.
- Lanzamos una solución al 80% a un solo grupo.
- Medimos cambio de comportamiento, no satisfacción.
- Calculamos el impacto en el negocio.
- Solo entonces se invierte en escalar.

Matar una idea temprano por falta de valor no es fracaso. Es disciplina.
Aquí mueren muchas ideas técnicamente brillantes, y está bien que mueran.

03 · PROBAR EL VALOR

Semana 2: validación con usuarios reales o se mata

Si en la segunda semana no hay gente real usándolo para trabajo real, paramos de construir. Nada de lo que hacemos pasa más de dos semanas sin que un usuario real lo toque.

Cuenta como validación

- Eligen la herramienta nueva sobre su flujo viejo.
- Uso y preguntas sin que nadie los empuje.
- "¿Cómo hago para tenerlo en mi equipo?"
- Un cambio de comportamiento que se puede medir.

No cuenta

- "Se ve muy bacano." "¡Qué buen demo!"
- "Deberíamos usar esto algún día."
- Uso forzado o pedido por el jefe.
- Satisfacción declarada sin uso real.

El calendario de decisión



Semana 0
la oportunidad

Semana 2
usuarios reales, trabajo real

Mes 1
30% de adopción o se pivota

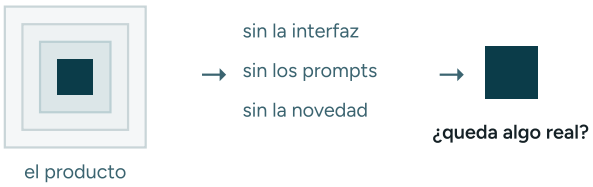
Trimestre 1
impacto medible o se apaga

Criterios para matar. No hay 3 usuarios con nombre al final de la semana 1. No hay uso ni preguntas en la semana 2. Cuando pueden escoger, vuelven al flujo viejo. Menos de 30% de adopción en el mes 1. Los definimos antes de arrancar, no después.

04 · ELEGIR EL PROBLEMA

Primeros principios: la prueba del toro

Picasso dibujó un toro una y otra vez, cada versión más simple, hasta dejar unas pocas líneas. Seguía siendo un toro. Esa es la prueba que le ponemos a cualquier producto de IA.



Lo que queda tiene que seguir cambiando una decisión, un comportamiento o un resultado. Si no, era decoración.

Nuestros no negociables

- **El impacto en el negocio es la unidad de valor.** Si no cambia una decisión, un hábito o un resultado, no vale.
- **Precisión sin uso es ruido.** Un modelo que nadie usa ni confía es un experimento, no un producto.
- **El entusiasmo interno no es valor externo.** Los demos no pagan facturas. Los resultados sí.

LA PREGUNTA **¿Qué supuesto, si resulta falso, mata toda la iniciativa? La mayoría de fracasos vienen de ahí, no de mal código.**

05 · CONSTRUIR

El carpintero: la herramienta correcta, no la más impresionante

Un carpintero escoge el martillo porque sirve, no porque impresione. Los productos de IA no se tratan de usar el modelo de punta. Se tratan de usar lo que resuelve el trabajo y lo que la gente va a adoptar.

No se trata de

- La herramienta más nueva o más de moda.
- La máxima sofisticación técnica.
- Una arquitectura que se vea bien en una presentación.

Se trata de

- Lo que resuelve el trabajo con el menor costo.
- Lo que los usuarios van a usar de verdad.
- Definir qué es "suficientemente bueno" para este caso.

Ejemplos reales

Bot de preguntas internas. Arrancó con cerca de 80% de precisión y salió así. Nunca llegó al 95% que parecía necesario, y no hizo falta: lo que subió fue la confianza y el uso.

Predicción de churn. Un modelo simple con buenas variables le gana a un modelo complejo con variables pobres.

La regla de calidad tiene dos lados. Para uso interno, 80% puede bastar. Para algo que ve el cliente, la calidad es binaria: "¿lo mandaría a su cliente más grande sin revisarlo?". Si no, no está listo.

06 · ELEGIR EL PROBLEMA

El dolor del jefe no es el dolor del usuario

Que un gerente diga que su equipo es ineficiente no significa que el equipo quiera la solución que el gerente imagina. Quien pide la herramienta muchas veces no es quien la va a usar todos los días.



El gerente ve síntomas, no la causa. La solución queda optimizada para la visibilidad del jefe, no para el flujo de quien hace el trabajo.

