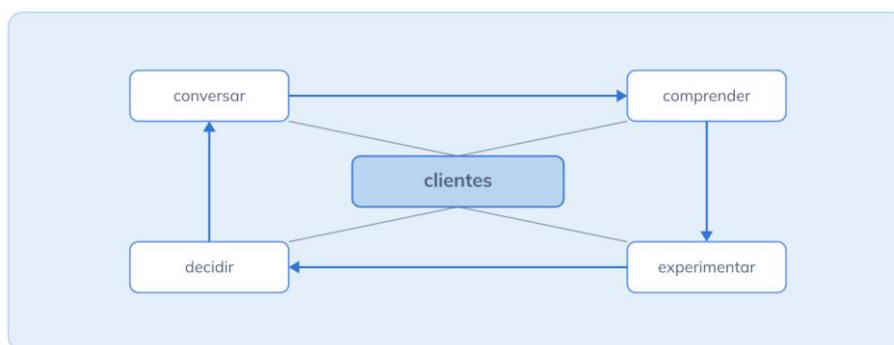


Guía

Descubrimiento de producto y validación con clientes

Desarrollo de los temas del State of the art



Actualizado a septiembre de 2026

Sobre este documento

Este documento desarrolla en profundidad los temas del mapa de referencia (State of the art) para practicar el descubrimiento de producto (*product discovery*) y validar con clientes qué merece construirse antes de construirlo, reduciendo la incertidumbre con evidencia útil de clientes y mercado, que está disponible en: [Skill Arena](#).

Úsalo como material de consulta para mantener y contrastar tu práctica con el conocimiento profesional actual.

Se complementa con la plataforma de entrenamiento y evaluación en Skill Arena. En el área [Descubrimiento de producto y validación con clientes](#) puedes realizar pruebas de entrenamiento para contrastar y mejorar tu nivel de conocimiento y si lo deseas también puedes obtener un diploma que acredita curricularmente la solvencia y vanguardia profesional en esta área.



Estado del conocimiento

El conocimiento sobre el descubrimiento de producto evoluciona de forma progresiva: su núcleo lleva años probado y cambia más por refinamiento de la práctica que por rupturas. Lo que se mueve deprisa es su instrumental. La inteligencia artificial ya transcribe, sintetiza y hasta entrevista, y con ella han llegado prácticas nuevas, evidencia todavía escasa y obligaciones legales que conviene conocer antes de usarla. En los distintos apartados del documento, las etiquetas (ESTABLECIDO, EN CONSOLIDACIÓN, EMERGENTE) ayudan a identificar la madurez de cada concepto:

ESTABLECIDO consenso asentado; conocimiento que se da por necesario.

EN CONSOLIDACIÓN gana adopción con rapidez; aún no universal pero ya relevante.

EMERGENTE frontera reciente; alta relevancia y alta volatilidad.

Índice

Capítulos del desarrollo, en el orden del State of the art.

Bloque A · Conversar y capturar señal

1. Espacio del problema y espacio de la solución
2. La entrevista que aprende, no la que confirma
3. A quién escuchar: segmento, muestra y reclutamiento
4. Notas de calidad: separar hecho, cita e interpretación
5. Consentimiento, datos y transparencia al conversar con clientes

Bloque B · De los datos a la comprensión

6. Síntesis: de la observación al insight accionable
7. Enmarcar el problema: el trabajo que el cliente quiere hacer

Bloque C · Diseñar experimentos para aprender

8. Mapeo de supuestos: los riesgos que hay que probar primero
9. Tests de prototipo: tareas como objetivos, no instrucciones
10. El MVP como experimento, no como producto reducido

Bloque D · Decidir con evidencia

11. Criterios de éxito y señales definidos de antemano
12. La fuerza de la evidencia y el control de sesgos
13. Perseverar, iterar, pivotar o parar

Bloque E · El descubrimiento como hábito del equipo

14. Descubrir cada semana: el trío de producto y el contacto continuo
15. El árbol de oportunidades: del resultado al experimento

Bloque F · Descubrir con inteligencia artificial

16. Entrevistar y sintetizar con IA: dónde ayuda y dónde erosiona
17. Lo sintético en la escala de la evidencia
18. Validar un producto cuyo núcleo es un modelo

BLOQUE A · CONVERSAR Y CAPTURAR SEÑAL

Capítulo 1. Espacio del problema y espacio de la solución · ESTABLECIDO

Descubrir exige entender a quién le pasa qué, en qué situación y qué se lo impide antes de comprometerse con una respuesta, y volver a ese espacio cada vez que una respuesta falla.

Introducción

Casi todo lo que sale mal en el descubrimiento de producto empieza por una confusión anterior a cualquier técnica. El equipo recibe una petición, una idea o una métrica que se ha torcido, y antes de haber entendido el problema ya está discutiendo cómo resolverlo. A partir de ahí la conversación con clientes se convierte en una forma de defender esa respuesta, y la evidencia se busca para confirmarla.

En este capítulo veremos qué distingue los dos espacios en los que se mueve el descubrimiento, cómo se cuela la solución dentro del problema sin que nadie lo advierta y cómo se trabaja en ambos sin mezclarlos. Es la primera capacidad que acredita esta *skill* y la que da sentido al resto de la guía.

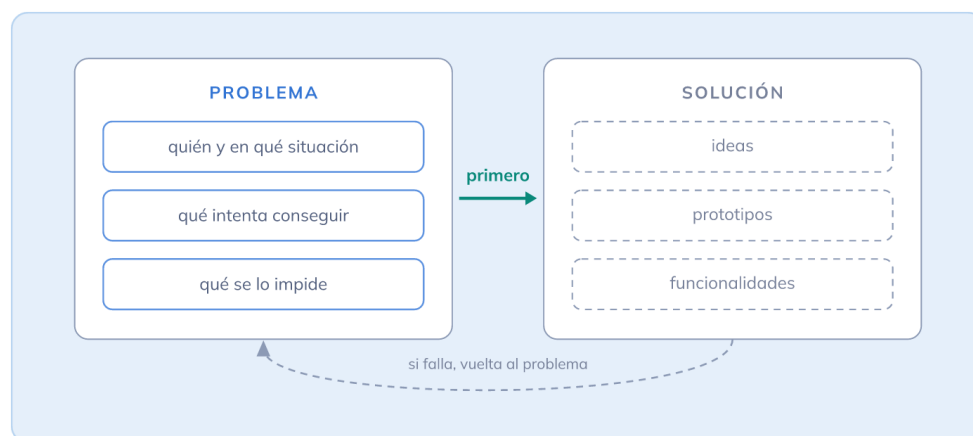


Figura 1. Dos espacios que no se mezclan: el problema se entiende antes de comprometerse con una solución, y se vuelve a él cada vez que una solución falla.

1.1 Dos espacios con preguntas distintas

Dan Olsen hizo explícita la distinción en *The Lean Product Playbook*. El espacio del problema recoge lo que le ocurre al cliente con independencia de lo que nosotros construyamos; el de la solución, todo lo que podríamos construir para responder. Cada espacio tiene sus propias preguntas, y conviene saber en cuál está uno antes de formular la siguiente.

Las preguntas del espacio del problema

- **Quién** tiene el problema, descrito por lo que hace y no solo por su perfil.
- **En qué situación** aparece, con sus disparadores y restricciones concretas.
- **Qué intenta conseguir** esa persona cuando se encuentra con él.

- **Qué se lo impide** hoy, y cuánto le cuesta ese impedimento.
- **Cómo lo resuelve ahora**, porque siempre hay una alternativa, aunque sea aguantarse.

Las preguntas del espacio de la solución

Aquí las preguntas cambian de objeto. Se trata de qué ideas podrían responder a la oportunidad elegida, qué prototipo permitiría comprobarlas, qué funcionalidad sería la mínima para aprender algo y qué supuestos tendrían que ser ciertos para que funcionara. Son preguntas legítimas y necesarias, siempre que lleguen después de las anteriores y no en su lugar.

La última pregunta del espacio del problema, cómo lo resuelve hoy la persona, merece atención especial. La alternativa actual (una hoja de cálculo, un correo, una llamada, o simplemente aguantarse) marca el listón que cualquier solución tendrá que superar y revela restricciones que nadie mencionaría si se le preguntara en abstracto. Un problema sin alternativa actual conocida suele ser un problema todavía mal explorado.

1.2 Cómo se cuela la solución dentro del problema

La mezcla rara vez es deliberada. Llega disfrazada de problema en frases que el equipo oye a diario, como «necesitan un panel de control», «falta la exportación a PDF» o «hay que mejorar el registro». Cada una contiene ya una respuesta, y aceptarla como punto de partida cierra la exploración antes de empezar.

Señales de que el problema ya trae la solución

- El enunciado **nombra una funcionalidad**, una pantalla o una herramienta concreta.
- El objetivo de la investigación usa verbos como **validar, confirmar o demostrar** referidos a una idea.
- El éxito se define como **aceptación de la propuesta** y no como comprensión de lo que le pasa al cliente.
- Nadie sabe decir **para qué** necesita el cliente lo que pide ni qué hace hoy en su lugar.

La corrección más útil es preguntar por el propósito que hay detrás de la petición. ¿Para qué necesitan el PDF, a quién se lo envían, qué hacen con él después? A menudo la respuesta revela un problema distinto (compartir datos con terceros, por ejemplo) que admite soluciones mejores que la pedida. El capítulo 7 desarrolla esta excavación con el marco de los trabajos que el cliente quiere hacer.

1.3 Trabajar en los dos espacios sin mezclarlos

Separar los espacios no significa recorrerlos en fases estrictas. En la práctica el equipo va y viene: una entrevista sugiere una idea, un prototipo revela algo del problema que nadie había visto, una prueba fallida obliga a replantear a quién se está sirviendo. Lo que importa es que cada actividad sepa en qué espacio trabaja y qué pregunta intenta responder.

Un prototipo, por ejemplo, puede servir en los dos espacios, aunque con propósitos distintos. Como sonda del problema, se enseña para que la persona cuente cómo encaja en su situación real; como prueba de la solución, se usa para ver si consigue lo que se pretende. El ritual de idear y prototipar en grupo tiene skill propia en *Scrum Manager*, Design Thinking; aquí interesa no confundir ambos usos al interpretar lo que se observa.

Un ejemplo ayuda a ver la diferencia. Un equipo que diseña una herramienta de planificación enseña a varios jefes de obra un boceto de calendario compartido. Si lo que busca es entender cómo coordinan hoy a sus subcontratas, el boceto es una excusa para que cuenten sus historias, y lo que anota son esas historias. Si lo que busca es saber si el calendario resuelve la coordinación, lo que observa es si consiguen hacer con él lo que necesitan. Confundir ambos usos lleva a tomar como validación de la solución lo que solo era una conversación sobre el problema.

Volver al problema cuando algo falla. Cuando una solución no funciona, la primera pregunta no es qué otra solución probar, sino si habíamos entendido bien el problema. Muchas iteraciones fallidas son síntomas de un problema mal enmarcado que ninguna solución iba a resolver.

1.4 Qué cambia cuando construir es barato

La inteligencia artificial ha abaratado el espacio de la solución como nunca. Marty Cagan señala que, con las herramientas actuales, hacer diez o veinte prototipos a la semana está al alcance de casi cualquiera, y Teresa Torres advierte que construir sea tan barato no significa que convenga construir cada idea. La tentación de saltarse el problema aumenta en la misma medida en que baja el coste de probar soluciones.

La consecuencia práctica es que la disciplina de este capítulo pesa más que antes. Un equipo puede generar en una tarde varias versiones funcionales de una respuesta a un problema que nunca exploró, y todas parecerán progreso. Separar los espacios es lo que permite aprovechar la velocidad para aprender del cliente en lugar de acumular soluciones sin dueño.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Distinguir en una conversación o un documento qué afirmaciones pertenecen al espacio del problema y cuáles al de la solución.
- Reformular una petición que trae la solución dentro en términos de quién, en qué situación, qué intenta conseguir y qué se lo impide.
- Detectar las señales de que un objetivo de investigación está planteado desde una solución.
- Indicar, ante un prototipo o una prueba, si se está usando para explorar el problema o para probar una solución.

- Volver a revisar el enmarcado del problema cuando una solución falla, antes de proponer la siguiente.

Errores y antipatronos frecuentes

- **Aceptar la petición como problema.** «Necesitan un panel de control» ya es una respuesta. Preguntar para qué lo necesitan y qué hacen hoy en su lugar devuelve la conversación al problema.
- **Investigar para defender una idea.** Cuando el equipo ya se ha encariñado con una solución, la investigación deriva hacia confirmarla. Formular el objetivo en términos del cliente, no de la idea, lo evita.
- **Iterar soluciones sobre un problema mal entendido.** Probar variante tras variante sin revisar el enmarcado suele encubrir que el problema estaba mal definido desde el principio.
- **Confundir velocidad con aprendizaje.** Generar muchas soluciones rápidas con IA parece avance, pero sin una oportunidad bien entendida detrás solo produce más cosas que decidir.

Para profundizar

- [Scrum Manager BoK — Product discovery](#)
- [Dan Olsen — The Lean Product Playbook](#)
- [Skill Arena — Design Thinking](#)

BLOQUE A · CONVERSAR Y CAPTURAR SEÑAL

Capítulo 2. La entrevista que aprende, no la que confirma

ESTABLECIDO

Una entrevista de descubrimiento explora el comportamiento pasado y el contexto real de la persona; no es una encuesta oral ni una venta encubierta.

Introducción

Todo el mundo recomienda hablar con los clientes, y casi nadie explica cómo hacerlo sin dejarse engañar. Ese es el punto de partida de Rob Fitzpatrick en *The Mom Test*, cuyo título resume el problema: hasta tu madre te mentirá para no herirte si le preguntas si tu idea es buena. La dificultad no está en las personas, sino en las preguntas que les hacemos.

En este capítulo veremos el propósito real de una entrevista de descubrimiento, qué tipo de preguntas generan señal y cuáles generan ruido, cómo se cuele el sesgo de confirmación en un guion y cómo se conduce la conversación para aprender algo que no sabíamos.

2.1 El propósito: aprender, no validar

Una entrevista de descubrimiento no busca que el cliente apruebe una idea. Busca entender cómo es su situación, qué hace hoy, dónde se atasca y qué intenta conseguir, para decidir después qué merece la pena construir. La diferencia entre una entrevista útil y una inútil suele decidirse antes de la primera pregunta, en lo que el equipo espera sacar de ella.

Una comprobación sencilla ayuda a fijar ese propósito antes de redactar el guion. Consiste en poder completar la frase «queremos aprender X y lo sabremos si observamos Y». Si no se sabe qué se quiere aprender, las notas saldrán difusas; si solo se sabe qué se quiere oír, la entrevista confirmará lo que ya se pensaba.

Una estructura sencilla para el guion

1. **Apertura:** explicar el propósito, pedir permiso para grabar y dejar claro que no hay respuestas correctas.
2. **Historia:** pedir una ocasión concreta relacionada con el objetivo y dejar que la persona la cuente entera.
3. **Profundización:** volver sobre los momentos de fricción, las alternativas que usó y lo que le costó cada una.
4. **Cierre:** preguntar si hay algo relevante que no haya salido y agradecer, sin presentar todavía ninguna solución.

El guion es una ayuda, no un cuestionario que haya que completar. Si la historia lleva a un terreno inesperado pero relevante, seguirla suele aportar más que volver a la siguiente pregunta prevista.

2.2 Preguntas sobre el pasado, no sobre el futuro

Las preguntas que generan evidencia apuntan al comportamiento pasado y concreto. Teresa Torres lo convierte en método con la entrevista basada en historias (*story-based interviewing*), que pide a la persona que cuente una ocasión específica, con su contexto y su secuencia, en lugar de generalizar sobre lo que suele hacer. El pasado se recuerda con detalles verificables; el futuro se inventa por cortesía.

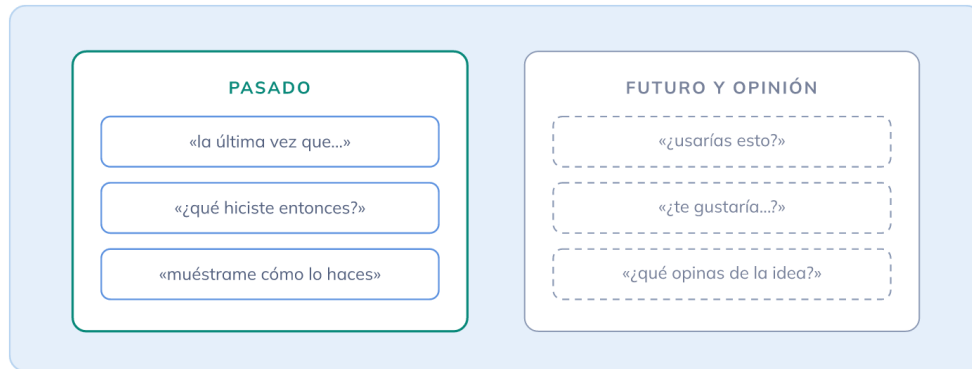


Figura 2. Preguntas que dan señal y preguntas que dan ruido: las primeras piden historias concretas del pasado y las segundas, predicciones u opiniones.

Fitzpatrick resume el mismo criterio en tres pautas: hablar de la vida de la persona y no de nuestra idea, preguntar por casos concretos del pasado y no por generalidades o intenciones, y escuchar mucho más de lo que se habla. Las preguntas de opinión («¿te gustaría esta función?», «¿no sería mejor tenerlo todo en un sitio?») piden estimaciones especulativas, empujan a la respuesta amable o la sugieren directamente.

Cómo reformular una pregunta de opinión

- «¿Usarías esto?» se convierte en «cuéntame la última vez que tuviste este problema».
- «¿Cuánto pagarías?» se convierte en «¿qué has usado hasta ahora para resolverlo y cuánto te cuesta?».
- «¿Te parece útil?» se convierte en «muéstrame cómo lo haces ahora».

2.3 La trampa del sesgo de confirmación

Cuando el equipo ya tiene una solución en mente, surge la tendencia a diseñar preguntas que la confirmen. Un objetivo como «validar si los usuarios quieren la nueva pantalla de alta» lleva la solución dentro y empuja hacia respuestas que la apoyen. El problema se ve menos en la entrevista que en el guion que se preparó antes.

Señales de alerta en un objetivo o un guion

- El objetivo **nombra una funcionalidad** concreta.
- Usa verbos como **validar, confirmar o demostrar** referidos a una solución.
- Define el éxito como **un número mínimo de respuestas positivas** .
- Las preguntas **presentan la idea** antes de haber explorado la situación.

La corrección es reformular hacia la exploración. «Validar la nueva pantalla de alta» pasa a ser «entender cómo se registran hoy, dónde se atascan y qué intentan conseguir». Con ese objetivo, la entrevista puede sorprender, que es precisamente lo que se le pide.

2.4 Conducir la conversación

Durante la entrevista, las preguntas de seguimiento valen más que las del guion. Cuando la persona menciona una fricción, conviene quedarse en ella («¿qué pasó entonces?», «¿cómo lo resolviste?») hasta tener una historia completa. Los silencios también ayudan, porque suelen traer el detalle que la primera respuesta omitió.

