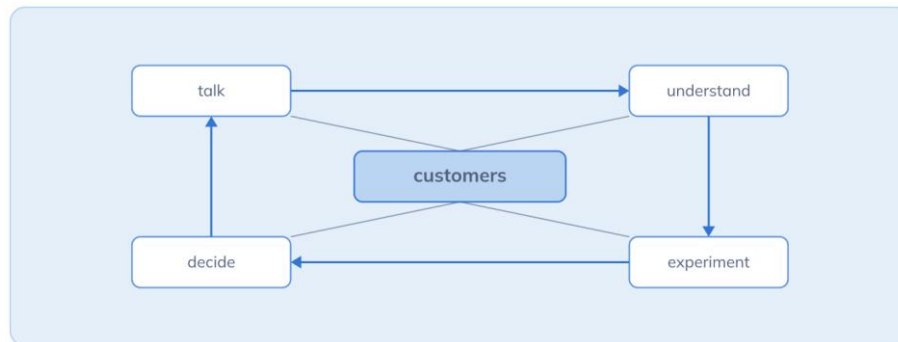


State of the art

Product discovery and customer validation

Reference map of current professional knowledge



Updated September 2026

Scrum Manager® · Skill Arena

About this document

This document is a curated map of the knowledge, practices and models in professional use as of its publication date for practicing product discovery and validating with customers what deserves to be built before building it, reducing uncertainty with useful evidence from customers and the market. For each entry, the map places it in context, indicates its degree of adoption in current professional practice and points to where to find reference information.

Use it as an observatory and reference point to check whether the professional knowledge you rely on is aligned with current practice and at the leading edge.

It is complemented by an in-depth companion guide that develops each concept and by the training and assessment platform in Skill Arena. In the [Product discovery and customer validation](#) area you can test your level of knowledge and, if you pass, obtain a diploma that provides curricular accreditation of professional competence and currency in this area.



State of knowledge

Knowledge about product discovery evolves progressively: its core has been proven for years and changes through refinement of practice rather than through breaks. What moves fast is its toolkit. Artificial intelligence already transcribes, synthesizes, and even interviews, and with it have come new practices, still scarce evidence, and legal obligations that are worth knowing before using it. Throughout the document, the labels (ESTABLISHED, CONSOLIDATING, EMERGING) help identify the maturity of each concept:

ESTABLISHED settled consensus; knowledge taken as required.

CONSOLIDATING gaining adoption fast; not yet universal but already relevant.

EMERGING recent frontier; high relevance and high volatility.

Block A — Talking to customers and capturing signal

The starting point of discovery: separating the problem from the solution, listening to the right people in order to learn rather than to confirm, and keeping a faithful record of what they said, with the care their data require. Without good input signal, everything that comes after is contaminated.

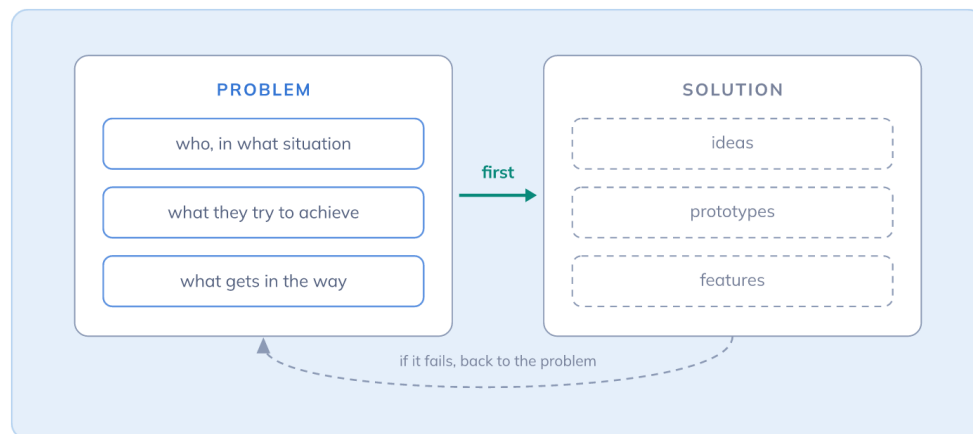


Figure 1. Two spaces that do not mix: the problem is understood before committing to a solution, and the team returns to it every time a solution fails.