Cómo lo evitamos

- No escribimos código hasta ver a 3 o más usuarios hacer el flujo completo, de principio a fin.
- Miramos, no solo preguntamos. Tomamos el tiempo de cada paso y separamos los cuellos de botella reales de los percibidos.
- En cada proyecto dejamos escrito quién lo pidió y quién lo usa. Son dos campos distintos.
- Construimos para el flujo del usuario, no para el tablero del gerente.

Bandera roja. Cuando quien pide no es quien usa, triplicamos la validación.

Excepción: si es un requisito regulatorio, o si quien lo pide también es el usuario principal y pone recursos para la adopción.

07 · LOGRAR QUE SE USE

El silencio de los usuarios está gritando

Si resolvimos un problema "crítico" y nadie lo usa, el problema era otro o la herramienta creó una fricción nueva. Poca adopción no es un tema de comunicación. Es una señal de producto.

Qué hacemos cuando hay silencio

- Dejamos de construir funcionalidades y nos sentamos a ver a la gente trabajar.
- Entrevistamos a 3 personas que no lo adoptaron: "¿qué tendría que cambiar para que lo use?"
- Buscamos el momento exacto en el que quieren volver al proceso manual.

Antes de construir, nos obligamos a responder

- ¿Qué tendría que ser cierto para que el 80% de los usuarios lo adopte en el mes 1?
- ¿En qué momento exacto de su día lo van a usar?
- ¿Se integra a su proceso actual o les pide cambiar todo?

Si no podemos responder con confianza, todavía no construimos.

08 · MEDIR Y DECIDIR

Optimizar o reimaginar

Optimizar supone que el sistema está bien y pregunta cómo mejorarlo. Si el flujo está mal o la métrica está mal, optimizar solo acelera el fracaso.

Optimizamos cuando

- El resultado es el correcto, la eficiencia no.
- Los usuarios lo quieren, pero es lento o caro.
- La base es sólida y hay un camino claro de mejora.

¿El resultado que buscamos es el correcto?

← SÍ NO →

Reimaginamos cuando

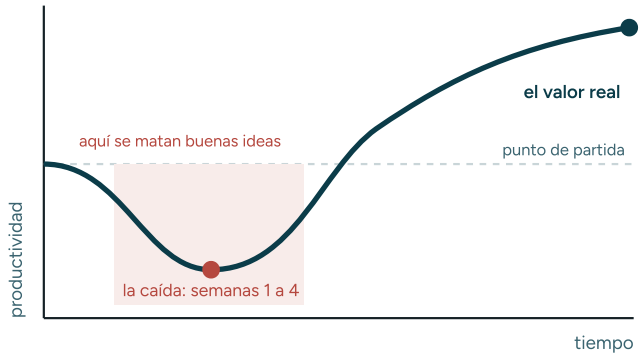
- El resultado en sí está mal.
- No lo adoptan aunque tenga buena calidad.
- Estamos resolviendo el problema equivocado o el flujo está roto de base.

La trampa: casi todos los equipos optimizan demasiado temprano porque se siente productivo y las métricas suben. Reimaginar se siente más lento y más riesgoso, pero muchas veces es el único camino al impacto. Nuestra regla: solo optimizamos después de validar el resultado. La infraestructura robusta llega después de la adopción, no antes.

09 · LOGRAR QUE SE USE

La curva J de adopción

Adoptar IA casi siempre empeora las cosas antes de mejorarlas. La herramienta nueva agrega carga mental, cambia la memoria muscular e introduce riesgo. Los equipos que ganan no evitan la caída: la esperan y la planean.



Cómo lo manejamos

- **Lo anunciamos antes.** Les decimos a los líderes que las primeras semanas van a verse peor, para que no confundan la caída con un fracaso.
- **Lo planeamos.** Entrenamiento, acompañamiento y un responsable claro dentro del flujo. Lanzar el modelo no es el final del trabajo.
- **Rediseñamos el flujo** para que la caída sea menos profunda y más corta.
- **Miramos la señal correcta.** Al comienzo no preguntamos si las métricas ya subieron. Preguntamos si el comportamiento está cambiando.

10 · ELEGIR EL PROBLEMA

Matriz de tareas para IA

Antes de decidir cómo usar IA en una tarea, la ubicamos en dos ejes: qué tipo de conocimiento exige y cuánto cuesta equivocarse. El cuadrante define cuánto control humano necesita.

