

Æ / WHITE PAPER / 10.2026

The Judgement Layer

Why creative thinking decides what AI is worth

AN Æ WHITE PAPER · 23 SEPTEMBER 2026

Executive summary

AI has made competent creative work abundant. The advantage now belongs to organisations whose AI is shaped by people with taste and critical judgement, and artists are trained in exactly that.

This paper sets out the evidence. It draws on peer-reviewed research, pre-registered field experiments with more than 1,500 professionals, meta-analyses covering thousands of campaigns and 24 years of effectiveness data, and documented outcomes from Heinz, Cadbury, Nutella, Coca-Cola, Google and Klarna. It addresses the findings that complicate the case as well as those that support it.

FIVE FINDINGS

01 **Creative quality is the largest single driver of advertising sales effect.**

Across nearly 500 campaigns, creative accounted for 47% of sales lift, more than reach, brand or targeting [1], a share unchanged in a later meta-study [2]. Creatively awarded campaigns have grown market share up to 11 to 12 times more efficiently than non-awarded ones [4][5].

02 **The market rewards standing out, not being competent.**

Top-quartile design performers grew revenue 32 percentage points faster than peers over five years, while the differences between the other three quartiles were marginal [3].

03 **Generic AI pulls work towards the average.**

In a controlled experiment published in Science Advances, AI assistance made individual stories better but made them more similar to one another [8].



04 **Human judgement decides whether AI helps or harms.**

In a field experiment with 758 BCG consultants, AI raised quality and speed on suitable tasks but made people 19 points less likely to reach a correct answer on an unsuitable one [9]. Higher confidence in AI is associated with less critical thinking [11].

05 **Audiences penalise AI that replaces judgement and reward AI that serves an idea.**

NielsenIQ found AI-generated ads drew negative responses and weaker memory activation [13]. Yet Heinz's AI campaign earned more than a billion impressions [19] and Cadbury's AI-personalised ads lifted sales 35% [20]. The difference was the idea, and the people directing it.

The pattern in the real world

Where AI executed a human creative idea, results were exceptional. Where AI was the idea, or replaced human judgement, brands pulled ads within days [27] or rehired the people they had cut [29]. Section 4 documents six cases.

What this means

If AI lifts everyone to the same baseline, the baseline stops being an advantage. What remains scarce is the judgement to frame the right problem, recognise originality and reject the plausible but generic. Æ builds AI products, agents and workflows designed by artists, designers and thinkers so that this judgement is built into the system rather than bolted on afterwards.

Who this paper is for: marketing leaders, creative agencies, procurement and operations teams, and investors. Section 8 speaks to each directly.



INFOGRAPHIC · THE ARGUMENT

The evidence, link by link



47%

01 · CREATIVE DRIVES SALES

of advertising sales lift comes from creative quality, more than reach, brand or targeting [1]

+32 pp

02 · DISTINCTIVE WINS

faster revenue growth for top-quartile design performers; the other quartiles barely differ [3]

Converge

03 · GENERIC AI CONVERGES

AI-assisted stories were individually better but more similar to one another [8]

−19 pts

04 · JUDGEMENT DECIDES

in correct answers when consultants used AI on a task outside its frontier [9]

+35%

05 · IDEAS, NOT TOOLS

Cadbury sales, where AI executed a human idea under creative direction [20]

SECTION

01

The abundance problem

Generative AI has solved the problem of producing competent creative work at speed. That success has created a new problem: when every competitor can produce competent work, competence no longer differentiates.

Adoption is mainstream, and mostly about efficiency. Industry reporting puts the share of connected-TV, social and online video ads made or enhanced with generative AI at 30% in 2025, up from 22% the year before [31]. Forecasts for generative AI in marketing alone put the market at roughly USD 6.6 billion in 2026, rising to around USD 18 billion by 2030 [32]. But a 2026 study by the World Federation of Advertisers and LIONS, covering more than 160 senior marketers and USD 141 billion of annual spend, found that only one in three is using AI to go beyond efficiencies and deliver better creative work, and only three in ten say their teams consistently deliver creative excellence at all [6]. Nearly two-thirds say they are still struggling to develop new ways of working around the technology [7].

The tools converge on the same output. Most teams use the same small group of foundation models, trained on overlapping data and prompted in similar ways. Controlled research shows that outputs become more alike as more people draw on the same model [8]. For brands, whose value depends on being distinctive and recognisable, this is a structural risk rather than a cosmetic one.

The efficiency trap has a precedent. The most-cited corporate AI success story of 2024 was Klarna, which reported its AI assistant doing the work of 700 agents. By May 2025 its chief executive said the company had focused too much on efficiency and cost, that the result was lower quality, and that it was hiring humans again [29]. The lesson generalises: gains that show up in speed and cost metrics can hide losses in quality that customers notice first.

The question for marketing leaders is therefore no longer whether to use AI. It is how to use AI without becoming indistinguishable from everyone else who uses it.