Conviene además separar el descubrimiento de la venta. Si la persona pregunta por la solución, se puede anotar y posponer hasta el final, pero presentar el producto a mitad de la conversación contamina todo lo que venga después. Cuando es posible, conviene que entrevisten dos personas, una que conduce la conversación y otra que toma notas, porque es difícil escuchar bien y registrar bien a la vez. Quién entrevista también importa, y el capítulo 14 trata cómo lo organiza un equipo que descubre cada semana.

La entrevista moderada por IA. Ya hay plataformas que conducen entrevistas sin entrevistador humano. El Nielsen Norman Group concluyó en enero de 2026 que sirven para recoger opiniones sobre un producto o cribar participantes, pero no para explorar un problema desconocido, que es justo lo que pide esta entrevista. El capítulo 16 desarrolla cuándo usarlas.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Formular el objetivo de aprendizaje de una entrevista antes de redactar el guion, en términos de qué se quiere aprender y qué observación lo indicaría.
- Distinguir una pregunta sobre comportamiento pasado de una de opinión o intención futura, y reescribir la segunda en términos de la primera.
- Pedir y seguir una historia concreta del pasado hasta obtener su contexto y su secuencia.
- Detectar en un objetivo o un guion las señales del sesgo de confirmación.
- Separar en la práctica la conversación de descubrimiento de la presentación de una solución.

Errores y antipatrones frecuentes

- **Preguntar «¿te gustaría...?».** Pide una predicción de cortesía sin valor. Reformular hacia el pasado, con «cuéntame la última vez que...», devuelve la conversación a hechos.

- **Mezclar venta y descubrimiento.** Presentar la solución a mitad de la entrevista contamina la señal. La propuesta, si se pide, se deja para el final o para otra conversación.
- **Salir con un objetivo vago.** «Recoger *feedback*» no dice qué supuesto se explora ni qué evidencia lo aclararía, así que las notas no serán accionables.
- **Conformarse con generalidades.** «Normalmente lo hago así» es una opinión sobre uno mismo. Pedir la última vez concreta convierte la generalidad en una historia que se puede analizar.

Para profundizar

- [Rob Fitzpatrick — The Mom Test](#)
- [Teresa Torres — Customer interviews](#)
- [Nielsen Norman Group — Interviewing users](#)

BLOQUE A · CONVERSAR Y CAPTURAR SEÑAL

Capítulo 3. A quién escuchar: segmento, muestra y reclutamiento · ESTABLECIDO

Lo que se aprende depende de a quién se escucha: el segmento se fija antes de reclutar, los participantes deben ser externos y representativos y la muestra se dimensiona por el propósito.

Introducción

Una entrevista impecable con la persona equivocada produce evidencia equivocada, y ninguna técnica posterior lo corrige. El sesgo de selección aparece igual en una entrevista, en un test de prototipo y en la validación de la demanda, porque nace en una decisión que se toma antes de cualquier conversación: a quién se invita.

En este capítulo veremos cómo se define el segmento, cómo se comprueba que un participante pertenece a él sin delatar lo que se busca, por qué la red propia es una fuente peligrosa y cuántas conversaciones hacen falta según lo que se quiera aprender.

3.1 Definir el segmento antes de reclutar

Reclutar empieza por decidir a quién interesa escuchar y por qué. El segmento se describe mejor por lo que las personas hacen que por su perfil demográfico: quien ha contratado un servicio parecido en los últimos seis meses, quien gestiona las facturas de un equipo, quien abandonó el alta a medias. Esa descripción se traduce después en criterios comprobables.

Un ejemplo muestra la diferencia. «Pymes del sector servicios» es un perfil que admite a casi cualquiera; «personas que en los últimos tres meses han preparado ellas mismas la declaración trimestral de su empresa» describe una situación comprobable y

relevante para quien quiera entender cómo se gestiona ese trámite. El segundo segmento produce menos candidatos, pero cada conversación vale mucho más.

Criterios de inclusión y de exclusión

- **De inclusión:** el comportamiento o la situación que hacen relevante a la persona para la pregunta de investigación.
- **De exclusión:** lo que la haría poco representativa, como trabajar en el sector, conocer el producto por dentro o participar en estudios con demasiada frecuencia.
- **De variedad:** lo que conviene repartir dentro de la muestra (antigüedad, uso intenso o esporádico, quienes se quedaron y quienes se fueron).

Los casos que no encajan con la mayoría merecen atención especial. El capítulo 6 explica por qué suelen revelar otro segmento con necesidades distintas, y por qué conviene haber reclutado con variedad suficiente para poder verlo.

3.2 El cuestionario de selección

Los criterios se comprueban con un cuestionario de selección (*screenener*) que se pasa antes de citar a nadie. Kathryn Whinton, del Nielsen Norman Group, insiste en que debe equilibrar dos objetivos: obtener la información necesaria y no revelar de qué trata el estudio, porque una pregunta demasiado directa invita a exagerar para ser elegido.

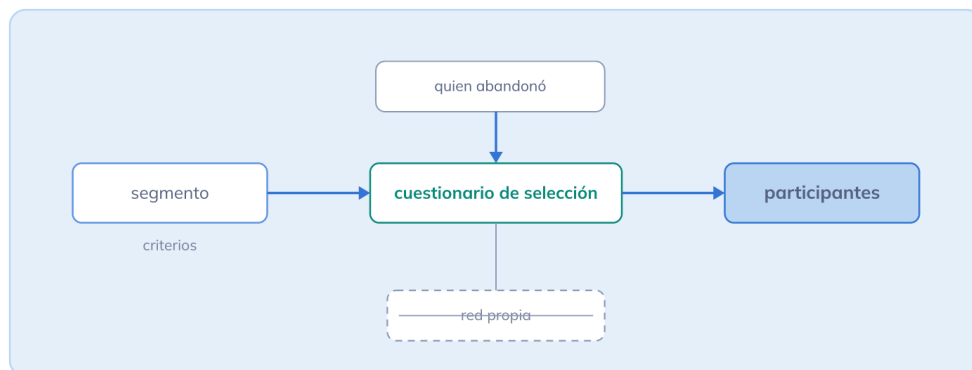


Figura 3. Del segmento a los participantes: los criterios se comprueban con un cuestionario de selección que no delata lo que se busca, y la red propia queda fuera.

Cómo redactarlo

- Preguntar de forma abierta o indirecta: en lugar de «¿juegas a juegos en línea?», «¿qué actividades haces en línea?».
- En las preguntas cerradas, incluir opciones distractoras plausibles junto a la que interesa.
- Preguntar por comportamiento reciente y concreto, no por intenciones.
- Desconfiar de quien marca todas las opciones una y otra vez, porque suele buscar el incentivo más que participar.

3.3 Fuera de la red propia

La fuente más cómoda de participantes es también la más sesgada. Amigos, compañeros, seguidores y clientes entusiastas conocen el contexto y tienen incentivos para agradar, así que su señal llega inflada. Reclutar solo entre quienes completaron un proceso añade además el sesgo de supervivencia, porque deja fuera a quienes abandonaron, que suelen ser los que más enseñan.

Fuentes que corrigen el sesgo

- Personas que se dieron de baja o abandonaron un proceso a medias.
- Usuarios de alternativas o de la competencia, incluida la alternativa de no hacer nada.
- Participantes captados desde el propio producto con criterios definidos de antemano.
- Paneles de reclutamiento, filtrados con el cuestionario de selección.

Cada vez reclutan más personas sin formación específica en investigación. Según Maze, en 2026 el 39 % de los encuestados afirma que en su organización los responsables de producto ya investigan con usuarios. Es una buena noticia para la frecuencia del contacto con clientes, siempre que la decisión de a quién escuchar se tome con el mismo cuidado que la de qué preguntar.

3.4 Cuántas conversaciones hacen falta

El tamaño de la muestra depende del propósito. Para descubrir patrones, problemas y oportunidades basta un número pequeño de conversaciones bien elegidas, que se amplía mientras sigan apareciendo cosas nuevas y se detiene cuando las entrevistas empiezan a repetir lo ya oído. En la investigación cualitativa se habla de saturación para describir ese punto, y conviene comprobarlo por segmento, porque un grupo puede estar saturado mientras otro apenas se ha explorado. Para medir cuántos clientes comparten un patrón, en cambio, hacen falta otros métodos y muestras mucho mayores.

Conviene no trasladar cifras de un contexto a otro. Los cinco usuarios de Jakob Nielsen se refieren a encontrar problemas de usabilidad en un test de un diseño concreto, no a explorar un problema en entrevistas ni a estimar la demanda. Lo honesto es declarar para qué sirve la muestra y no pedirle conclusiones que no puede dar.

En el descubrimiento continuo (*continuous discovery*), donde el equipo habla con clientes cada semana, la pregunta cambia de forma. Ya no se trata de cuántas entrevistas tiene un estudio, sino de si la cadencia permite escuchar a los distintos segmentos con la frecuencia suficiente para detectar cambios. El capítulo 14 trata cómo se sostiene ese reclutamiento semana a semana.

Una muestra pequeña descubre, no cuantifica. «Siete de doce entrevistados mencionaron la fricción» es un dato valioso para decidir qué explorar, pero no autoriza a afirmar que el 58 % de los clientes la sufre.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Traducir un segmento de interés en criterios de inclusión, exclusión y variedad basados en comportamiento.
- Redactar preguntas de un cuestionario de selección que comprueben los criterios sin delatar el objetivo del estudio.
- Identificar en una muestra el sesgo de la red propia y el de supervivencia, y proponer fuentes que los corrijan.
- Dimensionar la muestra según el propósito, distinguiendo descubrir patrones de medir su frecuencia.
- Explicar por qué una cifra de muestra válida para un tipo de estudio no se traslada a otro.

Errores y antipatrones frecuentes

- **Reclutar a quien está a mano.** Compañeros, amigos y clientes entusiastas quieren agradar y conocen el contexto. Su opinión puede orientar, pero no sustituye a participantes externos del segmento.
- **Delatar el estudio en la selección.** Una pregunta como «¿te interesaría una aplicación para X?» invita a exagerar. Preguntar por lo que la persona hace revela quién encaja sin darle pistas.
- **Escuchar solo a los que se quedaron.** Entrevistar únicamente a quienes completaron el proceso deja fuera a quienes abandonaron, que suelen explicar mejor qué falla.
- **Cuantificar con una muestra de descubrimiento.** Doce entrevistas sirven para encontrar patrones, no para afirmar porcentajes de la población. Declarar el alcance de la muestra evita conclusiones que no sostiene.

Para profundizar

- [Nielsen Norman Group — Screening questions](#)
- [Nielsen Norman Group — Why you only need to test with 5 users](#)
- [Maze — Future of User Research Report 2026](#)

BLOQUE A · CONVERSAR Y CAPTURAR SEÑAL

Capítulo 4. Notas de calidad: separar hecho, cita e interpretación · ESTABLECIDO

La síntesis solo es tan buena como las notas que la alimentan; separar lo que se dijo, lo que se hizo y lo que se interpreta desde la captura es lo que la hace posible.

Introducción

Una entrevista excelente con notas pobres equivale a una entrevista perdida. Los matices de una conversación se desvanecen deprisa, y lo que no quedó registrado con precisión se reconstruye después con los sesgos de quien escucha. La grabación y la transcripción automática han cambiado qué hay que anotar en vivo, pero no la necesidad de separar niveles.

En este capítulo veremos qué se captura y en qué orden, cómo se separa la observación de la interpretación, cómo se cierran las notas en caliente con el resumen de una página que propone Teresa Torres y qué cuidado piden los resúmenes que generan las herramientas.

4.1 Qué se captura y en qué orden

Durante la entrevista y justo después se separan cuatro tipos de registro. Cada uno sirve para algo distinto en la síntesis, y mezclarlos desde el inicio hace inviable un análisis riguroso.



Figura 4. Cuatro registros separados: cita, hecho, interpretación y pendiente se anotan por separado, y la interpretación va marcada como tal.

1. **La cita:** las palabras exactas de la persona, entre comillas.
2. **El hecho:** lo que hizo o contó que hizo, con su contexto de quién, cuándo y en qué situación, que es lo que permite segmentar después.
3. **La interpretación:** lo que infiere quien entrevista sobre el significado o la motivación, marcado claramente como inferencia propia.
4. **Lo pendiente:** las dudas que quedan abiertas y merecen otra pregunta u otra entrevista.

Cuando la conversación se graba y se transcribe, la cita literal deja de ser escasa, porque la transcripción la conserva entera. El valor de las notas tomadas en vivo se desplaza

entonces al contexto, a la interpretación marcada y a lo pendiente, que ninguna transcripción recoge por sí sola.

Por eso conviene que quien toma notas en vivo se concentre en lo que la grabación no captura. Qué hizo la persona al contar algo (si buscó un documento, si dudó, si abrió una herramienta para mostrarlo), qué contradicciones aparecieron entre lo que dijo y lo que hizo, y qué preguntas quedaron sin respuesta. Son precisamente los datos que más cuestan de reconstruir después.

4.2 Separar observación de interpretación

Anotar interpretaciones es legítimo y útil, siempre que estén marcadas. Un sistema de marcas sencillo basta, con una para la cita, otra para el hecho, otra para la interpretación y otra para lo pendiente. Así, al sintetizar, se distingue lo que la persona dijo e hizo de lo que el entrevistador supuso, y se evita tomar la inferencia por dato.

Una comprobación rápida ayuda a detectar el problema. Si las notas suenan a adjetivos («le encantó», «buena actitud», «contento con los tiempos»), probablemente falta evidencia concreta. «Paso dos horas al día en reuniones de seguimiento» es un hecho; «está contento» es una valoración sin sustento que nadie podrá revisar después.

Cuando entrevistan dos personas, comparar sus notas al terminar es una de las formas más baratas de detectar interpretaciones que se habían colado como hechos. Si una anotó «no le interesa la función» y la otra «dijo que ya usa otra herramienta para eso», la segunda es la observación y la primera, la conclusión que alguien sacó de ella. Resolver esas diferencias en caliente, con la conversación todavía fresca, mejora mucho el material que llegará a la síntesis.

4.3 Cerrar en caliente: el resumen de entrevista

Teresa Torres dio forma a la práctica de procesar cada conversación nada más terminarla con el resumen de entrevista (*interview snapshot*). Es una página por entrevista que la hace memorable, accionable y fácil de consultar cuando el equipo acumula decenas de conversaciones.

Qué contiene

- El nombre de la persona y unos datos rápidos de contexto (rol, tipo de organización).
- Una cita memorable que evoque su historia.
- Las oportunidades: necesidades no cubiertas, fricciones y deseos.
- Los hallazgos que no son oportunidades pero merecen recordarse.
- Un mapa de la experiencia que ordena los momentos clave de la historia.

Torres recomienda reservar el tiempo en la propia agenda de la entrevista, con 20 a 30 minutos de conversación y 15 a 25 más para el resumen, mientras las historias siguen frescas. Esperar días para cerrar las notas equivale a perder matices que ya no se

podrán recuperar. El capítulo 15 muestra cómo las oportunidades de estos resúmenes alimentan el árbol de oportunidades (*opportunity solution tree*).

4.4 Transcripciones y resúmenes generados

Las herramientas actuales transcriben la conversación y proponen un resumen en segundos. La transcripción es un registro fiel y muy valioso; el resumen, en cambio, es una interpretación más, hecha por un sistema que no estuvo en la conversación y que tiende a allanar lo que no encaja. Conviene tratarlo como tal y contrastarlo con la transcripción antes de apoyarse en él.

Grabar y transcribir tiene además obligaciones propias sobre los datos de la persona, que el capítulo 5 desarrolla. Y el uso de la IA para sintetizar varias entrevistas, con el riesgo de delegar el juicio, se trata en el capítulo 16.

Por último, las notas y los resúmenes solo sirven si el equipo los encuentra cuando los necesita. Guardarlos en un repositorio común, con el mismo formato y etiquetados por segmento y por oportunidad, permite volver a una conversación de hace meses cuando aparece una pregunta nueva, y evita que el conocimiento de clientes viva en los cuadernos de quien entrevistó. El capítulo 14 trata cómo se integra ese repositorio en el hábito semanal del equipo.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Separar en las notas de una entrevista la cita, el hecho con su contexto, la interpretación marcada y lo pendiente.
- Reescribir una interpretación encubierta como dato en una observación verificable con contexto.
- Detectar notas basadas en adjetivos y convertirlas en hechos concretos.
- Elaborar el resumen de una entrevista con sus oportunidades, hallazgos y mapa de la experiencia en el tiempo que sigue a la conversación.
- Contrastar un resumen generado por una herramienta con la transcripción antes de usarlo en la síntesis.

Errores y antipatrones frecuentes

- **Notas sin contexto.** Sin quién, cuándo y en qué situación no se puede derivar ningún *insight* ni segmentar después.
- **Interpretaciones que parecen datos.** Mezclarlas con los hechos produce conclusiones que nadie puede verificar. Marcar cada interpretación como tal permite revisarla.
- **Registrar solo lo que confirma.** Anotar lo que encaja con la hipótesis y olvidar lo que la contradice es sesgo de confirmación. Conviene preguntarse si se anotó algo que no gustó escuchar.

- **Fiarse del resumen automático.** Tomar el resumen de la herramienta como si fueran las notas pierde los matices y las excepciones. La transcripción es la fuente; el resumen, una lectura más.

Para profundizar

- [Teresa Torres — Interview snapshots](#)
- [Scrum Manager BoK — Interview snapshot](#)
- [Teresa Torres — Customer interviews](#)

BLOQUE A · CONVERSAR Y CAPTURAR SEÑAL

Capítulo 5. Consentimiento, datos y transparencia al conversar con clientes · EN CONSOLIDACIÓN

Grabar, transcribir y analizar una entrevista es tratar datos personales: se informa antes y durante, se pide consentimiento por grabación, se comprueba qué hace la herramienta con los datos y, si entrevista una IA, la persona lo sabe.

Introducción

La entrevista se ha llenado de herramientas. Se graba la videollamada, un servicio la transcribe, otro resume y un asistente ayuda a sintetizar decenas de conversaciones. Cada paso trata datos de la persona que habló con nosotros, y lo que antes era una cuestión de buena educación tiene hoy criterios de autoridad que conviene conocer.

En este capítulo veremos qué datos se tratan al conversar con clientes, qué hay que hacer antes, durante y después de grabar según los criterios que publicó en 2026 la Agencia Española de Protección de Datos, qué preguntar a las herramientas y qué cambia cuando quien entrevista o analiza es una IA. No es un capítulo de asesoramiento jurídico, sino de práctica profesional informada.

5.1 Qué datos se tratan al conversar

En una entrevista se recogen datos personales de varias formas. Está la identidad de la persona y su contexto, está lo que cuenta, que puede incluir información sensible sobre su trabajo o su vida, y está la propia grabación. La Agencia Española de Protección de Datos recuerda que la voz debe considerarse dato personal en la medida en que puede identificar a una persona.

De ahí se sigue que el Reglamento General de Protección de Datos se aplica a la grabación, a la transcripción, a las notas identificables y a lo que hagan con todo ello las herramientas. Conviene tenerlo presente desde el diseño del estudio, no descubrirlo cuando ya hay veinte grabaciones en una carpeta compartida.