Basada en el playbook de IA generativa de Anand y Wu (HBR, noviembre de 2025), adaptada a cómo la usamos.

	CONOCIMIENTO EXPLÍCITO (REGLAS, ESTRUCTURA)	CONOCIMIENTO TÁCITO (CRITERIO, CONTEXTO, CREATIVIDAD)
ERROR BARATO	**Zona sin arrepentimientos** Automatizar con decisión. La pregunta: ¿el ahorro compensa una pequeña pérdida de calidad? ¿Y qué podemos hacer ahora que antes era muy caro? Resúmenes, triage de tickets, reembolsos pequeños, filtrar hojas de vida.	**Catalizador creativo** La IA amplía la capacidad creativa. La pregunta: ¿le ahorra tiempo a los creativos y deja participar a quienes no lo son? Lluvia de ideas, primeros borradores, exploración de diseño.
ERROR CARO	**Control de calidad** La IA redacta, una persona verifica. La pregunta: ¿qué partes del flujo se pueden delegar y cuáles exigen un experto? Borradores legales, reportes financieros, documentos de cumplimiento.	**Primero la persona** La IA asiste, la persona decide. Nunca al revés. La pregunta: ¿qué tareas puede apoyar la IA para que el criterio humano sea mejor? Estrategia, contratación de cargos clave, crisis, temas sensibles de personas.

Revisar le gana a hacer

La regla más simple que conocemos para saber dónde usar IA: úsela donde revisar la respuesta sea más rápido que hacer el trabajo. La respuesta a un correo que se revisa en un minuto: sí. Una respuesta legal que nadie puede verificar: no.

La IA predice, no sabe. Por eso puede sonar muy segura y estar equivocada.

¿Quién lo revisa? La prueba del humo

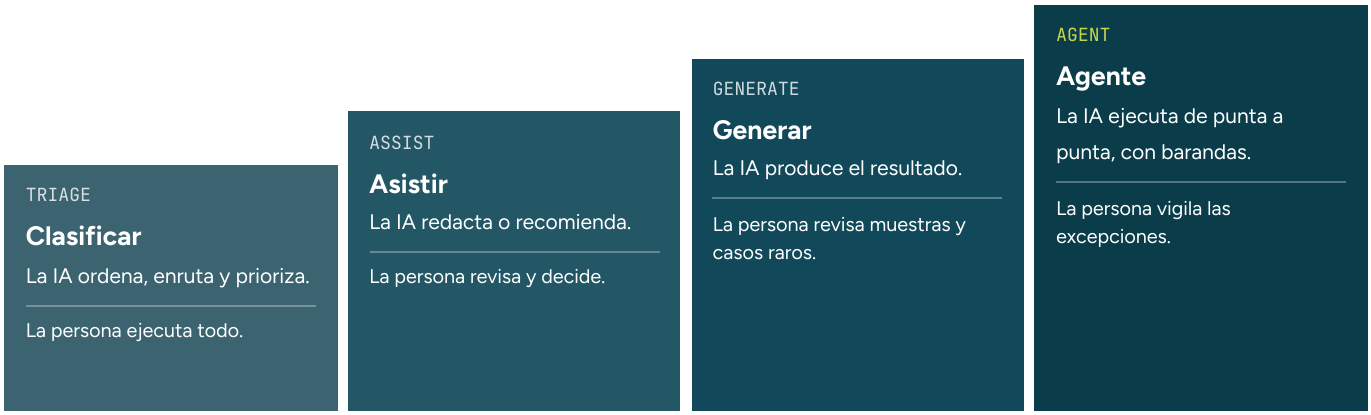
Harvard y BCG lo probaron con 758 consultores. En tareas donde la IA es fuerte, el trabajo salió cerca de un tercio mejor. En una tarea donde es débil, se equivocaron más, y las respuestas equivocadas sonaban más convincentes.

> Si la respuesta a "¿quién lo revisa?" es nadie, y equivocarse es caro, es humo o un riesgo que nadie le puso precio.

11 · CONSTRUIR

Escalera de autonomía: ganarse el derecho a subir

Toda iniciativa arranca abajo, en clasificar o asistir. Solo sube un escalón cuando la adopción y la confianza ya están probadas. Saltar directo a un agente autónomo es la forma más rápida de perder la confianza del usuario.