Problem space and solution space · ESTABLISHED · new since 2026

Discovery requires keeping two spaces apart. The problem space covers who is affected, in what situation, what they are trying to achieve, and what stands in their way; the solution space covers the ideas, prototypes, and features we could respond with. Mixing them too early makes the conversation revolve around one of our own ideas and turns the search for evidence into a defense of it. The discipline consists in understanding the problem precisely before committing to a solution, and in returning to it every time a solution fails.

Why it is here now. It is the first capability this skill certifies and the origin of almost every later mistake, because a request such as “they need a dashboard” arrives with the solution already inside. Dan Olsen made the distinction explicit in *The Lean Product Playbook*, and Teresa Torres's continuous discovery runs through it from start to finish. It is an established principle that no practice in the field disputes; what fails is sustaining it once the team has grown attached to an idea.

Where to look. [Scrum Manager BoK — Product discovery \(in Spanish\)](#) · [Dan Olsen — The Lean Product Playbook](#)

The interview that learns, not the one that confirms · ESTABLISHED

A discovery interview explores the person's past behavior and real context; it is neither an oral survey nor a disguised sales pitch. Before writing the script, it helps to be able to complete the sentence “we want to learn X, and we will know it if we observe Y.” Useful

questions ask for concrete stories from the past (“tell me about the last time...”), which is the basis of Teresa Torres's story-based interviewing; useless ones ask for predictions about the future or opinions about our idea.

Why it is here now. It is the foundation of the discipline and its best-documented lesson: confirmation bias, designing questions that validate a solution already imagined, is still the most common failure. The technique has not changed, but its context has. AI-moderated interviews are now sold as a product, and the Nielsen Norman Group concluded in January 2026 that they work for gathering opinions about a product or screening participants, but not for exploring an unknown problem, which is precisely what this interview is for (Block F).

Where to look. [Rob Fitzpatrick — The Mom Test](#) · [Teresa Torres — Customer interviews](#)

Who to listen to: segment, sample, and recruitment · ESTABLISHED

What you learn depends on whom you listen to. Before recruiting, you define the segment of interest, with inclusion and exclusion criteria that are checked through a screener written so that it does not give away the answer being sought. Participants must be external and representative: your own network, colleagues, and the most enthusiastic customers all have incentives to please. And the sample is sized by its purpose, because a few interviews are enough to discover patterns, but not to measure how many customers share them.

Why it is here now. Recruiting badly invalidates the signal at its root, and no later technique fixes it: selection bias shows up just the same in the interview, in the prototype test, and in demand validation. It is a decision taken before any conversation, and more and more people without specific research training are taking it. According to Maze, in 2026 39% of respondents say that product managers in their organization already do research with users.

Where to look. [Nielsen Norman Group — Screening questions](#) · [Nielsen Norman Group — Interviewing users](#)

Quality notes: separating fact, quote, and interpretation · ESTABLISHED

During the interview and right after it, you separate what the person said (the quote), what they did or said they did (the fact, with its context of who, when, and in what situation), what the interviewer interprets, marked as such, and what remains open to explore. When the conversation is recorded and transcribed, the literal quote is no longer scarce and the value of the notes shifts to context and marked interpretation. If the notes sound like adjectives (“she loved it,” “he's happy”), concrete evidence is missing.

Why it is here now. A synthesis is only as good as the material that feeds it. Teresa Torres gave shape to the practice of closing notes while they are fresh with the interview snapshot: a one-page summary with a memorable quote, the opportunities, the insights, and an experience map, filled in during the 15 to 25 minutes after the conversation. Automatic transcription makes this easier, as long as the summary the tool generates is treated as one more interpretation and checked against the transcript.