Dos principios del reglamento ayudan a decidir en la práctica. Uno es la minimización, que pide recoger solo los datos necesarios para el propósito del estudio: si

no hace falta el nombre completo de la empresa para entender el problema, no se anota. El otro es la limitación del plazo de conservación, que invita a fijar desde el principio cuánto tiempo se guardarán las grabaciones y a borrarlas cuando las notas y los resúmenes ya recogen lo aprendido.

5.2 Antes, durante y después de grabar

En abril de 2026 la Agencia publicó la segunda parte de sus criterios sobre la transcripción de voz con IA, centrada en la responsabilidad, los derechos y la transparencia. De ella se desprende una secuencia práctica para cualquier equipo que grabe entrevistas.

1. **Antes:** informar a la persona de que se va a grabar, para qué y quién accederá, y pedir un consentimiento específico para esa grabación; el consentimiento de una sesión no vale para las siguientes.
2. **Durante:** mantener un indicador visible y activo mientras se graba, como un aviso en pantalla, de modo que la persona sepa en todo momento que el tratamiento está ocurriendo.
3. **Después:** garantizar la exactitud de la transcripción y rectificarla sin demora si contiene errores, además de atender los derechos de acceso de la persona.

Una práctica que cabe en el guion. Abrir la entrevista explicando la grabación y pidiendo permiso, y cerrarla recordando cómo se usará lo que se ha contado, no alarga la conversación y deja constancia de algo que después cuesta reconstruir.

5.3 Qué hace la herramienta con los datos

La Agencia distingue dos operaciones que a menudo van juntas. Una es la transcripción que se encarga al servicio; otra, el uso de las muestras de voz para mejorar el propio sistema. Cuando el proveedor reentrena su modelo con los datos para sus propios fines, pasa a ser responsable de ese tratamiento y necesita una base jurídica distinta de la que ampara la transcripción.

Qué comprobar antes de adoptar una herramienta

- Si usa las grabaciones o las transcripciones para entrenar sus modelos, y si se puede desactivar.
- Dónde se almacenan los datos y durante cuánto tiempo.
- Si infiere emociones u otras categorías sensibles a partir de la voz o la imagen.
- Qué contrato de tratamiento de datos ofrece a la organización.

Las notas crudas merecen el mismo cuidado. Pegar transcripciones con nombres, empresas o datos identificables en un asistente de IA de uso general los envía a un tercero sin las garantías anteriores. Anonimizarlas antes es una costumbre barata, coherente con lo que recomienda la skill Historias de Usuario para el contexto que se entrega a un agente.

5.4 Cuando la IA entrevista o infiere emociones

El Reglamento europeo de IA añade obligaciones de transparencia que se aplican desde el 2 de agosto de 2026. Su artículo 50 exige que los sistemas que interactúan directamente con personas estén diseñados para informarles de que tratan con una IA, salvo que sea evidente, y obliga a quien despliega un sistema de reconocimiento de emociones a informar a las personas expuestas. Una entrevista moderada por IA entra de lleno en el primer supuesto.

La detección de emociones merece más cautela. La Agencia advierte que los servicios de transcripción que la incluyen quedan sometidos a un régimen estricto y pueden caer en usos que el Reglamento prohíbe, y su artículo 5 prohíbe inferir emociones en el lugar de trabajo y en centros educativos. Si se entrevista a empleados propios para una herramienta interna, conviene tratarlo como un riesgo serio y consultarlo antes de activar esas funciones.

Lo que no cambió con el aplazamiento. El paquete de simplificación que entró en vigor el 27 de julio de 2026 aplazó las obligaciones de alto riesgo y dio a los proveedores de sistemas ya comercializados hasta el 2 de diciembre de 2026 para el mercado del contenido sintético. Los deberes de transparencia de quien usa estos sistemas siguen vigentes desde agosto.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Identificar qué datos personales se tratan en una entrevista grabada, transcrita y analizada con herramientas.
- Aplicar la secuencia de información previa, indicador visible durante la grabación y consentimiento específico por grabación.
- Comprobar si una herramienta de transcripción o síntesis reentrena sus modelos con los datos y qué implica para su adopción.
- Anonimizar notas y transcripciones antes de usarlas en un asistente de IA de uso general.
- Reconocer cuándo el uso de una IA como entrevistador o de la detección de emociones activa obligaciones de transparencia o riesgos de prohibición.

Errores y antipatrones frecuentes

- **Grabar sin avisar de forma explícita.** Dar por hecho que aceptar la videollamada es aceptar la grabación incumple los criterios de la Agencia. Se informa antes y se mantiene el aviso visible.
- **Reutilizar un consentimiento.** El permiso para grabar una sesión no cubre las siguientes ni otros usos. Cada grabación lleva su propio consentimiento.
- **Pegar notas crudas en un asistente.** Nombres y datos identificables acaban en un tercero sin garantías. Anonimizar antes cuesta minutos.

- **Activar la detección de emociones por defecto.** Puede caer en usos restringidos o prohibidos, sobre todo con empleados propios. Se desactiva salvo que exista una justificación analizada.

Para profundizar

- [AEPD — Transcripción de voz con IA \(I\)](#)
- [AEPD — Transcripción de voz con IA \(II\)](#)
- [Reglamento de IA — artículo 50](#)
- [Morgan Lewis — Qué entró en vigor el 2 de agosto de 2026](#)

BLOQUE B · DE LOS DATOS A LA COMPRENSIÓN

Capítulo 6. Síntesis: de la observación al insight accionable

La síntesis convierte conversaciones sueltas en comprensión que guía decisiones, y solo lo consigue si sube por niveles y no confunde una cita con un insight.

Introducción

Tener muchas conversaciones con clientes no equivale a aprender de ellas. Entre la última entrevista y la próxima decisión de producto hay un trabajo que a menudo se hace deprisa o no se hace: ordenar lo escuchado, ver qué se repite, entender por qué ocurre y decidir qué significa para el producto. Ese trabajo es la síntesis, y en él se concentra buena parte del valor diferencial de quien descubre.

En este capítulo veremos los niveles por los que asciende el análisis cualitativo, cómo se pasa de las citas a los patrones contando fuentes, qué hacer con los casos que no encajan y cómo se formula un insight que de verdad oriente una decisión. El ritual de síntesis en grupo, con su sala y sus paredes llenas de notas, tiene skill propia en Scrum Manager, Design Thinking; aquí nos ocupamos de su vertiente analítica.

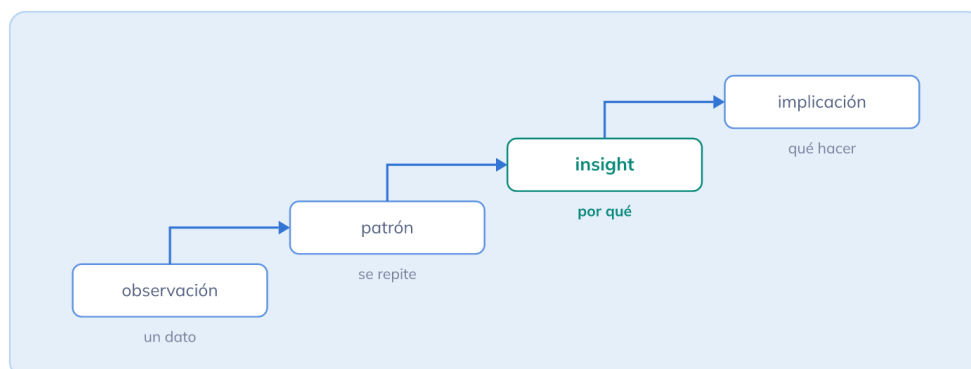


Figura 5. De la observación a la implicación: la síntesis asciende por niveles y el insight es el que conecta el patrón con una causa.

6.1 Los cuatro niveles del análisis

El análisis cualitativo asciende por niveles que se sostienen unos en otros. Nombrarlos ayuda a saber en qué punto está cada afirmación del equipo y a no presentar como conclusión lo que todavía es un dato suelto.

- **Observación.** Un dato de una fuente concreta: «Ana tarda treinta minutos en encontrar el contrato que necesita».
- **Patrón.** Lo que se repite en varias fuentes distintas: «seis de diez entrevistados describen fricción al buscar documentos».
- **Insight.** La comprensión nueva que conecta el patrón con una causa o una motivación: «pierden tiempo porque las carpetas reflejan la estructura del departamento y no la de los proyectos en que trabajan».
- **Implicación.** Lo que conviene hacer en consecuencia: «explorar una navegación por proyectos, además de la navegación por carpetas».

El error más frecuente consiste en llamar insight a una cita llamativa o a un patrón sin interpretar. Una comprobación sencilla ayuda a evitarlo: preguntarse qué hemos aprendido que no sabíamos antes y qué significa para el producto. Si falta la causa o falta la implicación, lo que tenemos todavía es una observación o, como mucho, un patrón.

6.2 De las citas a los patrones: agrupar y contar

El primer paso de la síntesis es agrupar las observaciones que tratan del mismo fenómeno. Conviene revisar todas las notas sin categorizarlas de antemano, dejar que los grupos aparezcan a partir del material y ponerles nombre después, que es la lógica del diagrama de afinidad (*affinity diagramming*) que describe el Nielsen Norman Group. Categorizar antes de leer hace que el material se ordene según lo que ya pensábamos.

Después se cuenta, y se cuentan fuentes distintas, no menciones. «Siete de doce entrevistados mencionaron el tiempo de espera» es mucho más útil que «varios usuarios hablaron de problemas de velocidad», porque permite comparar temas y priorizar. Como orientación, conviene ver un patrón en varias fuentes, tres o más, antes de afirmarlo; una sola voz, por intensa que sea, puede representar la necesidad de una minoría.

Conviene, además, que cada patrón conserve el rastro de las observaciones que lo sostienen, con la fuente y la fecha de cada una. Esa trazabilidad permite volver a la cita original cuando alguien discute el patrón, revisar la síntesis cuando llegan entrevistas nuevas y reutilizar lo aprendido meses después, que es la función del repositorio de evidencia del capítulo 14. Un patrón sin rastro acaba siendo una opinión del equipo con apariencia de dato.

Contar no es medir. Diez entrevistas bastan para descubrir que un patrón existe, pero no para estimar qué proporción de clientes lo comparte. Decir «el 70 % de los usuarios» a partir de siete de diez conversaciones convierte un indicio cualitativo en una cifra que no lo es. Para medir la magnitud hace falta otro instrumento, como una encuesta o un dato de uso.

6.3 Las excepciones como señal de segmento

Cuando la mayoría coincide y unos pocos casos no encajan, la tentación es descartarlos como ruido. Suele ser un error. Las excepciones revelan a menudo segmentos con necesidades distintas, usos que nadie había previsto o contextos organizativos que cambian el problema.

La pregunta útil no se dirige a cómo ignorar los casos que no encajan, sino a qué tienen en común entre sí. Si los tres entrevistados que no se quejan de la búsqueda trabajan en equipos pequeños, quizá el problema aparece con el volumen de documentos y la solución deba adaptarse a ese umbral. Documentar las excepciones con su contexto (rol, tamaño, situación) es lo que permite volver sobre ellas cuando el patrón principal deje de explicar lo que se observa.

A veces la excepción obliga a replantear el propio patrón. Si los entrevistados satisfechos son justo los que llevan más tiempo usando el producto, quizá la fricción está en el aprendizaje inicial y no en la búsqueda, y la oportunidad es otra. Por eso conviene revisar los grupos después de analizar las excepciones y no dar el mapa de patrones por cerrado a la primera.

6.4 Formular un insight y contrastarlo

Un insight bien formulado reúne tres piezas: un patrón cuantificado, una causa interpretada y una implicación concreta. «Observamos que cinco de ocho usuarios abandonan los contenidos largos porque sus jornadas fragmentadas no dejan sesiones extensas, y esto sugiere evaluar un formato de piezas breves» se puede discutir, comprobar y convertir en un experimento. «Los usuarios necesitan una experiencia más personalizada» no permite nada de eso, porque no tiene patrón contado, ni causa, ni implicación.

La causa es la pieza más frágil, porque es una interpretación del equipo y no un dato del cliente. Por eso conviene marcarla como tal y, cuando la decisión que depende de ella es cara, tratarla como un supuesto que hay que probar (capítulo 8). Una síntesis honesta distingue lo que se observó de lo que se infiere.

La síntesis con inteligencia artificial

Los asistentes de IA resumen transcripciones y agrupan temas en minutos, y eso ha creado una forma nueva del error de confundir cita e insight. Teresa Torres advierte contra la síntesis perezosa con IA (*lazy AI interview synthesis*), que consiste en volcar las transcripciones en un asistente y aceptar su resumen sin revisarlo. Su propuesta es

sintetizar primero, pedir después a la IA la misma síntesis y comparar las dos. El capítulo 16 desarrolla ese protocolo y sus límites.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Clasificar una afirmación del equipo en uno de los cuatro niveles del análisis (observación, patrón, insight o implicación).
- Distinguir un insight de una cita llamativa o de un patrón sin interpretar, comprobando que tiene causa e implicación.
- Agrupar observaciones por tema sin categorías previas y contar la recurrencia en número de fuentes distintas.
- Tratar los casos que no encajan como posible señal de otro segmento, identificando qué tienen en común entre sí.
- Formular un insight con patrón cuantificado, causa interpretada e implicación concreta, marcando la causa como inferencia.

Errores y antipatrones frecuentes

- **Lista de citas sin agrupar.** No deja ver patrones ni priorizar. Agrupa por tema y cuenta cuántas fuentes distintas respaldan cada uno.
- **Generalizar desde un caso aislado.** Una sola voz intensa puede representar a una minoría. Busca el patrón en varias fuentes antes de afirmarlo.
- **Insight sin implicación.** Si no orienta ninguna decisión, todavía no es accionable. Completa la frase con lo que conviene hacer en consecuencia.
- **Aceptar el resumen de la IA como síntesis.** El resumen generado es una interpretación más, sin la responsabilidad de nadie detrás. Haz tu síntesis, contrástala con la de la herramienta y decide tú.

Para profundizar

- [Nielsen Norman Group — Affinity diagramming](#)
- [Scrum Manager BoK — Affinity Mapping](#)
- [Scrum Manager BoK — Identify themes](#)
- [Skill Arena — Design Thinking](#)

BLOQUE B · DE LOS DATOS A LA COMPRENSIÓN

Capítulo 7. Enmarcar el problema: el trabajo que el cliente quiere hacer · ESTABLECIDO

Enunciar el problema como la falta de una funcionalidad esconde la solución dentro; enmarcarlo como el progreso que busca la persona deja abiertas varias soluciones.

Introducción

La forma en que se enuncia un problema condiciona todas las soluciones que se podrán imaginar después. El error más caro y más frecuente en producto es definirlo como la falta de una funcionalidad, porque cierra la exploración antes de empezar y convierte el trabajo del equipo en una cola de peticiones.

En este capítulo veremos por qué ese enunciado es una trampa, qué propone el marco de los *jobs to be done* (JTBD) y sus dos escuelas, cómo se escribe una *job story* útil y cómo se baja del síntoma que la persona describe a la causa que lo produce.



Figura 6. La job story y la alternativa actual: la situación, la motivación y el resultado describen el progreso que se busca, y la alternativa de hoy marca el listón que hay que superar.

7.1 La trampa de enunciar el problema como una funcionalidad

Cuando llega una petición como «necesitamos exportar a PDF», la tendencia es asumir que el problema es la ausencia de esa función. Pero «los usuarios no tienen un botón para exportar a PDF» lleva la solución dentro y cierra la indagación. Las preguntas que abren el problema son otras: para qué necesitan el PDF, a quién se lo envían y qué hacen con él después. Quizá lo que necesitan es compartir datos con personas de fuera, y un enlace de consulta resuelve mejor que un fichero.

Cuando un equipo solo recibe y ejecuta peticiones de este tipo, se convierte en lo que se ha llamado una fábrica de funcionalidades (*feature factory*): produce mucho y aprende poco, porque nunca pregunta qué progreso buscaba quien pidió cada cosa. Reformular la petición como «la persona intenta conseguir X, pero hoy se lo impide Y, porque Z» es el primer paso para salir de esa dinámica.

7.2 El marco de los trabajos por hacer y sus dos escuelas

La idea central del marco es que las personas no compran productos por sus características, sino que los contratan para hacer un progreso en una situación concreta. El foco se desplaza del perfil del cliente y de la funcionalidad pedida a la circunstancia y a lo que la persona intenta conseguir en ella. Bajo ese paraguas conviven dos escuelas que conviene distinguir, porque no usan las mismas herramientas.

- **El trabajo como progreso.** Clayton Christensen y Bob Moesta entienden el trabajo como el progreso que la persona intenta hacer en una circunstancia particular, y lo exploran reconstruyendo la historia de cómo alguien llegó a cambiar de solución. La obra de referencia es *Competing Against Luck* (2016).
- **El trabajo como resultado medible.** Tony Ulwick formula el trabajo como un verbo con su objeto y su contexto, y mide los resultados que el cliente desea obtener al hacerlo. Su método se llama innovación guiada por resultados (*Outcome-Driven Innovation*, ODI) y busca cuantificar qué resultados están peor servidos.

Un mismo trabajo tiene, además, varias dimensiones que conviene explorar en la entrevista: la funcional (lo que la persona necesita que ocurra), la emocional (cómo quiere sentirse) y la social (cómo quiere que la vean). Quien contrata una herramienta de gestión para «tener controlado el proyecto» busca a menudo también tranquilidad ante su responsable, y esa dimensión explica decisiones que la funcional no explica.

Para el trabajo diario de descubrimiento, la primera escuela aporta el lenguaje del progreso y de la situación, y la segunda, la disciplina de medir. No hace falta adherirse a una para aprovechar el marco, pero sí conviene saber de cuál procede cada herramienta, porque se citan a menudo mezcladas.

7.3 La job story y la alternativa actual

La forma más usada del marco en equipos de producto es la job story, que nació en Intercom con Paul Adams y que bautizó Alan Klement: «cuando [situación], quiero [motivación], para [resultado]». Sustituye el «quién eres» de la historia de usuario clásica por un «cuándo te pasa», porque en muchas necesidades la situación informa más que el perfil. Pertenece a la tradición del progreso, no a la de Ulwick.

«Cuando coordino entregas entre tres equipos en remoto, quiero ver el estado real de cada tarea, para anticipar bloqueos antes de la fecha comprometida» es específica y observable. «Cuando necesito gestionar mi trabajo, quiero una herramienta fácil, para ser más productivo» no dice nada: la situación es genérica y el resultado, circular.

A esas tres partes conviene añadirles la alternativa con que la persona resuelve hoy el problema: el correo, una hoja compartida, una llamada o, sencillamente, aguantarse. No forma parte de la plantilla original, pero es muy útil, porque marca el listón que cualquier solución tiene que superar y revela las restricciones reales del contexto.

Siempre hay una alternativa. Aunque el cliente diga que no hace nada, algo hace: convive con el problema, lo delega o lo resuelve a mano. Una solución que no supera claramente esa alternativa no provoca el cambio, por mucho que la idea guste en la entrevista.

7.4 Del síntoma a la causa raíz

Enmarcar bien exige excavar bajo el síntoma que la persona describe. El procedimiento es sencillo de enunciar y exigente de aplicar.