30%

of CTV, social and online video ads made or enhanced with generative AI in 2025 [31]

1 in 3

senior marketers using AI to go beyond efficiency and deliver better creative [6]

3 in 10

say their teams consistently deliver creative excellence at all [6]



INFOGRAPHIC · ADOPTION VS AMBITION

Adoption is up. Ambition isn't.



22%

2024



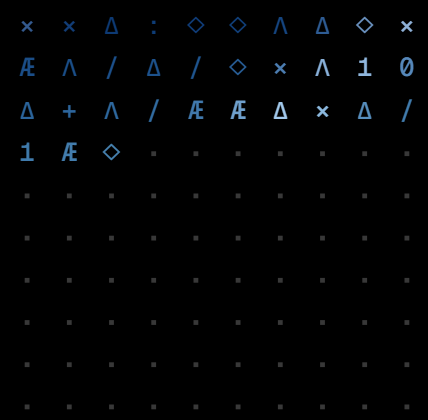
30%

2025

SHARE OF CTV, SOCIAL AND ONLINE VIDEO ADS MADE OR ENHANCED WITH GENERATIVE AI

SOURCE: IAB VIA SECONDARY REPORTING [31]

OF 160+ SENIOR MARKETERS CONTROLLING USD 141BN IN ANNUAL SPEND



1 in 3

use AI to go beyond efficiency and deliver better creative work [6]



3 in 10

say their teams consistently deliver creative excellence [6]



~2 in 3

are still struggling to develop new ways of working around AI [7]



SECTION

02

Why creative judgement is the missing layer

AI changes where human effort matters most: less on producing, more on directing and judging. That shift plays directly to the skills artists and designers spend their careers developing.

The work has moved from making to judging. A study of 319 knowledge workers, covering 936 real examples of AI use, found that AI shifts critical thinking towards verifying information, integrating AI responses and stewarding the task as a whole [11]. The human role becomes deciding what is worth keeping.

Judgement is a skill, and misplaced confidence erodes it. The same study found that people with higher confidence in AI reported less critical thinking, while those with higher confidence in their own abilities reported more [11]. A workforce that trusts the tool more than its own judgement will accept the plausible and generic. Google's 2024 Olympics ad, which showed a father asking AI to write his daughter's fan letter, tested well before airing and was pulled within days of broadcast [27]. Pre-testing measured emotional response; it did not measure whether the idea itself was wise.

How people work with AI varies, and the variation matters. A field study of 244 BCG consultants identified three distinct modes of working with AI, defined by who decides what needs doing and who decides how it is done [12]. Value comes from the structure of the collaboration, not from access to the tool.

Creativity pays only when it is directed well. Twenty-four years of IPA effectiveness data show that creatively awarded campaigns were around 12 times as efficient as non-awarded ones through 2008, but that this advantage collapsed to under four times as the industry shifted to short-term, activation-led work [4]. Creativity is not a magic ingredient. It produces returns when strategic judgement decides how it is used.

THREE CAPACITIES ARTISTS BRING

The following is Æ's argument, grounded in the evidence in Sections 3 and 4:



Framing

Deciding which problem is worth solving, and what an original answer would look like, before any prompt is written. Heinz's agency did not ask AI to make an ad; it asked AI what ketchup looked like, and built a global campaign on the answer [18].



Discernment

Recognising the difference between work that is merely competent and work that is distinctive, and rejecting the former. Coca-Cola's 2025 holiday ad was assembled by around 100 people from 70,000 AI-generated clips [25]; the value was in the selection.



Critical distance

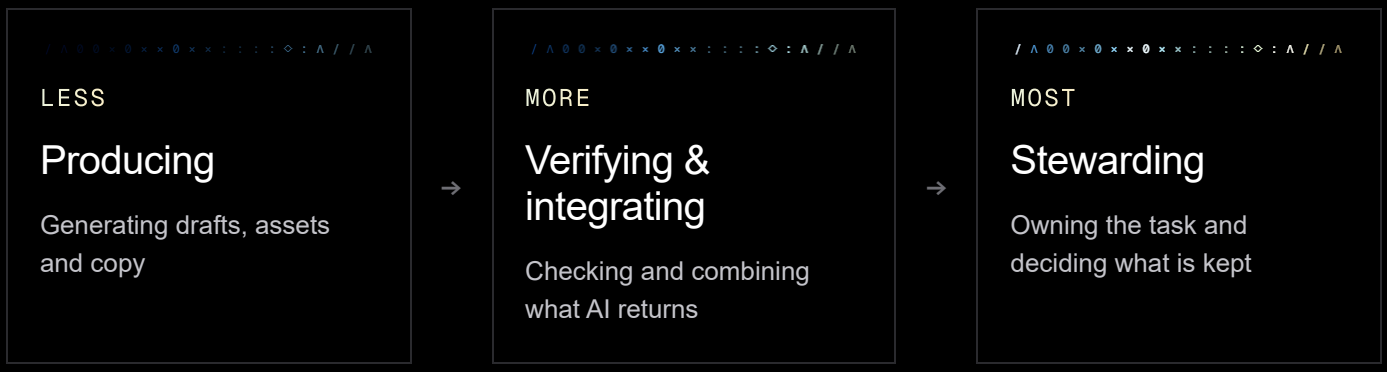
Questioning defaults, assumptions and the aesthetic tendencies built into a model, rather than inheriting them.