Where to look. [Teresa Torres — Interview snapshots](#) · [Scrum Manager BoK — Interview snapshot \(in Spanish\)](#)

Consent, data, and transparency when talking to customers ·

CONSOLIDATING · *new since 2026*

Recording, transcribing, and analyzing an interview means processing personal data, because the voice and what is said can identify the person. You must inform people before and during recording, ask for specific consent for each recording, and check what the tool does with the data: if it uses them to retrain its model, that is a different processing operation with its own legal basis. Notes containing identifiable data are not pasted into an AI assistant without anonymizing them. And if the interviewer is an AI, the person must know it.

Why it is here now. What used to be good practice has had authoritative criteria since 2026. The Spanish Data Protection Agency (AEPD) published its criteria on AI voice transcription in January and April: consent for each recording, a visible indicator while recording, accuracy and rectification of the transcript, and the warning that emotion detection may fall into uses that the EU AI Act restricts or prohibits. Since August 2, 2026, the same Act also requires systems that interact with people to inform them that they are an AI. It is consolidating because professional practice around these criteria is still taking shape.

Where to look. [AEPD — AI voice transcription \(I\)](#) · [AEPD — AI voice transcription \(II\)](#) · [EU AI Act — Article 50](#)

Block B — From data to understanding

Turning scattered conversations into actionable learning: detecting patterns that support an insight and framing the problem without hiding the solution inside it. This is where much of the professional's distinctive value lies.

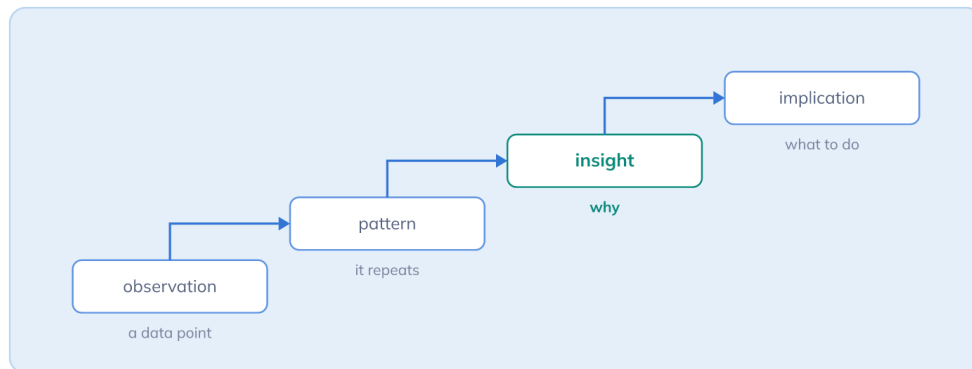


Figure 2. From observation to implication: synthesis climbs through levels, and the insight is the one that connects the pattern to a cause.

Synthesis: from observation to actionable insight · ESTABLISHED

Qualitative analysis climbs through four levels: the observation (a data point from one source), the pattern (what repeats across several sources), the insight (the understanding that connects the pattern to a cause), and the implication (what it makes sense to do as a result). An insight answers what we have learned that we did not know and what it means for the product, and it rests on recurrence across several sources, never on an isolated quote. Exceptions are treated as a possible signal of another segment, not as noise.

Why it is here now. Mistaking an interesting quote for an insight is the error that derails the most decisions, and AI has given it a new form. Teresa Torres warns against lazy AI interview synthesis, which consists of dumping the transcripts into an assistant and accepting its summary, and proposes synthesizing first, then asking the AI for the same synthesis and comparing the two (Block F). The group synthesis ritual has its own skill at Scrum Manager, Design Thinking; here the interest lies in its analytical side.