1. Recoger el síntoma tal como lo cuenta la persona («el sistema es lento»).
2. Explorar la situación con preguntas de seguimiento sobre el pasado concreto («¿qué intentabas hacer la última vez que lo notaste lento?»).
3. Enunciar la necesidad en términos de resultado («encontrar una información concreta a tiempo para decidir»).
4. Documentar el contexto y la alternativa actual.

A menudo «lento» resulta ser un problema de encontrar, no de rendimiento técnico: optimizar la base de datos no resolvería una frustración de búsqueda. Una comprobación útil es escribir, para el mismo enmarcado, tres soluciones que no se parezcan entre sí; si solo se nos ocurre una, o las tres son variantes de la misma funcionalidad, el enunciado todavía lleva la solución dentro. El enmarcado está bien hecho cuando deja abiertas varias soluciones posibles, y conviene comprobarlo explícitamente antes de pasar a idear. La reformulación creativa del reto, con sus preguntas del tipo «¿cómo podríamos...?», la desarrolla la skill Design Thinking.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Detectar un problema enunciado como la falta de una funcionalidad y reescribirlo sin prescribir la solución.
- Distinguir la escuela del trabajo como progreso (Christensen y Moesta) de la del trabajo como resultado medible (Ulwick) y atribuir a cada una sus herramientas.
- Redactar una job story con situación concreta, motivación y resultado observable, completada con la alternativa actual.
- Separar síntoma, causa y necesidad en una petición de cliente mediante preguntas de seguimiento sobre el pasado.
- Comprobar que un enmarcado deja abiertas varias soluciones posibles antes de pasar a idear.

Errores y antipatrones frecuentes

- **El problema como falta de funcionalidad.** «No tienen exportación a PDF» ya decidió la solución. Reformula hacia el objetivo: compartir datos con personas de fuera de forma que puedan consultarlos.

- **La job story circular.** «Quiero gestionar proyectos para ser más productivo» no orienta nada. Concreta la situación que dispara la necesidad y el resultado que se podría observar.
- **Atribuir la plantilla a quien no la creó.** La job story es de Intercom y Klement, en la tradición del progreso; el método de Ulwick formula el trabajo de otra manera. Mezclarlas confunde a quien quiera profundizar.
- **Ignorar la alternativa actual.** Sin ella no se sabe qué hay que superar. Pregunta siempre cómo resuelve hoy la persona el problema, aunque la respuesta sea que lo aguanta.

Para profundizar

- [Intercom — Job stories](#)
- [Christensen Institute — Jobs to Be Done](#)
- [Strategyn — Outcome-Driven Innovation](#)
- [Scrum Manager BoK — Job story](#)

BLOQUE C · DISEÑAR EXPERIMENTOS PARA APRENDER

Capítulo 8. Mapeo de supuestos: los riesgos que hay que probar primero · ESTABLECIDO

Toda idea descansa sobre supuestos que pueden resultar falsos; hacerlos explícitos y ordenarlos por importancia y evidencia evita probar lo fácil y dejar sin probar lo crítico.

Introducción

Antes de diseñar un experimento conviene saber qué hay que probar y en qué orden. Una idea de producto, por prometedora que parezca, se apoya en un conjunto de creencias que el equipo da por ciertas sin haberlas comprobado: que el problema existe, que alguien pagará, que se puede construir, que la gente sabrá usarlo. Si una de ellas es falsa y es de las que sostienen la idea, todo lo demás se viene abajo.

En este capítulo veremos qué tipos de supuesto distinguen las fuentes de referencia, cómo se priorizan con la matriz de importancia y evidencia de David Bland, por qué el orden habitual no es una regla fija y cómo se convierte un supuesto vago en uno que se puede probar.

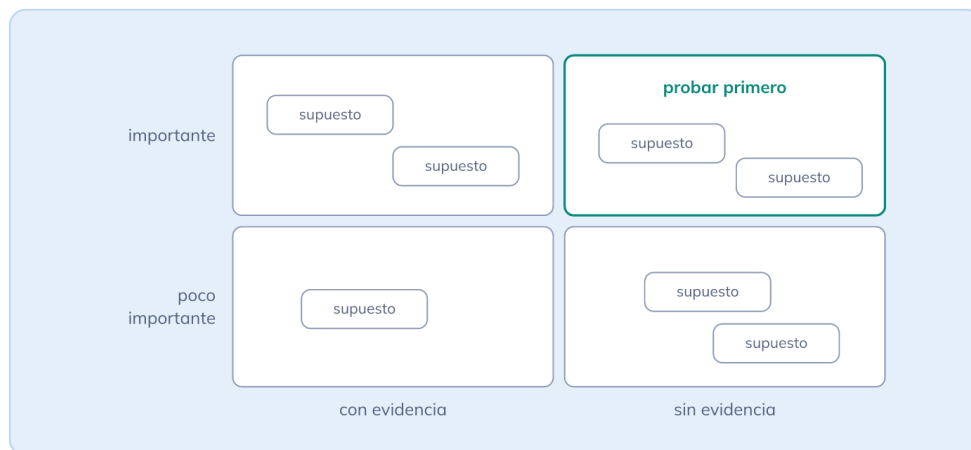


Figura 7. El mapa de supuestos: se prueba primero lo que es importante y todavía no tiene evidencia.

8.1 Los tipos de supuesto

La clasificación más difundida distingue tres tipos, que responden a tres preguntas sobre la idea.

- **Deseabilidad.** ¿Lo quieren? El problema es real y prioritario para alguien, y esa persona cambiará lo que hace hoy para adoptar la solución.
- **Viabilidad.** ¿Es un negocio? Hay quien pague el precio, el coste de conseguir clientes es asumible y los márgenes sostienen la inversión.
- **Factibilidad.** ¿Podemos construirlo? El equipo tiene las capacidades, la tecnología responde y la solución se integra con lo que ya existe.

Las fuentes de referencia amplían esta lista. Teresa Torres añade la usabilidad (si el cliente sabrá usar la solución) y la ética (si construirla puede causar algún daño). David

Bland incluye la adaptabilidad, que pregunta si la idea resistirá los cambios del entorno. Marty Cagan habla de cuatro riesgos (valor, usabilidad, factibilidad y viabilidad). Las listas no compiten entre sí; lo que comparten es la obligación de mirar más allá de la pregunta técnica.

8.2 La matriz de importancia y evidencia

El mapeo de supuestos (*Assumptions Mapping*) es el ejercicio que Bland difundió en *Testing Business Ideas*, escrito con Alexander Osterwalder, y que él mismo atribuye a su trabajo con Jeff Gothelf y Josh Seiden, autores de *Lean UX*. Consiste en listar los supuestos de la idea y colocarlos en una matriz con dos ejes, en tres pasos.

1. Escribir cada supuesto en una nota, uno por nota, completando la frase «para que esta idea funcione, tiene que ser cierto que...».
2. Colocar cada nota según su importancia y según la evidencia que hay a su favor.
3. Elegir los supuestos del cuadrante de lo importante sin evidencia y diseñar para cada uno la prueba más pequeña que aporte evidencia.

Los dos ejes se definen así.

- **Importancia.** Cuánto depende la idea de ese supuesto; si resultara falso, ¿fracasaría la iniciativa o bastaría con ajustarla?
- **Evidencia.** Cuánta evidencia tenemos hoy de que es cierto; desde ninguna, porque solo lo creemos, hasta mucha, porque lo hemos observado.

El cuadrante que manda es el de los supuestos importantes para los que no hay evidencia. Se prueban primero, aunque sean incómodos de probar, porque es ahí donde un error sale más caro. Los importantes con evidencia se vigilan, y los poco importantes pueden esperar. Hacer el ejercicio con todos los roles presentes (producto, diseño, desarrollo y negocio) saca a la luz supuestos que cada uno daba por obvios.

Los supuestos que nadie cuestiona. Los más peligrosos suelen ser los que vienen de la dirección y nadie se atreve a poner en la matriz: «los clientes cambiarán sin resistencia», «el boca a boca bastará para crecer». Colocarlos en el mapa los convierte en algo que se puede evaluar en lugar de un punto ciego.

8.3 El orden habitual y cuándo se invierte

Para la mayoría de los productos nuevos, el orden habitual es empezar por la deseabilidad, seguir por la viabilidad y dejar la factibilidad para el final. El motivo es práctico: el patrón de fallo más común consiste en probar primero lo técnico, porque es lo cómodo y se puede hacer sin salir del equipo, y descubrir tarde que nadie quería el resultado.

Ese orden, sin embargo, es una consecuencia de la matriz y no una regla que la sustituya. Si en un caso concreto el supuesto importante y sin evidencia es de factibilidad, se prueba primero ese. Pasa con frecuencia en los productos cuyo valor

depende de un modelo de IA, donde la pregunta de si el sistema responderá con suficiente calidad puede decidirlo todo; Torres ha llegado a presentar las evaluaciones de esos sistemas como un nuevo hábito del descubrimiento. El capítulo 18 trata ese caso.

Una consecuencia práctica de trabajar así es que se prueban supuestos y no ideas completas. En lugar de construir la solución entera para ver si funciona, se aísla el supuesto crítico y se busca la prueba más pequeña que lo ponga a prueba, que puede ser una pregunta en una entrevista, un prototipo de una pantalla o un dato que ya existe. El árbol de oportunidades del capítulo 15 organiza estas pruebas por solución y por oportunidad.

8.4 Del supuesto vago al supuesto que se puede probar

Un supuesto que no se puede confirmar ni refutar no sirve para decidir. «Los usuarios quieren una herramienta más sencilla» es una opinión razonable, pero no dice qué habría que observar para darla por buena o por falsa. Reformulado como «al menos seis de cada diez responsables de compras que prueben el flujo nuevo completarán un pedido sin ayuda», ya se puede diseñar una prueba y fijar su criterio de éxito (capítulo 11).

La reformulación pide nombrar a quién afecta el supuesto, qué comportamiento concreto lo pondría de manifiesto y a partir de qué umbral lo consideraremos cierto. Cuanto más cara sea la decisión que depende del supuesto, más precisa debe ser su formulación.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Clasificar un supuesto como de deseabilidad, viabilidad, factibilidad, usabilidad, ética o adaptabilidad.
- Colocar un conjunto de supuestos en la matriz de importancia y evidencia de Bland y justificar cuáles se prueban primero.
- Reconocer cuándo el orden habitual (deseabilidad primero) debe invertirse porque el supuesto crítico es de otro tipo.
- Hacer explícitos en el mapa los supuestos implícitos de la dirección o de la cultura del equipo.
- Reformular un supuesto vago en uno que nombra a quién afecta, qué comportamiento lo confirmaría y con qué umbral.

Errores y antipatrones frecuentes

- **Probar lo fácil y no lo crítico.** Los supuestos que pueden hundir la idea quedan sin probar. Deja que la matriz, y no la comodidad, decida el orden.
- **Convertir el orden habitual en dogma.** Aplicar siempre «deseabilidad primero» ignora los casos en que el supuesto letal es técnico. Manda el cuadrante de lo importante sin evidencia.

- **Supuestos implícitos que nunca se listan.** Producen puntos ciegos, sobre todo los que vienen de la dirección. Haz el mapeo con todos los roles y pon también esos supuestos en la matriz.
- **Supuestos que no se pueden probar.** Lo que no se puede confirmar ni refutar no ayuda a decidir. Reformúlalo con un comportamiento observable y un umbral.

Para profundizar

- [David Bland — Assumptions Mapping](#)
- [Teresa Torres — Assumption testing](#)
- [Marty Cagan — The four big risks](#)
- [Bland y Osterwalder — Testing Business Ideas](#)

BLOQUE C · DISEÑAR EXPERIMENTOS PARA APRENDER

Capítulo 9. Tests de prototipo: tareas como objetivos, no instrucciones · EN CONSOLIDACIÓN

Un test de prototipo mide el diseño solo si las tareas son objetivos, modera alguien neutral, participan personas representativas y se observa lo que hacen.

Introducción

Probar un prototipo con usuarios parece sencillo, pero muchos tests producen señal contaminada por defectos del propio test. Las prácticas correctas son conocidas desde hace años; lo que sigue siendo desigual es su aplicación disciplinada, y por eso esta ficha está en consolidación.

En este capítulo veremos las decisiones que separan un test que mide el diseño de uno que mide otra cosa: cómo se plantean las tareas, quién modera, qué se observa y qué fidelidad necesita el prototipo, también ahora que la IA genera prototipos pulidos en minutos. Quién participa en el test es una decisión previa que el capítulo 3 trata a fondo.

9.1 Tareas como objetivos, no como instrucciones

El error más común es redactar tareas que dicen cómo completarlas. «Pulsa el botón verde de arriba a la derecha, rellena el formulario y guarda con el icono del disquete» solo evalúa si la persona sabe seguir instrucciones. «Crea un proyecto llamado Prueba Q1» revela si el diseño se entiende, dónde se confunde la persona y qué modelo mental aplica.

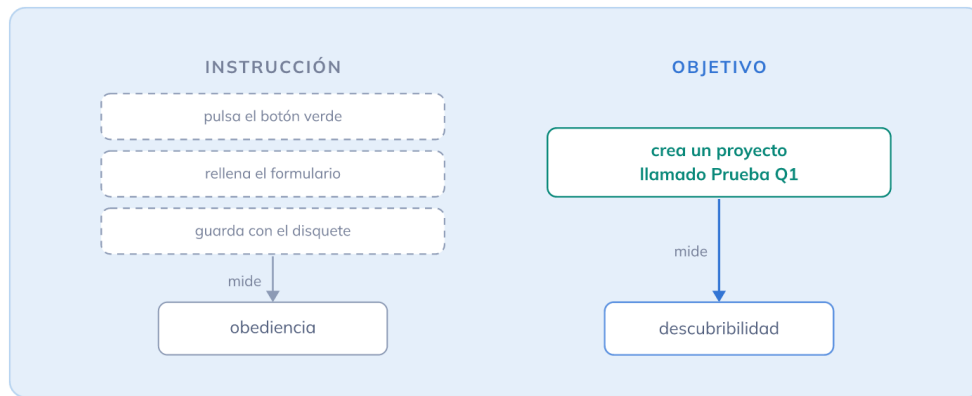


Figura 8. Dos formas de pedir la misma tarea: la instrucción paso a paso mide obediencia y el objetivo sin método mide si el diseño se deja descubrir.

La regla práctica es enunciar el objetivo sin indicar el camino y observar el que toma la persona. Conviene revisar el guion antes del test buscando las palabras que aparecen en la interfaz: si la tarea usa el mismo rótulo que el botón que hay que pulsar, está dando la respuesta.

9.2 Quién modera y quién observa

Cuando modera quien diseñó el prototipo, introduce sesgos sutiles que invalidan la señal.

- El interés emocional en que el diseño funcione.
- La guía inconsciente hacia el camino correcto, con gestos, silencios o preguntas que lo sugieren.
- La tendencia a atribuir la confusión a la persona y no al diseño.
- La ayuda excesiva en cuanto la persona se atasca, que borra justo lo que se quería observar.

La solución es separar los papeles: una persona neutral conduce la sesión y quien diseñó observa sin intervenir. Igual de importante es con quién se prueba. Compañeros, amigos o clientes entusiastas conocen el contexto y quieren agradar, así que el test necesita participantes externos y representativos del segmento, reclutados con los criterios que describe el capítulo 3.

En cuanto al número de sesiones, Jakob Nielsen defendió desde el Nielsen Norman Group que unas cinco personas por ronda bastan para encontrar la mayoría de los problemas de un diseño, y que conviene más hacer varias rondas pequeñas, corrigiendo entre una y otra, que una sola grande. La lógica es la del descubrimiento, porque el test sirve para encontrar problemas y no para medir con precisión cuántas personas los sufren.

9.3 Observar lo que la persona hace

El criterio de éxito de un test de usabilidad es el comportamiento observable, no la opinión. «¿Te gusta el diseño?» pertenece a otra conversación. Lo que se registra es otra cosa.

- Dónde pulsa primero y si ese primer paso lleva hacia el objetivo.
- Dónde se detiene, cuánto tarda y en qué punto abandona.
- Qué elementos no encuentra y cuáles interpreta de forma distinta a la prevista.
- Qué dice mientras trabaja, si se le pide que piense en voz alta.

Conviene preparar un guion breve para cada sesión, con una introducción que tranquilice a la persona (se evalúa el diseño y no a ella), un escenario realista que dé sentido a las tareas, las tareas en orden de dificultad y unas preguntas de cierre sobre lo que ha echado en falta. El guion se prueba antes con alguien del equipo para detectar tareas ambiguas o que delatan la respuesta.

El protocolo de pensar en voz alta (*think-aloud protocol*), en el que la persona verbaliza lo que busca y lo que espera mientras realiza la tarea, es la técnica clásica para acceder a su razonamiento sin preguntarle después por sus impresiones. Steve Krug lo resume en *Rocket Surgery Made Easy* como la forma más barata de encontrar los problemas graves de un diseño con pocas sesiones.

9.4 La fidelidad que responde a la pregunta

El nivel de fidelidad del prototipo debe corresponder a lo que se quiere aprender. Construir alta fidelidad para todo desperdicia tiempo y, además, cambia la conversación.

- Para saber si un concepto se entiende bastan bocetos o esquemas de pantalla.
- Para explorar un flujo de navegación, un prototipo navegable de fidelidad media.
- Para evaluar interacciones concretas, alta fidelidad.
- Para probar la propuesta de valor, a menudo basta una página que la describa (capítulo 10).

Prototipos y moderadores con inteligencia artificial

Las herramientas de IA generan hoy prototipos navegables de aspecto terminado en minutos, y eso trae una tentación nueva. Un prototipo que parece acabado invita a opinar sobre el acabado y a encariñarse con él, tanto a quien lo prueba como a quien lo hizo, así que la regla de la fidelidad mínima se vuelve más necesaria, no menos. El prototipado como forma de pensar lo desarrolla la skill Design Thinking.

También llegan agentes y moderadores automáticos que prometen evaluar la usabilidad sin participantes o sin moderador humano. La investigación que entrena agentes para esa tarea avanza, como muestra el trabajo de Gao y su equipo publicado en abril de 2026, pero hoy no sustituye la observación de personas reales del segmento, que es la que revela los malentendidos que nadie había previsto.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Reescribir una tarea de test redactada como instrucción paso a paso en una tarea planteada como objetivo.

- Asignar los papeles de moderación y observación de modo que quien diseñó el prototipo no conduzca la sesión.
- Definir para un test criterios de éxito observables (primer paso, tiempos, abandonos, elementos no encontrados) en lugar de preguntas de opinión.
- Aplicar el protocolo de pensar en voz alta sin sugerir el camino a la persona participante.
- Elegir la fidelidad mínima del prototipo que responde a la pregunta de investigación, también cuando la IA permite más.

Errores y antipatrones frecuentes

- **Tareas guiadas.** Si la tarea menciona ubicaciones o rótulos de la interfaz, el test mide obediencia. Reformúlala como objetivo y deja que la persona encuentre el camino.
- **El diseñador modera.** «Modero yo, que conozco mejor el diseño» introduce sesgo en cada gesto. Que modere alguien neutral y que el diseñador observe.
- **Pedir opiniones.** «¿Qué te parece?» no dice nada de la usabilidad. Diseña tareas que produzcan comportamiento observable.
- **Alta fidelidad porque es gratis.** Que la IA la genere en minutos no la hace adecuada: desplaza la conversación al acabado. Usa la fidelidad que responde a tu pregunta.