Un agente es IA que usa herramientas en un ciclo para lograr un objetivo. Hoy funciona bien en trabajos acotados, con una persona revisando.

12 · CONSTRUIR

Carril verde: aprobar por nivel de riesgo, no caso por caso

En empresas grandes, lo que más frena la IA no es la regulación, es la arquitectura de decisión: el mismo proceso largo de aprobación para un caso sin datos sensibles que para uno con datos de clientes. Proponemos acordar los niveles de riesgo una sola vez con cumplimiento y seguridad, y que cada caso entre al carril que le toca.

Nivel	Criterio	Ejemplos	Aprobación
Verde	Sin datos personales ni de clientes	Comunicados, resúmenes, formación	**Registro, el mismo día**
Amarillo	Datos internos o de empleados	Conciliaciones, reportes, onboarding	**Exprés, días**
Rojo	Datos de clientes o decisiones automatizadas	Casos que tocan clientes o sus contratos	**La ruta completa de siempre**

- Cada caso tiene un dueño, una línea base y una decisión a los 30, 60 y 90 días: escalar, ajustar o cerrar.
- Cumplimiento, legal y tecnología entran al equipo desde la primera semana, no al final.
- Medimos uso y horas liberadas, no licencias compradas.

13 · TRABAJAR CON TU EQUIPO

Cómo trabajamos con tu equipo: tres pilares

La IA solo rinde si la gente que la usa está motivada, entiende para qué sirve y ve el impacto de su trabajo. No funcionamos como una fábrica de software que solo administra un backlog. Por eso separamos tres trabajos distintos.

EL QUÉ

Producto

Decidir prioridades contigo. Enfocar el esfuerzo en lo que más impacto genera y decir que no a lo demás.

EL CÓMO

Proyecto

Cómo se entrega: tiempos, recursos, trade offs, y cuándo una solución ya es suficientemente buena.

EL QUIÉN

Personas

La gente que hace el trabajo, en el centro. Acompañamiento, entrenamiento y feedback directo, con Radical Candor.

Cómo alineamos

- Cada iniciativa se conecta con un objetivo del negocio: empresa → área → equipo → cada persona.
- Una métrica norte clara, acordada contigo desde el inicio.
- El éxito se mide por la adopción y el impacto, no por la cantidad de código.

Cómo mantenemos la innovación

- Pruebas de valor cortas y frecuentes, para que el equipo experimente sin arriesgar mucho.
- Lo que funciona se escala a solución completa. Lo que no, se cierra.
- Sistemas reutilizables por encima de soluciones de una sola vez.
- Todo documentado: decisiones, datos y procesos fáciles de encontrar, para que el conocimiento quede en tu equipo.

14 · CONSTRUIR

Dependencias que no controlamos

Si el éxito de un proyecto depende de lo que entreguen otros equipos, gestionar esas dependencias es gestionar el proyecto. Lo aprendimos con una iniciativa grande que dependía de bases de datos que nuestro equipo no controlaba.

- Al arrancar, mapeamos cada equipo o sistema externo del que depende la entrega, con un dueño por dependencia.
- Acuerdos explícitos: fechas, revisión semanal y ruta de escalamiento. Si una dependencia crítica se atrasa más de una semana, se escala.
- Si hay un tercero que aprueba, sus criterios de aceptación se definen por escrito con él. Sin criterios claros, no nos comprometemos a entregar.

Lo que aprendimos. Que el socio externo quede contento no significa que la empresa lo vaya a usar. Si la satisfacción del socio y la adopción interna empiezan a separarse, paramos y realineamos. Y cuando hay plata o contratos externos de por medio, el nivel de gobierno del proyecto sube por defecto.

15 · MEDIR Y DECIDIR

Proyectos ofensivos y defensivos

OFFENSIVE · DEFENSIVE (ENABLERS)

No todos los proyectos se miden igual. Desde el primer día clasificamos cada iniciativa como ofensiva o defensiva, porque eso cambia cómo la justificamos, cómo la medimos y cómo la priorizamos.

OFENSIVOS · MUEVEN LA AGUJA

Impacto directo y atribuible

Generan o protegen ingresos de forma que se puede atribuir: más revenue, menos churn, más expansión, horas ahorradas convertidas en plata.

- Se justifican con un caso de negocio y su retorno.
- Se miden con la cadena completa, hasta el impacto en el P&L.
- Ejemplos: score de churn, predicción de expansión, el bot que libera horas.