Where to look. [Nielsen Norman Group — Affinity diagramming](#) · [Skill Arena — Design Thinking](#)

Framing the problem: the job the customer wants to get done · ESTABLISHED

Defining the problem as “feature X is missing” puts the solution inside the problem and closes the exploration. The jobs to be done (JTBD) framework instead describes the progress the person is seeking in a specific situation, regardless of the solution. Its most widely used form in product work is the job story, which was born at Intercom with Paul Adams and named by Alan Klement: “when [situation], I want to [motivation], so I can

[outcome].” It is worth completing it with the alternative the person uses today, which sets the bar to beat.

Why it is here now. It is the bridge between evidence and product decisions, and the most effective defense against the feature factory. The framework has two schools worth distinguishing: that of Clayton Christensen and Bob Moesta, who understand the job as the progress a person is trying to make in a circumstance, and that of Tony Ulwick, whose Outcome-Driven Innovation (ODI) formulates the job as a verb with its object and measures the outcomes the customer desires. The job story template belongs to the first tradition.

Where to look. [Intercom — Job stories](#) · [Christensen Institute — Jobs to Be Done](#) · [Strategyn — Outcome-Driven Innovation](#)

Block C — Designing experiments to learn

Making risky assumptions explicit and designing the smallest test that puts them to the test: from assumptions mapping to the prototype test and to the minimum viable product (MVP) understood as an experiment. This is the area where activity is most often mistaken for learning.

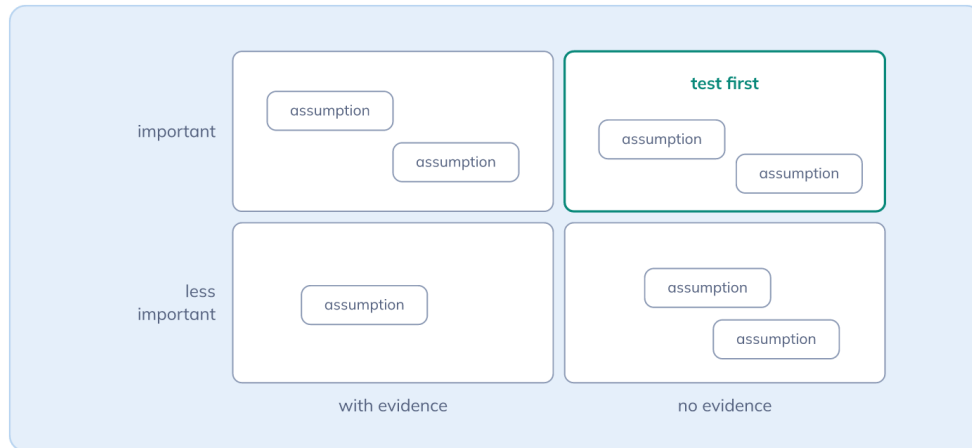


Figure 3. The assumptions map: what is important and still has no evidence is tested first.

Assumptions mapping: the risks to test first · ESTABLISHED

Every idea rests on assumptions that may turn out to be false. The three classic types are desirability (do they want it?), viability (is it a business?), and feasibility (can we build it?), and the reference sources add others, such as usability and ethics in Teresa Torres or adaptability in David Bland. They are prioritized in the importance-and-evidence matrix of Bland's Assumptions Mapping, which says to test first what is important and still lacks evidence, even if it is uncomfortable to test.

Why it is here now. The usual failure pattern is to test the easy part, almost always the technical one, and leave the critical part unexplored. Bland attributes the exercise to his work with Jeff Gothelf and Josh Seiden, authors of *Lean UX*, and spread it in *Testing Business Ideas*, with Alexander Osterwalder. Starting with desirability is the usual order, but the quadrant rules: in a product whose value depends on an AI model, feasibility may be the assumption that decides everything (Block F).