Para profundizar

- [Steve Krug — Rocket Surgery Made Easy](#)
- [Nielsen Norman Group — Usability testing 101](#)
- [Scrum Manager BoK — Think-aloud protocol](#)
- [Skill Arena — Design Thinking](#)

BLOQUE C · DISEÑAR EXPERIMENTOS PARA APRENDER

Capítulo 10. El MVP como experimento, no como producto reducido · EN CONSOLIDACIÓN

Un MVP es lo mínimo necesario para aprender algo concreto; a menudo se aprende mejor sin escribir código, y construir barato no exime de preguntar qué se quiere aprender.

Introducción

Pocos términos de producto se malinterpretan tanto como el de producto mínimo viable (*minimum viable product*, MVP). Entendido como un producto con menos funcionalidades, produce listas de componentes disfrazadas de experimento. Entendido como un instrumento para aprender, ahorra meses de desarrollo. La práctica gana

adopción, pero todavía convive con el reflejo de construir primero y ver qué pasa, y la IA ha dado a ese reflejo una fuerza nueva.

En este capítulo veremos el MVP como experimento, las alternativas que no requieren construir el producto, cómo se elige cada una según la hipótesis y por qué el abaratamiento de la construcción no cambia la pregunta de fondo.

10.1 Un experimento, no una versión pequeña de todo

Un MVP es el mínimo necesario para aprender algo específico, no el mínimo necesario para tener un producto. La diferencia se ve comparando dos planteamientos.

Un plan que incluye registro, perfil, notificaciones, panel de control e integraciones, con la hipótesis de que los usuarios usarán la plataforma y el criterio de publicarla en seis semanas, es una lista de componentes con una hipótesis genérica y sin criterio de validación. Un plan que supone que los equipos de finanzas pagarán por automatizar la conciliación bancaria, lo prueba con un servicio manual para cinco clientes durante dos semanas y da la hipótesis por buena si al menos tres de los cinco piden continuar y pagar es un experimento, con su hipótesis, su instrumento y su umbral fijado de antemano (capítulo 11).

La palabra «mínimo» se refiere, por tanto, a lo que hace falta para aprender y no a lo que hace falta para lanzar. Un MVP puede ser mucho más pequeño que la primera versión del producto, y a menudo no se parece en nada a ella; puede ser un formulario, una hoja de cálculo operada a mano o una llamada semanal con cinco clientes.

10.2 Las alternativas antes de escribir código

Antes de construir conviene preguntarse si la hipótesis se puede poner a prueba con algo más ligero. *Testing Business Ideas* recoge un repertorio amplio de experimentos; cuatro son especialmente útiles en el descubrimiento temprano.

- **Servicio manual a la vista (*concierge*).** Se entrega el servicio a mano y el cliente lo sabe. Pone a prueba la demanda real y la disposición a pagar antes de automatizar nada.
- **Mago de Oz.** Una interfaz aparentemente automática que opera por detrás una persona. Permite probar la experiencia sin haber desarrollado el sistema.
- **Página de captación (*landing page*).** Describe la propuesta y recoge una señal de interés, como un correo, una reserva o un pago por adelantado, antes de construir nada.
- **Una sola funcionalidad (*single feature MVP*).** El mínimo código necesario para poner a prueba el valor de una funcionalidad concreta.

Argumentar contra las alternativas manuales porque no escalan desperdicia la oportunidad de aprender deprisa y barato. Si nadie quiere el servicio manual, no hay nada que escalar; si lo quieren, se habrá validado la demanda antes de invertir en automatizar.

La cautela del Mago de Oz

El Mago de Oz oculta a la persona que hay alguien detrás, y esa ocultación tiene límites. Conviene que el experimento no induzca a decisiones que la persona no tomaría si supiera cómo funciona, que los datos que comparta se traten con el cuidado del capítulo 5 y que se le explique el funcionamiento al terminar. Si parte de las respuestas las genera de verdad un sistema de IA, desde el 2 de agosto de 2026 el artículo 50 del Reglamento europeo de IA obliga a que la persona sepa que interactúa con una IA.

10.3 Elegir el tipo según la hipótesis

Cada pregunta pide un instrumento distinto, y elegirlo según lo que se quiere aprender, y no por costumbre, es lo que mantiene el MVP como experimento.

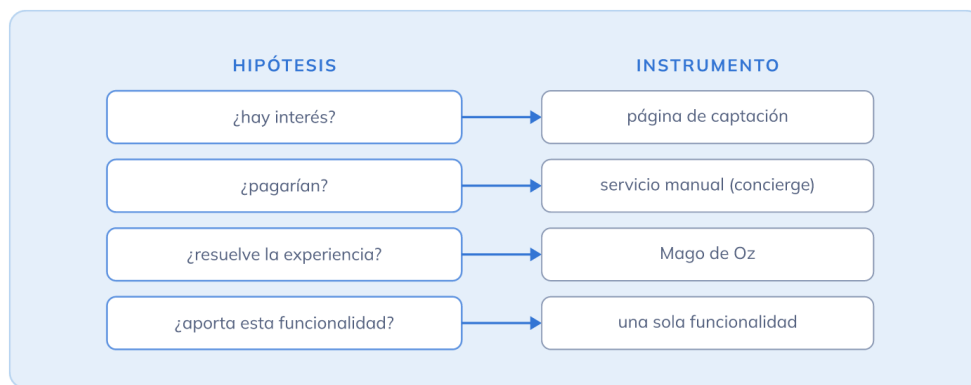


Figura 9. Cada pregunta, su instrumento: el tipo de MVP se elige por la hipótesis que se quiere poner a prueba y no por costumbre.

Conviene recordar que la señal que produce cada instrumento no pesa lo mismo. Una página de captación indica interés inicial, pero un correo dejado en un formulario es una señal débil; una reserva o un pago por adelantado pesan mucho más. El capítulo 12 desarrolla esa escala de la fuerza de la evidencia, que es la que permite interpretar el resultado de cualquiera de estos experimentos.

Los instrumentos se encadenan a menudo, de lo barato a lo más costoso. Una página de captación muestra si la propuesta despierta interés; si lo despierta, un servicio manual con unos pocos clientes muestra si pagan y qué esperan exactamente, y solo entonces tiene sentido construir la primera funcionalidad. Cada paso compra el derecho a invertir en el siguiente.

10.4 Construir barato no es aprender barato

Las herramientas de IA han abaratado la construcción de prototipos funcionales, y eso cambia el cálculo de coste que justificaba las alternativas manuales. Marty Cagan distingue construir para aprender, es decir, crear prototipos para evaluar los riesgos de valor, usabilidad, viabilidad y factibilidad, de construir para ganar, que es producir *software* con la calidad necesaria para venderlo y mantenerlo. Con las herramientas actuales, escribía en abril de 2026, diez o veinte prototipos a la semana están al alcance de casi cualquiera.

El riesgo es confundir velocidad con dirección. En julio de 2026 Cagan citaba a Hilary Gridley: como nunca ha sido tan rápido construir, nunca ha sido tan fácil correr diez veces más deprisa en la dirección equivocada. Para Teresa Torres, que construir sea muy barato no significa que debamos construir cada idea que tenemos.

El prototipo funcional generado con IA pasa a ser una opción legítima del repertorio, pero el criterio sigue siendo el mismo, lo mínimo para aprender algo concreto. Y aparece un peligro propio de este instrumento, que un prototipo que funciona se confunda con el producto y se empiece a desarrollar sobre él sin haber aprendido lo que se buscaba.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Distinguir un MVP planteado como experimento de uno planteado como producto reducido o como lista de componentes.
- Formular la hipótesis principal de un MVP y la señal que la confirmaría o refutaría antes de construir.
- Elegir entre servicio manual, Mago de Oz, página de captación, una sola funcionalidad o prototipo funcional según la hipótesis.
- Rebatir el argumento de que una alternativa manual no escala cuando se está en una etapa temprana.
- Aplicar las cautelas éticas y legales de un experimento que oculta cómo funciona o que incluye respuestas generadas por IA.

Errores y antipatrones frecuentes

- **El MVP como lista de funcionalidades.** Provoca crecimiento del alcance y pérdida de velocidad. Define primero la hipótesis y después el mínimo para ponerla a prueba.
- **Construir porque es barato.** Que la IA genere el prototipo en una tarde no dice qué se quiere aprender con él. Pregunta antes qué hipótesis pone a prueba.
- **Descartar las alternativas manuales.** Lleva a invertir cuando un servicio a mano habría bastado. Pregunta siempre si se puede aprender lo mismo sin código.
- **Tomar el interés por demanda.** Un correo en una página de captación es una señal débil. Interpreta el resultado con la escala de la fuerza de la evidencia.

Para profundizar

- [Marty Cagan — Build to learn vs build to earn](#)
- [Marty Cagan — The AI productivity paradox](#)
- [Bland y Osterwalder — Testing Business Ideas](#)
- [Scrum Manager BoK — Mínimo Producto Viable](#)

BLOQUE D · DECIDIR CON EVIDENCIA

Capítulo 11. Criterios de éxito y señales definidos de antemano

· ESTABLECIDO

El criterio de éxito se fija antes de lanzar el experimento, porque decidir después qué resultado parece bueno permite racionalizar cualquier resultado.

Introducción

Un experimento sin criterio previo de éxito se parece más a una historia que se contará a conveniencia cuando lleguen los datos que a una prueba. En el descubrimiento la tentación es especialmente fuerte, porque los equipos suelen tener pocas observaciones, mucha ilusión puesta en una idea y ninguna obligación formal de escribir antes lo que esperan ver.

En este capítulo veremos el principio de definición previa, la jerarquía de métricas que conviene distinguir, los límites que imponen el tamaño de la muestra y la ausencia de grupo de control, y cómo declarar con honestidad lo que un resultado permite concluir. Es donde el descubrimiento se vuelve riguroso o deja de serlo.

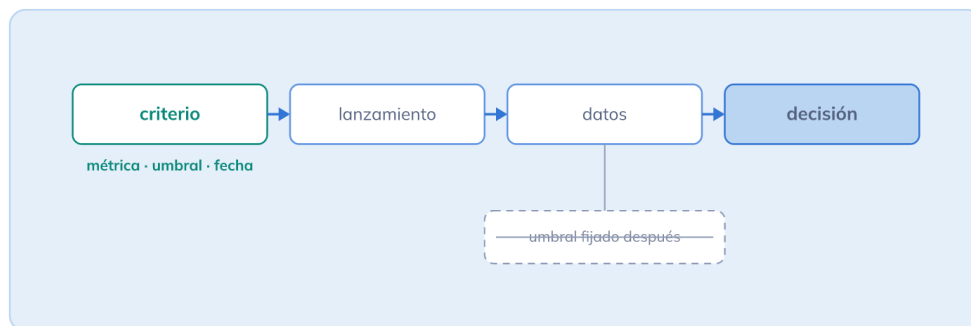


Figura 10. El criterio se fija antes de lanzar: el umbral de éxito queda escrito antes de que lleguen los datos.

11.1 El principio de definición previa

El criterio de éxito se escribe antes de lanzar, con su fecha, y nunca después de ver los resultados. La razón es sencilla. Con los datos delante, casi cualquier número parece bueno si se busca el ángulo adecuado, y el equipo que más desea que la idea funcione es el peor situado para fijar el listón a posteriori. Teresa Torres lo resume en la imagen de trazar una línea en la arena antes de probar el supuesto, de modo que el resultado caiga a un lado o al otro sin discusión.

Esta tentación tiene nombre en estadística. Se llama *p-hacking* a analizar los datos de varias maneras hasta dar con la que produce un resultado favorable, y *HARKing* a formular la hipótesis a la vista de los resultados como si se hubiera planteado antes; el término lo acuñó el psicólogo Norbert Kerr en 1998. Andrew Gelman y Eric Loken añadieron una observación incómoda con su «jardín de los senderos que se bifurcan»: no hace falta mala fe, porque basta con que las decisiones de análisis dependan de los datos para que la conclusión quede sesgada.

Un criterio completo. Lleva métrica primaria, umbral, tamaño mínimo de la muestra y fecha: «conversión de la página de reserva igual o superior al 5 % con al menos 400 visitas, medida del 3 al 17 de marzo y fijada el 1 de marzo». Lo contrario sería «mediremos visitas, clics y conversiones y decidiremos luego qué nos parece bueno».

Incluso un criterio bien escrito tiene límites que conviene conocer. Con 400 visitas, una conversión observada del 5 % es compatible con una conversión real que va aproximadamente del 3 al 7,5 %. El umbral sirve para decidir si se sigue adelante, y el resultado es una orientación de la demanda, que conviene leer como tal y no como una medición precisa.

11.2 La jerarquía de métricas

Un experimento bien diseñado distingue funciones entre las métricas que mira, y no todas pesan igual en la decisión.

- **Métrica primaria.** Es la que decide el éxito o el fracaso, y debe ser una sola, con su umbral.
- **Métricas secundarias.** Aportan contexto sobre el resultado sin cambiar la decisión.
- **Métricas de protección** (*guardrail metrics*). Son límites que no deben cruzarse aunque la primaria salga bien, y sirven para detectar daños colaterales, como una conversión que sube a costa de las devoluciones.
- **Métricas de diagnóstico.** Ayudan a entender por qué ocurrió lo que ocurrió y a preparar la siguiente iteración.

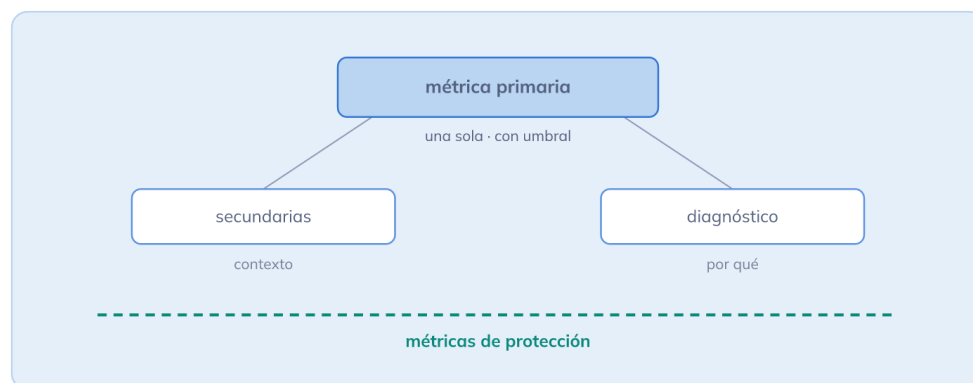


Figura 11. La jerarquía de métricas: una sola métrica primaria decide, las secundarias y las de diagnóstico dan contexto, y las de protección marcan lo que no debe romperse.

La confusión típica es tener varias métricas primarias. Con tres candidatas a decidir, basta con que una salga bien para declarar el éxito, y el criterio previo pierde su sentido por la puerta de atrás. Elegir una sola obliga además a una conversación valiosa, la de qué es exactamente lo que el experimento quiere aprender.

11.3 Tamaño de muestra y efecto mínimo detectable

Con muestras pequeñas, las fluctuaciones normales parecen señales. Antes de lanzar conviene preguntarse qué diferencia se podría detectar de forma fiable con la muestra esperada, lo que en experimentación se llama efecto mínimo detectable. Un umbral de «cualquier mejora» con cincuenta usuarios confunde el ruido con la señal, porque una variación de unos pocos puntos cabe de sobra en el azar.

En el descubrimiento temprano la muestra casi siempre es pequeña, y eso no invalida el experimento, aunque sí cambia lo que se puede afirmar con él. Un resultado con pocas observaciones orienta la siguiente decisión y dice poco sobre el tamaño del mercado. Declararlo así protege al equipo de la sobreconfianza y a la organización de inversiones apoyadas en un número frágil.

11.4 El grupo de control y la evidencia direccional

Sin grupo de control no se puede atribuir un cambio a la intervención. Un aumento de sesiones tras activar una funcionalidad puede deberse a la estacionalidad, a otra campaña o a factores que nadie controla. El diseño mínimo para atribuir causalidad incluye un grupo de tratamiento, un grupo de control equivalente y una asignación aleatoria entre ambos.

El matiz que conviene no perder es de escala. Un test A/B fiable necesita un tráfico que la mayoría de los descubrimientos tempranos no tiene. Ron Kohavi, Diane Tang y Ya Xu explican en *Trustworthy Online Controlled Experiments* cuánto tráfico exige un experimento controlado para que sus conclusiones sean fiables, y por qué con pocos usuarios las diferencias que interesan quedan por debajo de lo que el experimento puede distinguir.

Evidencia direccional. Cuando no hay control ni tráfico suficiente, lo honesto es declarar la evidencia como direccional: «observamos que..., lo que sugiere que...». Es una frase menos contundente, pero es la que el dato sostiene, y deja claro qué haría falta para convertir el indicio en una conclusión.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Definir el criterio de éxito de un experimento (métrica primaria, umbral, tamaño mínimo de la muestra y fecha) antes de lanzarlo.
- Clasificar las métricas de un experimento en primaria, secundarias, de protección y de diagnóstico, y exigir una sola primaria.
- Estimar si el tamaño de muestra previsto permite detectar el efecto buscado y ajustar en consecuencia lo que se afirmará con el resultado.
- Reconocer el p-hacking y el HARKing en un informe de resultados y reconstruir el criterio que se debió fijar antes.

- Distinguir un resultado causal, con grupo de control y asignación aleatoria, de uno direccional, y redactar cada uno con su grado de certeza.

Errores y antipatrones frecuentes

- **Fijar el umbral tras ver los datos.** Es la forma más común de p-hacking, y el listón acaba donde cayó el resultado. Documenta el criterio con fecha anterior al lanzamiento.
- **Varias métricas primarias.** Basta con que una salga bien para declarar el éxito. Elige una sola y deja las demás como contexto.
- **Umbral de «cualquier mejora».** Con muestras pequeñas confunde el ruido con la señal. Calcula antes qué diferencia puede detectar la muestra que vas a tener.
- **Presentar un indicio como causa.** «Tras activar X subió Y, luego X causó Y» es un salto sin grupo de control. Declara la evidencia como direccional.

Para profundizar

- [Teresa Torres — Assumption testing](#)
- [Kohavi, Tang y Xu — Trustworthy Online Controlled Experiments](#)

BLOQUE D · DECIDIR CON EVIDENCIA

Capítulo 12. La fuerza de la evidencia y el control de sesgos · ESTABLECIDO

La evidencia pesa más cuanto más se parece a una decisión real con coste para quien la toma, y cuatro sesgos frecuentes la erosionan.

Introducción

Recoger datos no equivale a tener evidencia. Un equipo puede acumular entrevistas, encuestas, comentarios y métricas y seguir sin saber si alguien pagaría por lo que propone, porque casi todo lo que ha recogido es lo que la gente dice y muy poco lo que la gente hace.

En este capítulo veremos cómo graduar la fuerza de la evidencia, qué señales parecen validación sin serlo, los cuatro sesgos que con más frecuencia la erosionan y dónde queda la frontera entre correlación y causalidad. Es el filtro que separa el descubrimiento riguroso del teatro de validación.

