AI AND RESEARCH

AI makes research faster, not truer: the method decides the quality

Collection, structuring and first drafts can be automated. Framing the questions, judging the sources and the final synthesis cannot.

KEY MESSAGE

AI makes research faster, not truer: method, sources and review decide the quality, so buy it from specialists who prove all three.

3 things to know

1. **AI output is a hypothesis to be tested** [1]; quality comes from method, not from the model.
2. **Errors depend on the task:** under 1% on short summaries, a 15x overestimate in one market sizing [1][2].
3. **Buy research from best-of-breed specialists and judge them on method:** sources, double sizing and senior review, not page count.

EXHIBIT 1 What AI research can and cannot do today (CarveLines view; dots = reliability)

TASK	EVIDENCE	RELIABILITY
Summarise	<1% errors on short, clean texts [2]	●●●
Collect, structure	Fast and broad; sources need checking	●●●
Competitor map	Good draft; misses private players	●●○
Synthetic survey	14.5 pt mean error vs real panel [3]	●○○
Market sizing	15x overestimate in one test [1]	●○○
Specialist query	69–88% errors, 2024 legal study [2]	●○○

Sources [1][2][3]; reliability dots are a CarveLines assessment.

15x
AI overestimate in a market-sizing test [1]

<1%
hallucination on short, clean summaries [2]

14.5 pts
mean error of AI-simulated survey answers [3]

What can be automated

Searching, reading and extracting: filings, websites, job postings, reviews, press, trade data. Structuring: tables of competitors, prices, products and deals. First drafts: market overviews and summaries of long documents.

These steps used to take most of an analyst's week. They now take hours, with wider coverage than any team could read.

What stays human

Choosing the questions that matter for the investment case. Deciding which source wins when two disagree. Building the sizing logic and testing it against bottom-up evidence. Writing the so-what.

AI can propose all four; it cannot own them. One CDD practitioner's rule: keep LLMs away from quantitative forecasting until their logic is traceable [1].

Where AI research adds most

Breadth: scanning hundreds of companies, reviews or job postings no team would read. Speed: a structured first view of an unfamiliar sector in a day. Consistency: the same checklist on every target.

Depth still comes from people: expert calls, customer interviews, site visits, and the judgement of someone who has seen the sector before.

Why errors hide

Benchmarks flatter. The best models hallucinate below 1% on short, clean documents, but that is measured under controlled conditions [2]. Deal research is long, messy and specialised.

Across a workstream that touches hundreds of sources, small error rates compound [1]. A wrong number reads as convincingly as a right one; only traceability to a source exposes it.

Sources on page 3.

CarveLines © 2026

A method that holds

Every figure tied to a source and a date. Estimates labelled as estimates. Two independent routes to each key number, top-down and bottom-up. A reviewer who reads the sources, not just the slides.

Synthetic surveys, if used at all, serve to design real ones: a 2025 study found a 14.5-point mean error against a real panel of 2,000 adults [3]. Whether research is done in-house or bought, these are the standards to demand from every provider.

EXHIBIT 2 **Three ways to get a market fact base (illustrative, CarveLines view)**

	TRADITIONAL RESEARCH	GENERIC AI ASSISTANT	SPECIALIST AI-NATIVE
Elapsed time	Weeks	Minutes to hours	**Days**
Source coverage	Deep on chosen sources	Broad, unchecked	**Broad and checked**
Key numbers	Modelled by analysts	Often unreconciled	**Two routes, reconciled**
Auditability	Partial footnotes	Links may not exist	**Every fact with URL and date**
Error detection	Senior review	Rarely	**Automated checks plus senior review**
Investment framing	Strong	Generic	**Built on the deal's questions**
Cost	Highest	Lowest	**In between**

Illustrative comparison, not measured results. Qualities vary by provider; test any provider on closed deals.

> "The secondary research that AI produces is a hypothesis to be tested, not a conclusion to be delivered."
>
> **Rick Claar**, President, Martec Group · 2026 [1]

> "Until commercial LLMs improve the traceability of their logic, forget about using AI for quantitative analysis and forecasting."
>
> **Nicolas Mialaret**, CDD practitioner, quoted by Martec Group · 2026 [1]

5 hints: the average fund vs full power

	THE AVERAGE FUND	FULL POWER
1	Asks for a market overview	**Asks for a question tree, then evidence for each branch**
2	Trusts a single sizing	**Sizes twice, top-down and bottom-up, and explains the gap**
3	Reads the AI summary	**Reads a sample of the sources behind it**
4	Uses synthetic surveys as data	**Uses them only to design real surveys**
5	Keeps findings in the deck	**Stores verified facts in a reusable fact base**

Watch out

Polished output. A 60-page AI report with clean charts invites a trust it has not earned. Judge it on its sources, not its layout.

Questions for the partners' meeting

1. For our last three IC memos, can every key number be traced to a source?
2. Who reviews AI-produced research, and against which checklist?
3. Where do we rely on AI for numbers rather than for collection?
4. What did AI get wrong on a recent deal, and what rule did we add?

Sources on page 3.

CarveLines © 2026

7 recipes: who, how, what you get

01 Question tree first

DEAL TEAM

- State the investment question in one sentence
- AI proposes a question tree
- Prune to the ten questions that change the decision
- Give each question a source plan

Output research plan
Time 1 hour

02 Sourced fact base

RESEARCH PARTNER + ANALYST

- Brief a best-of-breed research provider with the question tree
- Require every fact with URL and date, estimates labelled [E]
- Analyst checks a random 20% at source
- Store in a shared fact table

Output fact table
Time 48 hours

03 Double sizing

ANALYST

- Top-down from reports
- Bottom-up from customers, volumes and prices
- Explain any gap above 30%
- Present a range, not a point

Output sizing with a range
Time 1 day

04 Source audit

ASSOCIATE

- Pick 20% of citations at random
- Click through and check each claim
- Log errors by tool and task
- Add a rule where errors cluster

Output error rate per tool
Time 2 hours

05 Competitor map at scale

RESEARCH PARTNER

- AI scans registries, websites and job posts for all players
- Classify by segment and size
- Check the top 30 by hand
- Publish the map with sources and date

Output competitor map
Time 2 days

06 Customer voice

RESEARCH PARTNER

- AI reads public reviews and forums by segment
- Codes themes and counts mentions
- Pulls verbatim examples
- Compares with management's claims

Output themes with quotes and counts
Time half a day

07 Expert-call synthesis

DEAL TEAM

- Map each transcript to the question tree
- Flag contradictions between experts
- Update answers and confidence levels
- Decide whether another call is needed

Output updated answers
Time 1 hour per call

Monday morning

Take the last market sizing your team presented at IC. Ask an AI tool to rebuild it from scratch with sources, then compare. Where the two differ is where your method needs a rule.

Sources: [1] Martec Group, The impact of AI on commercial due diligence, 2026; [2] Development Corporate, AI hallucination rates are a due diligence crisis (citing Vectara and Stanford RegLab/HAI), Mar 2026; [3] Verasight study (n=2,000 US adults), Dec 2025, cited by Martec Group. Quotes verbatim from the sources cited. Figures as published; survey scopes and dates vary. Vendor-reported figures are labelled; [E] = CarveLines estimate; 'CarveLines view' = our interpretation.

CarveLines © 2026