Where to look. [David Bland — Assumptions Mapping](#) · [Teresa Torres — Assumption testing](#) · [Marty Cagan — The four big risks](#)

Prototype tests: tasks as goals, not instructions · CONSOLIDATING

A prototype test only yields a usability signal if the tasks are framed as goals ("create a project called Q1 Test") rather than step-by-step instructions, if someone other than the prototype's designer moderates, and if the participants are external and representative. What counts is what the person does (where they click, where they get stuck, what they cannot find), not what they think. Fidelity is matched to the research question, and the minimum that answers it is enough.

Why it is here now. The theory is stable, but its disciplined execution is not universal, and 2026 adds two temptations. AI tools generate high-fidelity prototypes in minutes, and a prototype that looks finished invites opinions about the polish and attachment to it. Automated agents and moderators that promise to evaluate usability are also arriving; the research that trains them is advancing, but today they do not replace observing real people from the segment.

Where to look. [Steve Krug — Rocket Surgery Made Easy](#) · [Nielsen Norman Group — Usability testing 101](#) · [Scrum Manager BoK — Think-aloud protocol \(in Spanish\)](#)

The MVP as an experiment, not a reduced product · CONSOLIDATING

An MVP is the minimum needed to learn something specific, not a small version of everything. Before building, lighter alternatives are considered, chosen according to the hypothesis: a manual service in full view of the customer (concierge), human operation hidden behind an apparent interface (Wizard of Oz), a landing page, or a single-feature MVP. That the manual service does not scale is irrelevant at first, because if nobody wants it there is nothing to scale.

Why it is here now. The misunderstanding of the MVP as a list of components is still widespread, and AI reinforces it. Marty Cagan distinguishes building to learn from building to earn (April 2026) and notes that, with today's tools, ten or twenty prototypes a week are within almost anyone's reach, so the bottleneck is no longer building. In July he quoted Hilary Gridley: since it has never been faster to build, it has never been easier to run ten times faster in the wrong direction. Building cheaply does not exempt you from asking what is the minimum needed to learn.

Where to look. [Marty Cagan — Build to learn vs build to earn](#) · [Bland and Osterwalder — Testing Business Ideas](#) · [Scrum Manager BoK — Minimum Viable Product \(in Spanish\)](#)

Block D — Deciding with evidence

Closing the loop: setting the success criterion before looking at the data, weighing the strength of evidence against biases, and deciding whether to persevere, iterate, pivot, or stop. This is where discovery turns into decisions the team can defend.

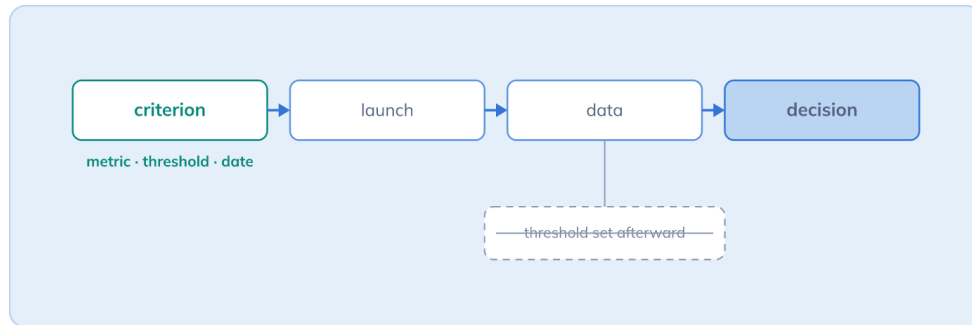


Figure 4. The criterion is set before launch: the success threshold is written down before the data arrive.

Success criteria and signals defined in advance · ESTABLISHED

The success criterion is set before launching the experiment: a single primary metric with its threshold, secondary metrics for context, and guardrail metrics, limits that must not be crossed even if the primary metric comes out well. Deciding after seeing the data which result looks good, or formulating the hypothesis in light of the results, is what statistics calls p-hacking and HARKing. And you have to recognize what the sample size allows you to conclude, because without a control group the evidence is an indication, not a causal relationship.