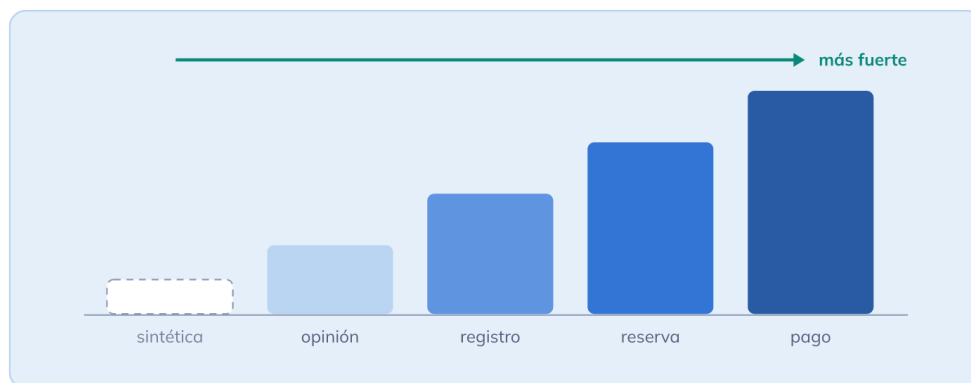


Figura 12. La fuerza de la evidencia: la evidencia pesa más cuanto más se parece lo observado a una decisión real con coste para quien la toma.

12.1 La evidencia como gradiente

Alexander Osterwalder formuló en 2017 una idea que ordena todo lo demás. Lo que los clientes dicen y lo que hacen son cosas distintas, y la evidencia se fortalece a medida que se acerca a una compra real. Entre una opinión expresada en una entrevista y un pago hay muchos peldaños intermedios, y conviene saber en cuál está cada dato antes de decidir cuánto pesa.

- **Lo que se dice.** Opiniones, intenciones declaradas y respuestas a encuestas. Son rápidas y baratas de obtener, y también las más débiles.
- **Lo que se registra.** Un correo dejado en una página de captación o una inscripción en una lista de espera. Hay un pequeño gesto, casi sin coste para quien lo hace.
- **Lo que se reserva.** Una reserva, una carta de intención o una cita comprometida. El cliente pone en juego tiempo, reputación o dinero futuro.

- **Lo que se paga.** Un prepago o una compra, idealmente de alguien que no sabe que participa en un experimento. Es lo más cercano a la decisión real.

La evidencia débil tiene su uso. Las entrevistas, que ocupan el peldaño más bajo como evidencia de demanda, son la mejor fuente para entender el problema, como vimos en el bloque A. Lo que pide el gradiente es emplear cada instrumento para la pregunta que responde bien y no sumar sus resultados como si pesaran lo mismo.

El gradiente resuelve además una contradicción frecuente. La captura de correos sirve para medir el interés inicial, como vimos en el capítulo 10, y a la vez es una evidencia débil de demanda. Las dos cosas son ciertas, porque el correo es el instrumento adecuado para saber si a alguien le interesa la propuesta y un instrumento insuficiente para saber si la pagará.

12.2 Señales que parecen validación

Algunas señales de bajo coste se confunden a menudo con validación. Los me gusta y los comentarios en redes no implican uso ni disposición a pagar, y cuarenta y siete reacciones a una publicación dicen muy poco sobre la demanda. El feedback de la red propia (amigos, contactos, suscriptores) llega con un incentivo social para ser amable, y las aperturas de un correo indican una lista activa, lo que informa poco del interés por un producto nuevo.

La pregunta de calidad. Ante cualquier dato que se presente como validación conviene preguntar si es un comportamiento observable con coste para quien lo produce o una opinión que no le cuesta nada. La respuesta decide cuánto peso merece.

En 2026 ha aparecido una fuente nueva de falso positivo, las respuestas de usuarios simulados por IA. Son opinión sin coste y, además, complaciente, así que ocupan el peldaño más bajo del gradiente, por debajo incluso de la opinión de un cliente real. El capítulo 17 las trata con detalle.

12.3 Cuatro sesgos frecuentes

Muchos sesgos cognitivos afectan a la evidencia de producto, y cuatro aparecen con tanta frecuencia que conviene tenerlos presentes en cada experimento. La selección de estos cuatro es un criterio de esta guía, y cada uno tiene su mitigación.

- **Confirmación.** Lleva a ver solo los datos que apoyan la creencia previa. Se mitiga buscando activamente lo que la contradice y anotando lo que no gustó escuchar.
- **Supervivencia.** Analiza solo los casos que llegaron al final, como cuando se encuesta únicamente a quien completó un proceso. Se mitiga incluyendo a quienes abandonaron o fracasaron.
- **Selección.** Usa una muestra que no representa al segmento, como la que sale de reclutar en la propia red. Se mitiga fijando los criterios de selección antes de reclutar (capítulo 3).

- **Deseabilidad social.** Produce respuestas pensadas para agradar, del tipo «lo usaría» o «lo pagaría». Se mitiga preguntando por el comportamiento pasado, cuánto pagó la persona la última vez en lugar de cuánto pagaría.

La página del BoK de Scrum Manager sobre sesgos cognitivos recoge otros con su definición. Conviene repasarla cuando un resultado parece demasiado bueno para ser cierto, que es justo cuando menos ganas hay de dudar de él.

12.4 Correlación y causalidad

En producto es fácil confundir correlación con causalidad. Que las conversiones subieran tras añadir una funcionalidad no implica que la funcionalidad las causara, y que los usuarios que usan una función retengan más no implica que esa función cause la retención; quizá los más comprometidos simplemente usan más funciones. Para afirmar causalidad hacen falta grupo de control, asignación aleatoria y control de los factores de confusión, como vimos en el capítulo 11.

Sin esas condiciones, lo honesto es decir «observamos una correlación». Es una afirmación más débil y también más útil, porque orienta el siguiente experimento sin comprometer una inversión que el dato no justifica.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Situar un dato en el gradiente de la fuerza de la evidencia, desde lo que se dice hasta lo que se paga, y asignarle un peso coherente con su posición.
- Distinguir una señal de validación débil (me gusta, feedback de la red propia, aperturas de correo) de una señal con coste para quien la produce.
- Identificar cuál de los cuatro sesgos frecuentes afecta a una muestra o a un dato y proponer su mitigación.
- Reformular una pregunta de intención futura en una de comportamiento pasado para reducir la deseabilidad social.
- Diferenciar correlación de causalidad en un resultado concreto y pedir evidencia adicional cuando la disponible no sostiene la conclusión.

Errores y antipatrones frecuentes

- **Tomar reacciones por demanda.** Cuarenta y siete me gusta no validan nada. Busca un comportamiento con coste, como una reserva, un pago o tiempo invertido.
- **Encuestar solo a quien terminó.** Excluye a quien abandonó, que es precisamente quien mejor explica el problema. Incluye a los que fracasaron.
- **Descartar los datos incómodos.** Tratar a dos usuarios negativos como casos aislados pierde información. Investiga qué los diferencia del resto.
- **Pesar igual lo que se dice y lo que se hace.** Una intención declarada y un pago no son la misma evidencia. Sitúa cada dato en su peldaño antes de sumarlos.

Para profundizar

- [Alexander Osterwalder — How strong is your innovation evidence?](#)
- [Rob Fitzpatrick — The Mom Test](#)
- [Scrum Manager BoK — Sesgo cognitivo](#)

BLOQUE D · DECIDIR CON EVIDENCIA

Capítulo 13. Perseverar, iterar, pivotar o parar · EN CONSOLIDACIÓN

Ante la evidencia se abren cuatro caminos, y la falacia del coste hundido es el sesgo más peligroso a la hora de elegir entre ellos.

Introducción

El descubrimiento culmina en una decisión, y decidir bien cuando la evidencia es mixta es una competencia propia. Muchos equipos plantean la decisión como un sí o un no, seguir o abandonar, y pierden así las opciones intermedias, que es donde caen la mayoría de los casos reales.

En este capítulo veremos las cuatro opciones que se abren ante un resultado, la falacia del coste hundido, qué exige un pivote responsable y qué factores conviene ponderar cuando la evidencia no es concluyente. Su estado es en consolidación porque el espectro de opciones está aceptado, aunque su uso explícito todavía madura en muchos equipos.

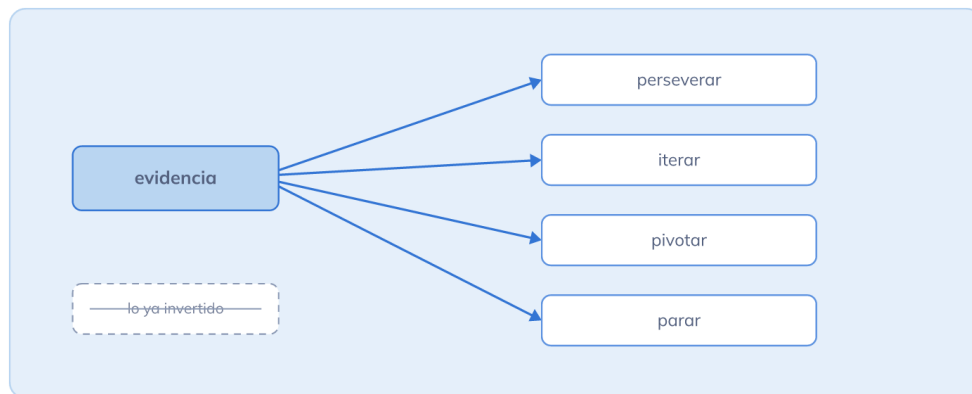


Figura 13. Cuatro caminos ante la evidencia: la decisión se abre en cuatro opciones, y lo ya invertido no figura entre los criterios.

13.1 Las cuatro opciones

Ante los resultados de un experimento, el equipo elige entre cuatro caminos.

- **Perseverar.** La evidencia es lo bastante positiva para seguir en la dirección actual.
- **Iterar.** Se ajustan elementos concretos (un mensaje, un precio, un paso del flujo) manteniendo el rumbo general.

- **Pivotar.** Se cambia un aspecto fundamental, como el segmento, la propuesta de valor o el canal, conservando lo aprendido.
- **Parar.** La evidencia indica que la iniciativa no es viable y que los recursos rinden más en otra parte.

La expresión clásica procede de Eric Ries, que en *The Lean Startup* plantea la decisión de pivotar o perseverar y describe varios tipos de pivote, entre ellos el cambio de segmento de clientes. Por eso esta guía no trata «segmentar» como una opción aparte. Descubrir que otro segmento responde mejor y reorientarse hacia él es una forma de pivote, y conviene someterlo a la misma disciplina que los demás.

La frontera entre iterar y pivotar no siempre es nítida. Una regla práctica es preguntarse si cambia la hipótesis de fondo o solo su ejecución. Ajustar el precio de una oferta que ya interesa es iterar; descubrir que quien paga es un cliente distinto del que se entrevistó es pivotar, porque la hipótesis sobre el segmento era falsa y todo lo que dependía de ella hay que revisarlo.

13.2 La falacia del coste hundido

El sesgo más peligroso en estas decisiones es justificar que se continúa por el tiempo o los recursos ya invertidos. «Llevamos seis meses, el equipo está muy comprometido y los dos usuarios que terminaron fueron entusiastas» suena razonable y no dice nada sobre el futuro. Lo ya invertido es irrelevante para la decisión; lo relevante es cuál es la mejor inversión de los próximos recursos dada la evidencia actual.

Las señales de alerta son fáciles de reconocer cuando se buscan.

- Argumentos sobre el esfuerzo invertido en lugar de sobre la evidencia.
- Atención a los pocos casos positivos y silencio sobre los negativos.
- Resistencia emocional a considerar el pivote o la parada, a menudo porque la idea tiene un promotor con peso en la organización.

La mejor defensa es pactar de antemano, junto con el criterio de éxito del capítulo 11, qué resultado llevaría a pivotar o a parar. Si el umbral de parada está escrito antes de empezar, la decisión deja de depender del ánimo del equipo en el peor momento y de la influencia de quien más ha apostado por la idea.

13.3 Capturar el aprendizaje antes de cambiar de rumbo

Un pivote responsable es una decisión documentada. Antes de cambiar de rumbo, el equipo escribe qué aprendió y por qué no funcionó, formula la nueva hipótesis a partir de ello, la pone a prueba antes de invertir y explica el razonamiento a quienes dependen de la decisión.

La frase que ordena el pivote. «Aprendimos que [hipótesis] era falsa porque [evidencia]; la nueva hipótesis es [X] y la probaremos con [experimento]». Si el equipo no puede completarla, todavía no está listo para pivotar.

Parar también se documenta. Una iniciativa que se detiene con su aprendizaje escrito deja al equipo mejor situado que una que se abandona en silencio, porque la siguiente idea en ese espacio ya no empieza de cero y las entrevistas y experimentos hechos siguen sirviendo como evidencia.

13.4 Decidir con evidencia mixta

Cuando la evidencia no es claramente positiva ni negativa, conviene ponderar varios factores antes de decidir. Los que siguen son un criterio de trabajo de esta guía, útil para ordenar la conversación del equipo y hacer explícito lo que de otro modo se decidiría por intuición.

- **Calidad de la evidencia.** Si es robusta o si hay sesgos en la muestra (capítulo 12).
- **Magnitud de las señales.** Si son lo bastante grandes para distinguirse del ruido (capítulo 11).
- **Coste de oportunidad.** Qué deja de hacer el equipo mientras sigue en esta línea.
- **Reversibilidad.** Si la decisión se puede deshacer con poco coste o compromete recursos difíciles de recuperar.
- **Incertidumbre residual.** Cuánto queda por saber y cuánto costaría averiguarlo.

Si la incertidumbre residual es alta y el experimento adicional es barato, lo sensato es diseñar una prueba más específica antes de decidir. Así la decisión enlaza con el descubrimiento continuo del bloque E, donde decidir es un acto recurrente de cada semana y deja de ser un hito final.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Situar una decisión en el espectro perseverar, iterar, pivotar o parar, sin reducirla a una alternativa binaria.
- Reconocer la falacia del coste hundido en un argumento y reorientar la decisión hacia la mejor inversión de los próximos recursos.
- Tratar un cambio de segmento como un tipo de pivote y aplicarle la misma disciplina de hipótesis y validación.
- Documentar un pivote o una parada con lo aprendido, la nueva hipótesis y el experimento que la pondrá a prueba.
- Ponderar la evidencia mixta por su calidad, su magnitud, el coste de oportunidad, la reversibilidad y la incertidumbre residual, y decidir si hace falta otro experimento.

Errores y antipatrones frecuentes

- **Plantear la decisión como binaria.** Seguir o abandonar ignora iterar y pivotar, que es donde caen la mayoría de los casos reales.
- **Ceder al coste hundido.** Lo invertido no pesa en la decisión futura. Pregunta por la mejor inversión de los próximos recursos.

- **Pivotar sin capturar lo aprendido.** Cambiar de rumbo sin documentar por qué arriesga repetir los mismos errores en la dirección nueva.
- **Decidir a ciegas con una prueba barata a mano.** Si un experimento asequible aclararía la situación, decidir sin él es precipitado.

Para profundizar

- [Eric Ries — The Lean Startup](#)
- [Teresa Torres — Product discovery basics](#)
- [Scrum Manager BoK — Product discovery](#)

BLOQUE E · EL DESCUBRIMIENTO COMO HÁBITO DEL EQUIPO

Capítulo 14. Descubrir cada semana: el trío de producto y el contacto continuo · EN CONSOLIDACIÓN

El descubrimiento continuo mantiene el contacto con clientes semana a semana, en manos del mismo equipo que construye el producto.

Introducción

Durante años el descubrimiento se entendió como una fase, con una investigación al principio del proyecto, un informe y después meses de desarrollo sin volver a hablar con clientes. El problema es que los mercados cambian y las hipótesis de producto caducan, de modo que un equipo que descubrió hace seis meses y desde entonces solo entrega acumula riesgo sin saberlo.

En este capítulo veremos qué es el descubrimiento continuo y quién lo sostiene, cómo se organiza la cadencia semanal, cómo encaja con el trabajo del *sprint* y cómo se conserva lo que se aprende. El qué del descubrimiento continuo lo enseña la certificación Product Owner de Scrum Manager; aquí nos ocupamos de cómo mantenerlo en marcha semana tras semana.



Figura 14. Una cadencia semanal: el trío de producto habla con clientes cada semana y lo aprendido se acumula en un repositorio común.

14.1 El descubrimiento continuo y el trío de producto

Teresa Torres definió el descubrimiento continuo en *Continuous Discovery Habits* (2021) como el contacto al menos semanal con clientes del equipo que construye el producto, mediante actividades pequeñas de investigación orientadas a un resultado (*outcome*). Cada parte de la definición importa. La frecuencia semanal impide que el contacto se convierta en un acontecimiento, que lo haga el propio equipo evita depender del informe de terceros, y el resultado da dirección a lo que se investiga.

Lo sostiene el trío de producto (*product trio*), formado por quien lleva el producto, quien diseña y quien desarrolla. Los tres participan en las entrevistas y en las decisiones de descubrimiento, porque cada uno ve cosas distintas en la misma conversación y porque lo que se aprende de primera mano se aplica mejor que lo que se lee en un resumen. El trío tampoco excluye al resto del equipo ni a las partes interesadas (*stakeholders*); es el núcleo que garantiza la continuidad.

La práctica es canónica en el área y sigue siendo minoritaria. En la encuesta de ProductPlan de finales de 2025, publicada como *State of Product Management 2026* con respuestas de unas 250 personas, los equipos describían así su forma de recoger y usar el conocimiento de clientes.

- Un 34 % lo recoge con regularidad y lo usa para priorizar.
- Un 18,9 % lo recoge, pero le cuesta convertirlo en decisiones.
- Un 19,3 % vive sobre todo de peticiones sueltas o de escaladas de clientes.
- Un 10,2 % apenas investiga y se apoya en perspectivas internas.
- Un 5,3 % usa herramientas de IA para sintetizar o interpretar el feedback de clientes.
- Un 12,3 % no tiene ningún proceso consistente.

14.2 Sostener la cadencia

La dificultad del descubrimiento continuo está en sostenerlo cuando llegan las urgencias, y tres decisiones prácticas lo hacen viable.

Quién entrevista

Entrevista el trío, idealmente al completo y como mínimo dos de sus miembros, para que la interpretación no dependa de una sola persona. Rotar quién conduce la conversación y quién toma notas reparte la habilidad y evita que la entrevista se convierta en el trabajo de un especialista al que el resto no acompaña.

Cómo se recluta cada semana

El cuello de botella suele ser el reclutamiento. Torres recomienda automatizarlo para que cada semana haya conversaciones en la agenda sin que nadie tenga que organizarlas de nuevo, por ejemplo con una invitación dentro del propio producto o con un acuerdo estable con el equipo que atiende a clientes.

Automatizar el reclutamiento no exime de cuidar a quién se escucha. Los criterios de segmento, el cuestionario de selección y la precaución con la red propia son los del capítulo 3, y conviene revisarlos de vez en cuando para que la agenda semanal no se llene siempre con el mismo tipo de cliente, el más dispuesto a hablar.