¿Cuánto gana o ahorra el negocio, y en cuánto tiempo?

DEFENSIVOS · ENABLERS

Necesarios, difíciles de atribuir

El negocio no puede operar sin ellos, pero es difícil atribuirles un impacto directo: cumplimiento, regulación, seguridad, gobierno y calidad de datos, infraestructura.

- Se justifican con el costo de no hacerlos: multas, incidentes, proyectos bloqueados.
- Se miden por el riesgo que evitan y por lo que habilitan.
- Ejemplos: carril verde de riesgo, protección de datos personales, pipelines, observabilidad de modelos.

¿Qué riesgo evita, o qué hace posible que sin esto no existiría?

Cómo cambia la gestión

	Ofensivos	Defensivos (enablers)
Métricas	Revenue, retención, churn, expansión, horas en US$	Riesgo evitado, incidentes, tiempo de aprobación, cobertura, proyectos ofensivos que destrabó
Alcance	Empezar chico, probar valor y escalar lo que funciona	El mínimo que cumple la norma o el criterio de aceptación, no el ideal
Error común	Medir solo el entregable y declarar éxito	Pedirle ROI directo y matarlo, o dejarlo crecer sin límite porque "es obligatorio"

Un portafolio sano tiene los dos. Solo ofensivos: se crece sobre una base que se cae. Solo defensivos: el equipo se vuelve un centro de costo y pierde la conversación con el negocio.

Lo defensivo bien hecho habilita lo ofensivo. Por eso reportamos cada enabler junto a los proyectos ofensivos que destraba. Así su valor se vuelve visible sin inventarle un ROI.

LA PREGUNTA **¿Este proyecto es ofensivo o defensivo, y lo estamos midiendo como tal?**

TODO JUNTO

El flujo completo de decisión

Así pasa una solicitud o una oportunidad por los modelos, de la idea a escalar o cerrar.

1	**Primeros principios** ¿Qué problema estamos resolviendo de verdad? ¿Quién lo pide y quién lo usa? ¿Están el problema, los datos y el valor?	modelos 00 · 04 · 06
2	**Prueba de valor** Si esto funciona, ¿qué cambia? ¿En qué cuadrante de riesgo está?	modelos 02 · 10
?	**¿Podemos ver a 3 o más usuarios sufrir con el proceso actual?**	sí → sigue no → no se construye
3	**Forzar la semana 2** ¿Cómo probamos valor en dos semanas? ¿Quiénes son los primeros usuarios comprometidos?	modelo 03
4	**Criterios para matar** ¿Qué nos haría parar? Se escriben antes de arrancar.	modelo 03
5	**Construir lo mínimo que sirve** La herramienta correcta, el escalón más bajo de autonomía, el carril de riesgo que le toca.	modelos 05 · 11 · 12
6	**Acompañar la adopción** Esperar la curva J, escuchar el silencio, rediseñar el flujo.	modelos 07 · 09
7	**Medir el resultado, no el entregable** ¿Cambió el comportamiento? ¿Dónde estamos en la cadena? Si es defensivo: ¿qué riesgo evitó y qué habilitó?	modelos 01 · 15
?	**¿Hay impacto?** Si el resultado es el correcto, optimizamos. Si no, reimaginamos o cerramos.	sí → escalar y optimizar no → reimaginar o cerrar

Checklist antes de aprobar una iniciativa

- [ ] Sabemos en qué cuadrante de riesgo está y cuánto error se tolera.
- [ ] Arranca en clasificar o asistir, no salta a agente.
- [ ] Podemos nombrar 3 personas que lo van a usar cada semana.
- [ ] El resultado está definido como una decisión o un hábito que cambia, no como "mejores insights".
- [ ] Hay criterios para matarlo antes de la semana 4.
- [ ] Sabemos en qué punto de la cadena estamos y qué tiene que ser cierto para avanzar.

LOS MODELOS EN LA PRÁCTICA

Un caso que funcionó, dos que no

Funcionó · Un bot para las preguntas que se repetían

Los equipos se ahogaban en mensajes de chat con las mismas preguntas una y otra vez: resultados de clientes, metodología de medición, políticas internas. Una respuesta podía tardar casi una hora, no por falta de ganas sino porque la gente estaba haciendo su trabajo.