Why it is here now. Without a prior criterion, almost any result gets rationalized as a success. Teresa Torres sums it up as drawing a line in the sand before testing an assumption. The nuance not to lose is one of scale: a reliable A/B test with a control group requires traffic that most early discovery work does not have, as Ron Kohavi and his coauthors explain in *Trustworthy Online Controlled Experiments*, and in that case the honest thing is to declare the evidence directional.

Where to look. [Teresa Torres — Assumption testing](#) · [Kohavi, Tang, and Xu — Trustworthy Online Controlled Experiments](#)

The strength of evidence and bias control · ESTABLISHED

Evidence is stronger the more closely what is observed resembles a real decision with a cost for the person making it: an opinion weighs less than a sign-up, a sign-up less than a reservation, and a reservation less than a payment. Four common biases erode it (confirmation, survivorship, selection, and social desirability), and low-cost signals, such as likes, comments from your own network, or email opens, are often mistaken for validation.



Figure 5. The strength of evidence: evidence weighs more the more closely what is observed resembles a real decision with a cost for the person making it.

Why it is here now. It is the filter that separates rigorous discovery from validation theater. Alexander Osterwalder framed it as a gradient: evidence grows stronger the closer it gets to a real purchase, because what people say and what they do are different things. In 2026 a new source of false positives has appeared, the responses of AI-simulated users, which occupy the lowest rung of this scale (Block F).

Where to look. [Alexander Osterwalder — How strong is your innovation evidence?](#) · [Rob Fitzpatrick — The Mom Test](#) · [Scrum Manager BoK — Cognitive bias \(in Spanish\)](#)

Persevere, iterate, pivot, or stop · CONSOLIDATING

Faced with the evidence, there are four options: persevere in the current direction, iterate by adjusting specific elements, pivot by changing a fundamental aspect (the segment, the value proposition, the channel) without losing what was learned, or stop. Changing segment is one of the types of pivot Eric Ries described in *The Lean Startup*. The sunk cost fallacy, which justifies continuing because of what has already been invested, is the most dangerous bias; what counts is the best investment of the next resources.

Why it is here now. It connects discovery with the real product decision and with continuous discovery, where deciding is a recurring act and not a final milestone. It is consolidating because the range of options is accepted, but its explicit use is still maturing: documenting what was learned before changing course and weighing mixed evidence by its quality, its magnitude, the opportunity cost, and the reversibility of the decision.

Where to look. [Eric Ries — The Lean Startup](#) · [Teresa Torres — Product discovery basics](#)

Block E — Discovery as a team habit

Discovery works better as a weekly habit of the team that decides what to build than as a phase before development. This block covers how that habit is organized and the tool that connects what is learned with the outcome being pursued.

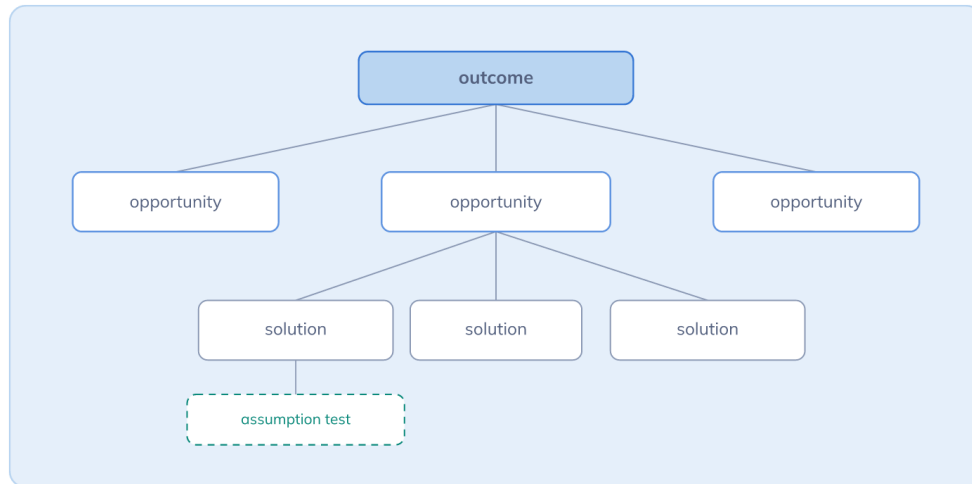


Figure 6. The opportunity solution tree: the outcome breaks down into opportunities, each opportunity into solutions, and each solution into tests of its assumptions.