Cómo se protege el tiempo

Una franja fija en el calendario del equipo, con dos o tres conversaciones a la semana, resiste mejor que la intención de encontrar hueco. Lo que se protege es la continuidad, y una entrevista cada semana a lo largo de un trimestre deja al equipo más cerca de sus clientes que diez entrevistas concentradas en un mes y seguidas de silencio.

14.3 Doble vía: descubrir y entregar a la vez

El campo llamó doble vía (*dual-track*) a la idea de que el descubrimiento y la entrega corren en paralelo: mientras el equipo entrega lo que ya tiene validado, sigue investigando las oportunidades siguientes. El encuadre es útil siempre que no se lea como dos equipos ni como un traspaso en el que una vía produce especificaciones que la otra ejecuta, porque entonces reaparece la fase de descubrimiento con otro nombre.

En un equipo que trabaja con scrum, el descubrimiento continuo no necesita un evento propio. Las conversaciones con clientes alimentan el refinamiento del *product backlog*, lo aprendido informa la planificación del sprint y la revisión del sprint es una ocasión natural para contrastar con las partes interesadas lo que se ha descubierto.

La Guía de Scrum no prescribe ninguna práctica de descubrimiento, así que el encaje es una decisión del equipo que conviene hacer explícita. Lo importante es que descubrir forme parte del trabajo ordinario del sprint, con su tiempo reservado, y que no quede como una actividad voluntaria que se sacrifica en cuanto aprieta la entrega.

14.4 Conservar la evidencia

Lo que se aprende en una entrevista se pierde si solo vive en la memoria de quien estuvo. Un repositorio de evidencia, accesible a todo el equipo, guarda los resúmenes de entrevista (capítulo 4), las citas etiquetadas por oportunidad y los resultados de los experimentos, de modo que cualquier decisión pueda rastrearse hasta lo que la sostiene.

La necesidad crece con la democratización de la investigación. Según el informe de Maze de 2026, el 39 % de los encuestados afirma que en su organización los responsables de producto ya investigan con usuarios, pero solo el 49 % de las organizaciones mantiene un repositorio de investigación. Cuando investigan más personas y nadie guarda lo aprendido, cada equipo repite entrevistas que otro ya hizo y las conclusiones dependen de quién las recuerde.

Cadencia y memoria. El descubrimiento continuo necesita las dos cosas, contacto frecuente con clientes y un lugar donde lo aprendido se acumule. Sin la primera, el equipo decide con información vieja; sin la segunda, cada semana empieza de cero.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Explicar los componentes de la definición de descubrimiento continuo de Teresa Torres (frecuencia semanal, equipo que construye, actividades pequeñas y resultado) y el papel de cada uno.
- Organizar la participación del trío de producto en las entrevistas, con roles de conducción y de notas que roten.
- Diseñar un reclutamiento automatizado que mantenga conversaciones semanales sin renunciar a los criterios de segmento.
- Integrar lo aprendido en el refinamiento del product backlog, la planificación y la revisión del sprint sin crear un traspaso entre descubrir y entregar.
- Establecer un repositorio de evidencia que permita rastrear una decisión hasta los resúmenes de entrevista y los experimentos que la sostienen.

Errores y antipatrones frecuentes

- **Descubrir por campañas.** Una ronda intensa de entrevistas cada seis meses deja al equipo decidiendo con información caducada. Protege una franja semanal.
- **Delegar las entrevistas.** Si solo entrevista un especialista, el resto del trío lee resúmenes y aprende menos. Que participen al menos dos de sus miembros.
- **Convertir la doble vía en traspaso.** Una vía que entrega especificaciones a la otra reproduce la fase de descubrimiento con otro nombre. Descubren y entregan las mismas personas.
- **Aprender sin guardar.** Sin repositorio, cada equipo repite entrevistas y las conclusiones dependen de la memoria. Conserva los resúmenes y los resultados donde todos puedan consultarlos.

Para profundizar

- [Teresa Torres — Continuous Discovery Habits](#)
- [Scrum Manager BoK — Dual-track discovery & delivery](#)
- [Skill Arena — Product Owner](#)
- [Maze — Future of User Research Report 2026](#)

BLOQUE E · EL DESCUBRIMIENTO COMO HÁBITO DEL EQUIPO

Capítulo 15. El árbol de oportunidades: del resultado al experimento · EN CONSOLIDACIÓN

El árbol de oportunidades conecta el resultado que persigue el equipo con lo que aprende de sus clientes y con las pruebas que decide hacer.

Introducción

Un equipo que entrevista cada semana acumula pronto más oportunidades de las que puede atender, y cada conversación parece sugerir una funcionalidad nueva. Sin una estructura que ordene lo aprendido, el descubrimiento continuo degenera en una lista de peticiones y el equipo vuelve a trabajar como una fábrica de funcionalidades.

En este capítulo veremos el árbol de oportunidades y soluciones de Teresa Torres, sus cuatro niveles, las decisiones que obliga a tomar, cómo ha cambiado su construcción con la IA y dónde termina su terreno frente a las skills vecinas. Es la herramienta que une lo que hemos visto en el bloque A, conversar, con lo que vimos en el bloque C, experimentar.

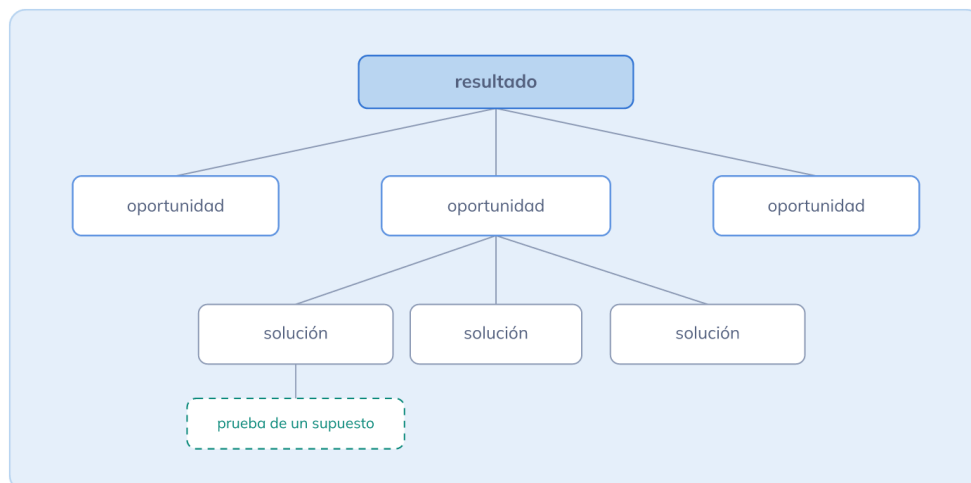


Figura 15. El árbol de oportunidades: el resultado se descompone en oportunidades, cada oportunidad en soluciones y cada solución en pruebas de sus supuestos.

15.1 Los cuatro niveles

El árbol de oportunidades y soluciones ordena el descubrimiento de arriba abajo en cuatro niveles.

- **El resultado.** En la raíz está el resultado que el equipo persigue, formulado como un cambio medible en el comportamiento de los clientes o en el negocio. Suele venir de los objetivos y resultados clave de la organización.
- **Las oportunidades.** Debajo van las necesidades, fricciones y deseos de los clientes que, si se atendieran, moverían el resultado. Salen de las entrevistas y se expresan desde el punto de vista del cliente, nunca como funcionalidades.

- **Las soluciones.** Para cada oportunidad elegida se plantean varias soluciones posibles que compiten entre sí.
- **Las pruebas.** Para cada solución se prueban los supuestos de los que depende, en lugar de construirla entera para ver si funciona.

Las oportunidades pueden descomponerse en otras más pequeñas, y esa jerarquía es parte del valor del árbol. Una oportunidad grande y vaga («les cuesta coordinarse») se divide hasta llegar a una lo bastante concreta para idear sobre ella («no saben qué ha cambiado desde la última reunión»), y en ese nivel las soluciones empiezan a ser comparables.

La raíz merece cuidado. Torres distingue los resultados de negocio, como los ingresos o la retención, de los resultados de producto, que el equipo puede mover directamente con su trabajo, como que más clientes completen su primera configuración sin ayuda. Un resultado de producto ordena mejor el árbol, porque está lo bastante cerca de lo que el equipo decide cada semana para que las oportunidades se puedan conectar con él.

15.2 Las decisiones que obliga a tomar

El árbol es útil sobre todo por las decisiones que impone y por el orden en que las impone.

- **Elegir una oportunidad antes de idear.** Comparar oportunidades entre sí, en lugar de comparar ideas, evita que la solución favorita del equipo decida por la puerta de atrás qué problema se ataca.
- **Comparar varias soluciones.** Torres recomienda trabajar con varias soluciones para la oportunidad elegida, porque con una sola el equipo tiende a confirmarla en lugar de ponerla a prueba.
- **Probar supuestos en lugar de construir.** Cada solución se descompone en los supuestos de los que depende (capítulo 8) y se prueban primero los más arriesgados, con el mínimo coste.

El árbol también hace visible cuándo cambiar de rama. Si las pruebas de todas las soluciones de una oportunidad fallan, quizá el problema esté en la oportunidad elegida más que en las soluciones, y el árbol permite volver un nivel arriba sin perder lo aprendido.

15.3 El árbol con IA

En 2026 han aparecido herramientas que generan borradores del árbol a partir de entrevistas. Teresa Torres presentó en febrero una que convierte tres entrevistas reales en un primer árbol en minutos, sintetizando primero cada entrevista por separado y buscando después los patrones entre ellas.

La misma autora advierte contra el extremo opuesto, los árboles de un clic que se generan describiendo un mercado y se rellenan con datos inventados. Su argumento es que el sentido del árbol está en sintetizar lo que se aprende de clientes reales, así que un

árbol sin entrevistas detrás es una lista de ideas con forma de árbol. El borrador generado con IA sirve como punto de partida que el trío revisa con su experiencia y su contexto, con las reglas de síntesis del capítulo 16.

Revisar el borrador tiene un método sencillo. Cada oportunidad del árbol debe poder señalar al menos una entrevista de la que procede, y las que no pueden señalar ninguna son hipótesis del equipo o invención del modelo, que se marcan como tales o se retiran antes de elegir por dónde seguir.

El árbol vive. Un árbol se actualiza cada semana con lo que traen las entrevistas y los experimentos. Si no ha cambiado en un mes, lo más probable es que el equipo haya dejado de descubrir.

15.4 Su lugar entre las skills vecinas

El árbol toca a varias disciplinas y conviene saber qué corresponde a cada una. El resultado de la raíz se define con objetivos y resultados clave, cuyo diseño tiene skill propia en Skill Arena, y la skill de Historias de Usuario usa el árbol para decidir dónde intervenir antes de escribir historias. Lo que sale de las pruebas validadas alimenta después el product backlog.

Aquí nos ocupamos del tramo central, que es el propio del descubrimiento. Cómo se construyen las oportunidades a partir de las entrevistas, cómo se elige entre ellas y cómo se idean y prueban las soluciones que compiten por cada una.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Construir un árbol de oportunidades a partir de entrevistas reales, con un resultado en la raíz y oportunidades formuladas desde el punto de vista del cliente.
- Descomponer una oportunidad vaga en oportunidades más concretas sobre las que se pueda idear.
- Elegir una oportunidad comparándola con otras antes de generar soluciones, y justificar la elección.
- Idear varias soluciones para la oportunidad elegida y descomponer cada una en los supuestos que hay que probar.
- Revisar un borrador del árbol generado con IA contra las entrevistas de las que procede, detectando oportunidades sin respaldo.

Errores y antipatrones frecuentes

- **Oportunidades con forma de funcionalidad.** «Necesitan un panel de control» es una solución disfrazada. Reformula desde lo que el cliente intenta conseguir.
- **Una sola solución por oportunidad.** Sin alternativas, el equipo confirma su idea favorita. Plantea varias y compáralas.

- **Árbol sin entrevistas detrás.** Un árbol generado describiendo un mercado es una lista de ideas con forma de árbol. Constrúyelo desde lo que dijeron clientes reales.
- **Árbol congelado.** Si no cambia en semanas, el equipo ha dejado de descubrir. Actualízalo con cada entrevista y con cada prueba.

Para profundizar

- [Teresa Torres — Opportunity solution trees](#)
- [Teresa Torres — AI opportunity solution trees](#)
- [Skill Arena — OKRs y estrategia de producto](#)
- [Skill Arena — User stories](#)

BLOQUE F · DESCUBRIR CON INTELIGENCIA ARTIFICIAL

Capítulo 16. Entrevistar y sintetizar con IA: dónde ayuda y dónde erosiona · EN CONSOLIDACIÓN

La IA acelera la parte mecánica del descubrimiento, que es transcribir, resumir y agrupar, y erosiona la parte de juicio cuando se le delega; su uso sensato consiste en comparar su trabajo con el propio.

Introducción

En muy poco tiempo, la inteligencia artificial ha pasado de ser una curiosidad a formar parte del trabajo diario de quien investiga con clientes. Transcribe las entrevistas, resume lo que se dijo, agrupa temas entre conversaciones y propone borradores de síntesis, y ya hay plataformas que conducen la entrevista entera sin nadie al otro lado. La pregunta profesional ha dejado de ser si usarla y se ha convertido en dónde confiar en ella.

En este capítulo veremos qué dice la evidencia sobre su adopción, cómo usarla en la síntesis sin perder lo que la síntesis aporta, cuándo tiene sentido que modere una entrevista y qué cuidados añade al tratamiento de los datos. Partimos de lo visto en los bloques A y B, porque la IA cambia el instrumental y deja intactas las dos disciplinas de fondo, la entrevista que aprende y la síntesis que llega a un insight.

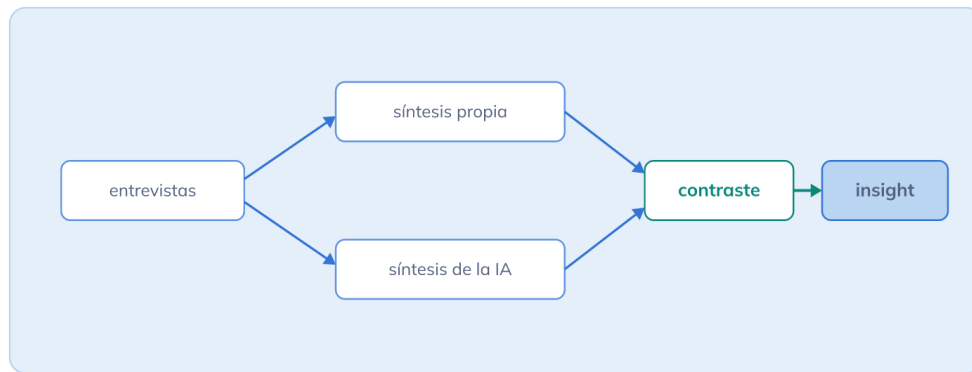


Figura 16. Comparar, no delegar: la síntesis propia y la de la IA se hacen por separado y se contrastan antes de afirmar un insight.

16.1 Qué dice la evidencia sobre su adopción

La adopción está medida por más de una fuente. El *Future of User Research Report 2026* de Maze, publicado el 11 de marzo de 2026 a partir de una encuesta a casi quinientos profesionales de la investigación, el diseño y el producto, da un 69 % que usa IA en al menos parte de sus estudios, un 19 % más que el año anterior. El mismo informe recoge un límite claro: el 82 % sostiene que interpretar el matiz y la emoción sigue siendo trabajo humano, y el 80 % dice lo mismo de las decisiones éticas.

La encuesta de ProductPlan *State of Product Management 2026*, hecha a finales de 2025 con cerca de doscientas cincuenta personas, permite ver en qué se usa. Quienes la emplean en el descubrimiento lo hacen sobre todo en las primeras tareas del proceso:

- resumir el feedback de clientes (47,1 %);
- redactar resúmenes o informes de investigación (40,2 %);
- convertir el material en bruto en hallazgos estructurados (40,2 %);
- identificar temas o patrones (39,8 %).

En cambio, solo un 14,3 % la usa para priorizar problemas de clientes, y un 27,9 % no la usa en el descubrimiento. El dibujo que resulta es el de una herramienta que ayuda a procesar y ordenar la información, mientras la decisión sobre qué problema importa sigue en manos del equipo. Es un reparto sensato, y el resto del capítulo explica por qué conviene mantenerlo.

16.2 Sintetizar con la IA sin delegarle el juicio

El riesgo principal tiene nombre. Teresa Torres llama síntesis perezosa con IA a volcar las transcripciones en un asistente y aceptar su resumen sin revisarlo. El resumen suele sonar convincente, porque la IA escribe bien, y precisamente por eso es fácil tomarlo por un insight cuando a menudo es una observación bien redactada o un patrón sin causa.

La alternativa que propone Torres se ordena en tres pasos, y el orden forma parte del método:

1. sintetizar primero, a partir de las propias notas y transcripciones, sin mirar lo que haya producido la herramienta;

2. pedir después a la IA la misma síntesis, sobre el mismo material;
3. comparar las dos y examinar las diferencias, tanto lo que la IA vio y el equipo no como lo que el equipo vio y la IA pasó por alto.

La comparación es donde está el valor. Cada método capta cosas que el otro pierde, y cada discrepancia obliga a volver a las fuentes, que es justo el hábito que la síntesis perezosa elimina. El orden importa porque quien lee primero el resumen de la herramienta tiende a anclarse en él y a encontrar en las notas lo que ya espera encontrar.

La transcripción también interpreta. El resumen que genera una herramienta de transcripción ya selecciona y reformula lo que se dijo. Conviene tratarlo como una interpretación más, con la misma marca que las inferencias propias del capítulo 4, y contrastarlo con la transcripción literal antes de citarlo como evidencia.

El ritual de síntesis en grupo, con sus dinámicas de sala, tiene skill propia en Scrum Manager, Design Thinking. Aquí interesa otra cosa, el uso de la IA como segunda lectura del material, que acelera el trabajo sin quitarle al equipo la responsabilidad de decidir qué ha aprendido.

16.3 Cuándo puede entrevistar una IA

Las plataformas de entrevista moderada por IA conducen la conversación sin entrevistador humano. Formulan las preguntas, repreguntan según las respuestas y entregan la transcripción con un resumen. El Nielsen Norman Group las puso a prueba en un estudio que Maria Rosala publicó el 30 de enero de 2026, con diez participantes de ocho países y dos plataformas, Marvin y UserFlix, y sus conclusiones separan con claridad los usos.

Dónde encaja

- recoger opiniones sobre un producto, sobre todo después de lanzarlo;
- entrevistas de criba para seleccionar participantes;
- equipos sin personas dedicadas a la investigación que necesitan una primera señal;
- estudios en varios idiomas sin intérpretes.

Dónde no encaja

- temas que exigen un conocimiento profundo del dominio;
- la exploración de un espacio de problema desconocido, que es precisamente la entrevista de descubrimiento del capítulo 2;
- investigaciones en las que cada entrevista debe tomar un camino distinto;
- estudios que requieren compartir pantalla o ver a la persona hacer una tarea.