El camino tentador era construir el mejor sistema de preguntas y respuestas posible y llevar la precisión a 95%. En cambio preguntamos qué cambiaría si respondíamos al instante. Hicimos un sprint de cuatro semanas: un canal, un bot, cero pulido, cerca de 80% de precisión. La hipótesis no era sobre el modelo, era sobre el flujo de trabajo.

La adopción arrancó dispareja, la curva J. Pasamos tres semanas sentados con los equipos que no lo usaban y aprendimos que el valor no era responder rápido: era quitarle el cambio de contexto a quienes antes respondían a mano. Rediseñamos cómo encajaba el bot en su día, y la adopción cambió. Solo después de eso invertimos tres sprints en los pipelines de datos.

De una hora	Varios equipos	Cientos	Cientos de horas
a segundos por respuesta	usándolo cada día	de preguntas respondidas	liberadas al año

La precisión nunca llegó al 95% que parecía necesario, y no hizo falta. Lo que mejoró fue la confianza, la adopción y la integración al trabajo diario. Modelos en juego: 02, 03, 05, 07, 08, 09.

No funcionó · Una herramienta de forecasting

Funcionaba bien técnicamente y casi nadie la usó, así que la apagamos. Fue un entregable sin resultado. En el siguiente proyecto pusimos usuarios reales en la semana dos y cortamos lo que ignoraban. De ahí sale nuestra regla: nada pasa más de dos semanas sin que un usuario real lo toque.

Modelos que nacieron de aquí: 01, 03.

No funcionó · Reportes semanales automáticos

Construimos rápido y validamos lento. Resolvimos el dolor de los gerentes, no el de quienes hacían los reportes, y la calidad quedó por debajo de lo que se le puede mandar a un cliente. Los usuarios nunca se quejaron: simplemente no lo usaron.

Modelos que nacieron de aquí: 05, 06, 07.

> **La lección que une los tres casos:** optimizamos para construir rápido cuando había que optimizar para aprender rápido. La meta no es lanzar funcionalidades. Es lanzar soluciones que la gente de verdad use.

PARA TRABAJAR CON NOSOTROS

Cómo trabajamos

- **Sesgo a la acción.** Preferimos decidir con información incompleta y corregir rápido que esperar certeza.
- **Recomendaciones con opinión.** Traemos una propuesta concreta y los puntos abiertos marcados, no una lista de opciones sin postura.
- **Validación barata primero.** Antes de gastar en serio, buscamos la prueba más pequeña que diga si vale la pena.
- **La adopción es la métrica.** Un producto excelente que nadie usa vale cero.
- **Honestidad con los números.** Si la adopción es baja, lo decimos primero, y el desarrollo para hasta entender por qué.
- **Medir, no adivinar.** La percepción no es prueba: hay desarrolladores que sentían que la IA los hacía 20% más rápidos cuando los hacía 19% más lentos (METR, 2025).

Las preguntas que siempre hacemos

01 ¿Si esto funciona, ¿qué cambia de verdad?
La pregunta detrás de todo el documento.

02 ¿Qué supuesto, si resulta falso, mata toda la iniciativa?
Primeros principios.

03 ¿Quién lo pidió y quién lo va a usar todos los días?
Dolor del jefe ≠ dolor del usuario.

04 ¿Podemos nombrar 3 personas que lo usarán cada semana?
Semana 2: validación o se mata.

05 ¿Qué nos haría parar, y cuándo lo sabremos?
Criterios para matar, escritos antes de arrancar.

06 ¿Quién lo revisa, y cuánto cuesta equivocarse?
Matriz de tareas y la prueba del humo.

07 ¿Estamos optimizando algo cuyo resultado todavía no está probado?
Optimizar o reimaginar.

08 ¿En qué punto de la cadena estamos midiendo, y por qué no más a la derecha?
Esfuerzo → Entregable → Resultado → Impacto.

09 ¿Es ofensivo o defensivo, y lo estamos midiendo como tal?
Proyectos ofensivos y defensivos.

AL FINAL **Nuestro trabajo no es maximizar la inteligencia de un modelo. Es acortar la distancia entre el esfuerzo y el impacto.**

Fuentes: Dell'Acqua y otros, Harvard y BCG (*Organization Science*, 2026); METR, estudio de productividad de desarrolladores (2025); Anand y Wu, "The Gen AI Playbook for Organizations" (HBR, 2025); Shane Parrish sobre primeros principios. Los casos vienen de proyectos reales; omitimos nombres y cifras internas.