Discovering every week: the product trio and continuous contact ·

CONSOLIDATING · new since 2026

Continuous discovery is, in Teresa Torres's definition, at least weekly contact with customers by the team building the product, through small research activities in pursuit of an outcome. It is sustained by the product trio: the person who leads the product, the person who designs, and the person who develops, who discover together instead of handing each other a report. Discovery and delivery run in parallel, what the field called dual-track, and what is learned is kept in a research repository the whole team can access.

Why it is here now. It has been the canonical practice of the field since *Continuous Discovery Habits* (2021) and it is still a minority one: in ProductPlan's survey of late 2025, only 34% of teams regularly collect customer insights and use them to prioritize, compared with almost 20% who rely on ad hoc requests and 12% with no process at all. The what of continuous discovery is taught in Scrum Manager's Product Owner certification; here the interest lies in how the cadence is sustained, who interviews, how recruitment happens every week, and how evidence is kept.

Where to look. [Teresa Torres — Continuous Discovery Habits](#) · [Scrum Manager BoK — Dual-track discovery & delivery \(in Spanish\)](#) · [Skill Arena — Product Owner](#)

The opportunity solution tree: from outcome to experiment ·

CONSOLIDATING · new since 2026

Teresa Torres's opportunity solution tree organizes discovery into four levels: at the top, the outcome the team is pursuing; below it, the opportunities (needs, pain points, and

desires that emerge from the interviews); then, the possible solutions for each opportunity; and at the bottom, the tests of each solution's assumptions. It forces the team to choose an opportunity before ideating, to compare several solutions for it, and to test assumptions instead of whole features.

Why it is here now. It is the tool that connects what is learned in interviews with the business outcome and keeps every conversation from triggering a feature. In 2026, AI-generated drafts of the tree built from real interviews have appeared. Torres defends them as a starting point the team reviews and warns against one-click trees made from invented data, because the point of the tree lies in synthesizing what is learned from real customers. The outcome at the root is defined with objectives and key results, which have their own skill.

Where to look. [Teresa Torres — Opportunity solution trees](#) · [Scrum Manager BoK — Opportunity solution tree \(in Spanish\)](#) · [Skill Arena — OKRs and product strategy](#)

Block F — Discovering with artificial intelligence

Artificial intelligence transcribes, summarizes, interviews, and even simulates customers. This block treats it as a source and as a filter of evidence: where it speeds discovery up, where it erodes it, and how to validate a product whose value depends on a model.

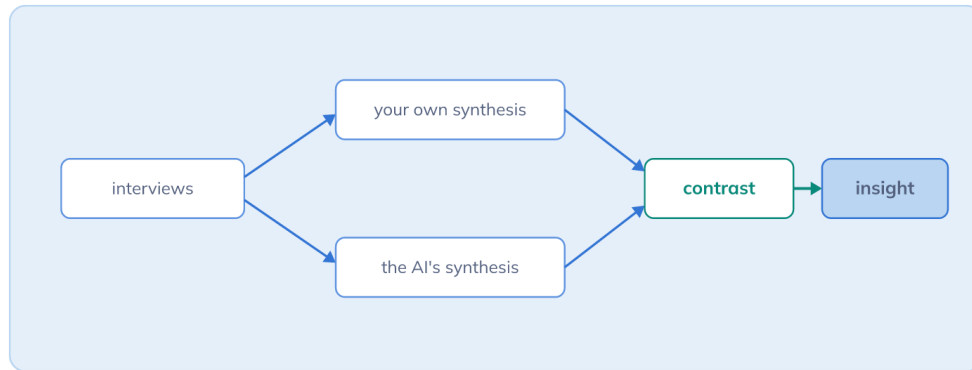


Figure 7. Compare, do not delegate: your own synthesis and the AI's are done separately and contrasted before asserting an insight.