La experiencia de los participantes también cuenta. En el estudio, solo tres de diez estuvieron de acuerdo en que la conversación les había parecido natural, y cinco de diez dijeron sentirse cómodos. Algunos se sintieron escuchados; otros describieron la

situación como artificial o sintieron que perdían el control. Para una entrevista cuyo valor depende de que la persona cuente con confianza una historia concreta, ese dato pesa.

Cuando entrevista una IA se añaden obligaciones. Desde el 2 de agosto de 2026, el artículo 50 del Reglamento europeo de IA exige que los sistemas que interactúan con personas les informen de que son una IA, y si la herramienta reconoce emociones, quien la usa debe informar a las personas expuestas. El capítulo 5 desarrolla estas obligaciones y las de la grabación y la transcripción, que se aplican igual cuando la entrevista la conduce una máquina.

16.4 Datos y procedencia

Usar un asistente de IA sobre material de entrevistas plantea primero una cuestión de datos. Las notas y transcripciones en bruto llevan nombres, cargos, empresas y detalles que identifican a las personas, y pegarlas en una herramienta sin comprobar qué hace con ellas es un tratamiento que nadie ha autorizado. La pauta práctica, coherente con la que enseña la skill Historias de Usuario, consiste en anonimizar antes de analizar y en usar herramientas cuyo tratamiento de los datos esté comprobado y cubierto por el consentimiento obtenido.

La otra cuestión es la procedencia. Un hallazgo que sale de la síntesis debe poder rastrearse hasta las citas y los hechos que lo sostienen, y eso vale igual si lo propuso la IA. Si el equipo no puede señalar en qué entrevistas aparece un patrón, todavía no puede afirmarlo, por muy bien escrito que esté el resumen que lo enuncia.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Situar un uso concreto de la IA en el descubrimiento entre los que la evidencia respalda (transcribir, resumir, estructurar, agrupar temas) y los que sigue reservando al juicio humano (interpretar el matiz, priorizar problemas, decidir).
- Aplicar el flujo de síntesis comparada, que empieza por la síntesis propia, sigue con la de la IA sobre el mismo material y termina en el análisis de las diferencias.
- Decidir si una entrevista puede moderarla una IA según su propósito, distinguiendo el feedback sobre un producto y la criba de la exploración de un problema desconocido.
- Tratar el resumen de una herramienta de transcripción como una interpretación y contrastarlo con la transcripción literal antes de usarlo como evidencia.
- Preparar material de entrevistas para su análisis con IA anonimizando los datos identificables y comprobando el tratamiento que hace la herramienta.

Errores y antipatrones frecuentes

- **Síntesis perezosa con IA.** Volcar las transcripciones en un asistente y aceptar su resumen convierte una observación bien redactada en un falso insight. Sintetiza primero y compara después.

- **Leer primero el resumen de la herramienta.** La lectura posterior de las notas queda anclada en lo que propuso la IA. El orden de la comparación forma parte del método.
- **Usar un entrevistador automático para explorar.** La moderación por IA no persigue lo inesperado ni cambia de camino según la historia que escucha, que es lo que pide la entrevista de descubrimiento. Resérvala para opiniones sobre un producto, criba o estudios multilingües.
- **Pegar notas en bruto en un asistente.** Nombres, cargos y detalles identificables acaban en una herramienta cuyo tratamiento nadie ha comprobado. Anonimiza antes de analizar.

Para profundizar

- [Nielsen Norman Group — AI-moderated interviews](#)
- [Teresa Torres en airfocus — AI product management](#)
- [Maze — Future of User Research Report 2026](#)
- [ProductPlan — State of Product Management 2026](#)

BLOQUE F · DESCUBRIR CON INTELIGENCIA ARTIFICIAL

Capítulo 17. Lo sintético en la escala de la evidencia · EMERGENTE

Los usuarios sintéticos (synthetic users) sirven para preparar el trabajo con clientes reales, y sus respuestas ocupan el peldaño más bajo de la escala de la evidencia, por debajo de cualquier opinión humana.

Introducción

Entre las herramientas de IA que se ofrecen para el descubrimiento, ninguna promete tanto como los usuarios sintéticos, perfiles generados por un modelo de lenguaje que responden como lo haría un segmento de clientes. La promesa es evidente, investigación en minutos, sin reclutar ni esperar y con tantas respuestas como se quieran. Y la tentación también, porque las respuestas llegan redactadas, coherentes y sin contradicciones.

En este capítulo veremos qué son y por qué convencen, qué dice la evidencia acumulada hasta 2026 sobre su validez, dónde se sitúan en la escala de la fuerza de la evidencia del capítulo 12 y qué uso legítimo les queda. Adelantamos la conclusión, porque ordena todo el recorrido: sirven para preparar el trabajo con clientes reales y resultan peligrosos cuando lo sustituyen.

17.1 Qué son y por qué convencen

Un usuario sintético es un perfil que un modelo de lenguaje interpreta a partir de una descripción (edad, profesión, contexto, necesidades) y que responde a las preguntas de una entrevista o de una encuesta como lo haría esa persona. Algunas herramientas generan decenas de perfiles a la vez y agregan sus respuestas en informes que se parecen mucho a los de una investigación real.

Convencen por las mismas razones que los hacen peligrosos. Responden siempre, lo hacen con fluidez y rara vez se contradicen, mientras que las personas reales dudan, cambian de opinión, se van por las ramas y a veces no saben explicar lo que hacen. Esa aspereza de lo humano es justamente la información que busca el descubrimiento, y el perfil sintético la elimina.

17.2 Qué dice la evidencia

La evidencia acumulada es negativa y cada vez más precisa. El Nielsen Norman Group ya advirtió en 2024, en un artículo de Maria Rosala y Kate Moran, de dos problemas de fondo. Los usuarios sintéticos son complacientes, porque tienden a agrandar y describen conductas ideales, como terminar todos los cursos que empiezan cuando las personas reales los abandonan, y son superficiales, porque enumeran muchas necesidades sin distinguir cuáles importan de verdad.

En agosto de 2025, Raluca Budiu, de la misma organización, revisó los estudios que comparan conducta simulada y conducta real y llegó a una conclusión que resume bien el problema: los usuarios sintéticos captan las tendencias generales del

comportamiento humano, pero fallan en la magnitud de los efectos y en su variabilidad. Y la variabilidad es la que revela los subgrupos, los casos extremos y las opiniones polarizadas que cambian una decisión de diseño.

2026 ha traído evidencia académica más directa, con dos estudios publicados en arXiv en julio:

- Zihan Chen, Di Zhu y Lei Nico Zheng compararon respuestas simuladas y humanas en encuestas de actitudes sociales y de valores de varias culturas. Ningún modelo superó a los métodos de referencia que no usan modelos de lenguaje, todos trataban los rasgos demográficos como mucho más predictivos de lo que son en personas reales y, en tareas de segmentación, inflaban entre dos y cuatro veces las diferencias entre segmentos y llevaban a elegir el segmento equivocado en la mitad de los casos estadounidenses.
- Mantas Lukauskas y Viktorija Šarkauskaitė auditaron 37 modelos frente a una encuesta respondida por 263 empleados. Midieron un sesgo de aquiescencia, la tendencia a mostrarse de acuerdo, de 0,84 desviaciones típicas, y comprobaron que las respuestas de los modelos se parecían más entre sí que a las humanas. Concluyen que las muestras generadas por modelos no pueden reemplazar sin más a los datos humanos.

La opinión de los profesionales va en la misma dirección. En la encuesta de User Interviews *State of Synthetic Users*, de mayo de 2026, con 150 respuestas, el 47 % se declaraba escéptico, el 17,3 % contrario a su uso, el 79 % temía que las partes interesadas confiaran de más en hallazgos generados por IA y el 62,7 % trabajaba en organizaciones sin ninguna pauta sobre ellos.

17.3 Su lugar en la escala de la evidencia

En el capítulo 12 ordenamos la evidencia por su parecido con una decisión real con coste para quien la toma. Una respuesta sintética no tiene coste para nadie, y tampoco procede de nadie, porque es la predicción de un modelo sobre lo que diría una persona. Por eso ocupa el peldaño más bajo de la escala, por debajo incluso de la opinión declarada de un cliente real, que al menos pertenece a alguien.

Una comprobación sencilla. Ante un hallazgo, conviene preguntar de quién procede. Si procede de un modelo, puede orientar la siguiente pregunta de investigación, pero todavía no dice nada sobre los clientes.

Esta posición no depende de que los modelos mejoren. Aunque un modelo acertara la dirección de una conducta, la decisión de producto necesita saber cuánto, en quién y con qué variación, y eso es lo que la evidencia de 2025 y 2026 muestra que los perfiles simulados no reproducen. Es también la razón de su estado emergente, porque la herramienta se ofrece con insistencia, la práctica está en disputa y su permanencia no está garantizada.

17.4 Qué sí se puede hacer con ellos

Descartarlos como evidencia no obliga a descartarlos del todo. Las mismas fuentes que los critican les reconocen usos de preparación, en los que su respuesta nunca se toma por un dato sobre clientes:

- investigación documental para familiarizarse con un dominio nuevo antes de hablar con nadie;
- generar hipótesis que después se contrastan con personas reales;
- ensayar un guion de entrevista para detectar preguntas confusas, sugerentes o que invitan a opinar sobre el futuro;
- construir borradores de perfiles que la investigación real confirmará o corregirá.



Figura 17. Preparar, no sustituir: los usuarios sintéticos ayudan a preparar el trabajo con clientes reales, y la evidencia sale solo de estos.

La regla que une todos estos usos es de dirección. Lo sintético puede empujar hacia las personas reales, abriendo preguntas que merece la pena hacer, pero nunca puede ocupar su lugar en una decisión. Cuando un informe sintético llega a una reunión de priorización como si fuera investigación, el problema ya no es técnico: es de honestidad sobre la procedencia de la evidencia.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Explicar qué es un usuario sintético y por qué sus respuestas resultan convincentes aunque no procedan de ningún cliente.
- Identificar en un hallazgo sintético los fallos que documenta la evidencia, como la complacencia, la superficialidad, el exceso de peso de lo demográfico y la falta de variabilidad.
- Situar una respuesta sintética en la escala de la fuerza de la evidencia y justificar por qué ocupa el peldaño más bajo.
- Distinguir los usos de preparación (investigación documental, hipótesis, ensayo de guiones, borradores de perfiles) de su uso indebido como evidencia en una decisión.

- Señalar la procedencia sintética de un hallazgo cuando se presenta en una decisión de producto.

Errores y antipatrones frecuentes

- **Tomar lo sintético por investigación.** Un informe con muchas respuestas bien redactadas parece un estudio, pero no recoge la opinión de ningún cliente. Llámalo por su nombre y úsalo para preparar.
- **Segmentar con perfiles simulados.** Los modelos exageran las diferencias entre segmentos y pueden llevar a elegir el equivocado. La segmentación se valida con personas reales.
- **Buscar confirmación en lo sintético.** Un modelo complaciente tiende a dar la razón a quien pregunta, así que confirma casi cualquier idea. Es el sesgo de confirmación con otro disfraz.
- **Ocultar la procedencia.** Mezclar hallazgos sintéticos y reales en el mismo informe sin distinguirlos hace que las partes interesadas confíen de más. Indica siempre de dónde sale cada hallazgo.

Para profundizar

- [Nielsen Norman Group — Synthetic users](#)
- [Nielsen Norman Group — Evaluating AI-simulated behavior](#)
- [Chen, Zhu y Zheng — When synthetic users fail \(arXiv, 2026\)](#)
- [Lukauskas y Šarkauskaitė — Plausible but not valid \(arXiv, 2026\)](#)

BLOQUE F · DESCUBRIR CON INTELIGENCIA ARTIFICIAL

Capítulo 18. Validar un producto cuyo núcleo es un modelo

EMERGENTE

Cuando el valor depende de un modelo de IA, el descubrimiento incluye comprobar pronto si ese valor llega con una fiabilidad aceptable, y el análisis de errores con clientes reales se convierte en una actividad de descubrimiento.

Introducción

Hasta aquí hemos tratado la IA como herramienta del descubrimiento. Este capítulo trata el caso inverso, el de los productos cuyo valor depende de un modelo de IA, como un asistente que redacta, un sistema que clasifica o un agente que actúa en nombre del usuario. En ellos, a las preguntas de siempre sobre si el cliente lo quiere, si es negocio y si podemos construirlo, se suma otra que cambia el orden del trabajo, la de si el sistema acierta lo bastante a menudo, y falla de forma aceptable, para que el valor llegue de verdad.

Veremos por qué esa pregunta convierte la factibilidad en un supuesto que se prueba pronto, cómo el análisis de errores funciona como actividad de descubrimiento y dónde termina este capítulo y empiezan las skills que tratan la operación de estos sistemas. Es un terreno emergente, con la práctica todavía en formación, y conviene recorrerlo con la prudencia correspondiente.

18.1 La fiabilidad como parte del valor

En un producto convencional, una funcionalidad hace lo mismo cada vez que se usa, y la pregunta de valor se puede separar de la técnica. En un producto basado en un modelo, la misma entrada puede producir salidas distintas, y algunas serán erróneas. El valor que percibe el cliente depende entonces de la proporción y del tipo de errores que tolera en su contexto: un borrador con algún fallo puede ahorrarle tiempo, mientras que una clasificación equivocada en un proceso crítico puede destruir su confianza.

Por eso la pregunta de descubrimiento se amplía. Ya no basta con saber si el cliente quiere el resultado; hay que saber con qué fiabilidad lo necesita, qué errores le resultan tolerables y cuáles le hacen abandonar el producto. Esas respuestas solo salen de observar el sistema en manos de clientes reales, con sus datos y en sus situaciones.

18.2 La factibilidad como supuesto que se prueba pronto

En el capítulo 8 vimos que empezar por la deseabilidad es el orden habitual del mapeo de supuestos, y que el que manda es el cuadrante. Este es el caso en que el cuadrante altera el orden. Si la calidad del modelo en el dominio del cliente es incierta y todo el valor depende de ella, la factibilidad es un supuesto importante y sin evidencia, y conviene probarlo al principio y con el problema real.

Probarlo al principio no exige construir el producto. Un prototipo que conecte el modelo con un puñado de casos reales del cliente, o una versión en la que una persona

revise cada salida antes de entregarla, permiten ver pronto cuánto acierta el sistema y qué tipo de fallos comete. El capítulo 10 describe estas alternativas ligeras; la diferencia aquí es que lo que se mide incluye la calidad de las respuestas del modelo, además del interés del cliente.

18.3 El análisis de errores como actividad de descubrimiento

Teresa Torres y Petra Wille lo formularon en una conversación publicada en septiembre de 2025. Las prácticas de descubrimiento informan cada paso de la evaluación de un producto de IA, y la más productiva es el análisis de errores, que consiste en revisar con detalle las salidas reales del sistema, identificar sus modos de error y convertirlos en criterios de evaluación (*evals*) que permitan medir si el sistema mejora. Torres ha llegado a llamar a los evals un nuevo hábito del descubrimiento.

El paralelismo con la síntesis del capítulo 6 es directo. Igual que un insight se apoya en patrones que se repiten entre entrevistas, un modo de error se apoya en fallos que se repiten entre casos, y la tarea es la misma: agrupar, nombrar y decidir qué importa. Los modos de error más valiosos son los que solo aparecen con clientes reales, porque dependen de su lenguaje, de sus datos y de lo que esperan del sistema.

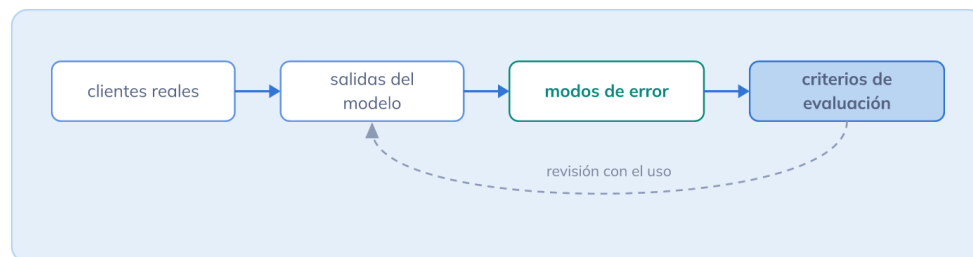


Figura 18. El análisis de errores como descubrimiento: los fallos que aparecen con clientes reales se agrupan en modos de error, que se convierten en criterios de evaluación que se revisan con el uso.

Los criterios cambian. Torres y Wille advierten de la deriva de los criterios, porque lo que el equipo considera una buena respuesta cambia a medida que aprende del uso real. Los evals necesitan mantenimiento continuo, igual que el resto del conocimiento que produce el descubrimiento.

18.4 Dónde termina este capítulo

La frontera con otras skills es deliberada. Aquí interesa un tramo concreto, el de validar que el producto aporta valor con una fiabilidad aceptable antes de comprometer la inversión. Cómo se diseñan, se automatizan y se operan los evals antes y después del lanzamiento lo tratan la certificación Product Owner de Scrum Manager y la skill Harness Engineering, que conviene consultar cuando el producto pasa de la validación a la entrega.

Lo que el descubrimiento aporta a esa operación es el origen de los criterios. Unos evals que no nacen de observar a clientes reales miden lo que el equipo imaginó, y

reproducen en la evaluación el mismo error que el capítulo 1 describía en el producto, el de mezclar la solución que se desea con el problema que existe.

Lo que debes saber hacer

Al dominar este tema, un profesional es capaz de:

- Formular, para un producto basado en un modelo, la pregunta de valor que incluye la fiabilidad, es decir, qué errores tolera el cliente en su contexto y cuáles le hacen abandonar.
- Reconocer cuándo la calidad del modelo es un supuesto importante y sin evidencia, y situarlo en el mapa de supuestos para probarlo pronto.
- Diseñar una prueba ligera que mida la calidad de las salidas del modelo con casos reales del cliente antes de construir el producto.
- Aplicar el análisis de errores a salidas reales, agrupando fallos, nombrando modos de error y convirtiéndolos en criterios de evaluación.
- Delimitar la validación del valor, propia del descubrimiento, frente a la operación de los evals, que corresponde a otras disciplinas.

Errores y antipatrones frecuentes

- **Dejar la calidad del modelo para el final.** Si todo el valor depende de ella, descubrir tarde que no alcanza obliga a tirar el trabajo hecho. Pruébala pronto y con casos reales.
- **Evaluar con casos inventados por el equipo.** Los criterios reflejan lo que el equipo imagina y no lo que el cliente necesita. Obtén los casos del uso real.
- **Tratar los evals como una lista cerrada.** Los criterios derivan a medida que se aprende. Revísalos con la misma cadencia que el resto del descubrimiento.
- **Confundir una demostración con una validación.** Una demostración con casos elegidos enseña lo que el modelo puede hacer; la validación muestra lo que hace con los casos del cliente, incluidos los difíciles.

Para profundizar

- [Torres y Wille — Al evals & discovery](#)
- [Teresa Torres — Assumption testing](#)
- [Skill Arena — Harness Engineering](#)
- [Skill Arena — Product Owner](#)

© 2026 Scrum Manager®. Esta obra se publica bajo licencia Creative Commons Atribución – No Comercial 4.0 Internacional (CC BY-NC 4.0). Los formadores y centros oficiales de Scrum Manager quedan licenciados bajo los términos CC BY 4.0 para su actividad formativa.