Interviewing and synthesizing with AI: where it helps and where it erodes · CONSOLIDATING · new since 2026

In discovery work, AI already transcribes, summarizes, groups themes, and proposes draft syntheses, and there are platforms that moderate interviews without a human interviewer. Sensible use follows two rules. In synthesis, you compare instead of delegating: first your own synthesis, then the AI's, and finally the contrast between the two. In interviews, automated moderation works for opinions about a product, screening, and studies in several languages, but not for exploring an unknown problem.

Why it is here now. Adoption has been measured. According to Maze, in 2026 69% of research professionals use AI in at least some of their studies, and 82% hold that interpreting nuance and emotion is still human work. ProductPlan finds the same pattern: AI is used above all to summarize customer feedback (47%) and very little to prioritize problems (14%). The Nielsen Norman Group tested two AI interview platforms, and only 3 in 10 participants found the conversation natural. It is consolidating because the criteria for use already have sources, although the tools change every few months.

Where to look. [Nielsen Norman Group — AI-moderated interviews](#) · [Teresa Torres on airfocus — AI product management](#) · [Maze — Future of User Research Report 2026](#)

Where synthetic users sit on the scale of evidence · EMERGING · new since 2026

Synthetic users are AI-generated profiles that respond the way a customer segment would. They are useful for preparing the real work (desk research, hypotheses, draft interview scripts), but their answers are not evidence about customers: they are costless opinion, sycophantic and overly consistent, that captures general trends and fails on

their magnitude and variability. On the scale of the strength of evidence, they occupy the lowest rung.

Why it is here now. The tool is being offered insistently, and the evidence that is building up is negative. A study published on arXiv in July 2026 found that the models inflated differences between segments two to four times over and led to choosing the wrong segment in half of the US cases; another from the same month, with 37 models, measured an acquiescence bias of 0.84 standard deviations. 79% of researchers fear that stakeholders will overtrust these findings. It is emerging because the practice is in dispute and its permanence is not guaranteed.

Where to look. [Nielsen Norman Group — Synthetic users](#) · [Chen, Zhu, and Zheng — When synthetic users fail \(arXiv, 2026\)](#) · [User Interviews — State of Synthetic Users](#)

Validating a product whose core is a model · EMERGING · *new since 2026*

When the product's value depends on an AI model, one more question joins the usual ones: whether it delivers value with a reliability that is acceptable to the person using it. Discovery expands with error analysis, which consists of observing with real customers where the system fails, turning those failures into evals, and accepting that the criteria change as the team learns. Feasibility stops being a technical question settled at the end and becomes an assumption tested early.

Why it is here now. Teresa Torres and Petra Wille framed it in September 2025: error analysis works as a discovery activity, and Torres has gone as far as calling evals a new discovery habit. It is emerging because the practice is still taking shape. Here the interest lies only in the stretch of validating value with acceptable reliability; the operation of evals before and after launch is covered by the Product Owner certification and the Harness Engineering skill.

Where to look. [Torres and Wille — AI evals & discovery](#) · [Skill Arena — Harness Engineering](#) · [Skill Arena — Product Owner](#)

A note on the references

The links included were available and verified as of the update date. Given the pace of change in this area, some may change; each revision of the map also updates its references.

© 2026 Scrum Manager®. This work is published under a Creative Commons Attribution-NonCommercial 4.0 International licence (CC BY-NC 4.0). Official Scrum Manager trainers and centres are licensed under the terms of CC BY 4.0 for their training activity.