These are not soft extras. They are the capacities the research identifies as deciding whether AI improves or degrades the work.



INFOGRAPHIC · WHERE THE EFFORT GOES

Effort moves from making to judging



319 KNOWLEDGE WORKERS, 936 REAL EXAMPLES OF AI USE · SOURCE: LEE ET AL. (2025) [11]

INFOGRAPHIC · CREATIVITY'S MULTIPLIER

The payoff depends on direction



~12×

Awarded vs non-awarded efficiency, to 2008



<4×

After the shift to short-term activation

Creativity is not a magic ingredient. It earns its multiple when strategic judgement decides how it is used.

IPA DATABANK, ~600 CASES OVER 24 YEARS [4]

SECTION

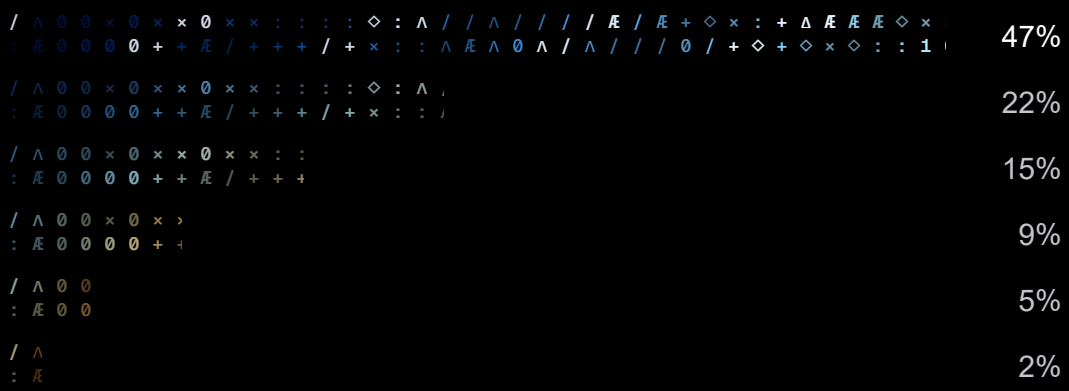
03

The evidence

Five bodies of evidence, taken together, support the case. None alone proves that artist-led AI outperforms, but each establishes one link in the chain, and Section 4 shows the chain holding in the market.

F1 Creative quality drives commercial results

Creative quality is the single largest driver of the sales effect of advertising. Nielsen's analysis of nearly 500 campaigns attributed 47% of sales lift to creative, more than reach, brand and targeting combined with recency [1]. The finding has held over time. A later NCSolutions meta-study of around 450 CPG campaigns found creative's contribution unchanged at 49% [2]. Nielsen also found that weak creative produces weak sales lift regardless of media spend [1].

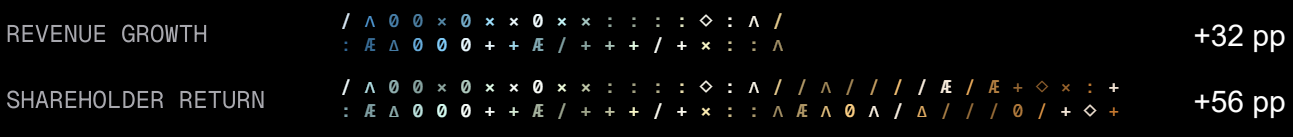


SHARE OF SALES LIFT BY DRIVER · SOURCE: NIELSEN (2017) [1]

The IPA's effectiveness databank, covering almost 600 case studies over 24 years, adds the long-term view. Creatively awarded campaigns grew market share around 11 to 12 times more efficiently than non-awarded ones through 2008 [4][5]. That multiple fell to under four as the industry shifted towards short-term activation, which is a warning as much as an endorsement: creativity delivers returns when it is deployed with strategic judgement, and loses them when it is not [4].

F2 The market rewards distinctiveness, not competence

McKinsey tracked design practices and financial results across 300 listed companies over five years. Top-quartile design performers achieved 32 percentage points higher revenue growth and 56 points higher shareholder return growth than peers [3].



TOP-QUARTILE DESIGN PERFORMERS VS PEERS, 5 YEARS · SOURCE: MCKINSEY (2018) [3]

The critical detail is that the differences between the second, third and fourth quartiles were marginal [3]. The market disproportionately rewarded companies that genuinely stood out. Being good was not enough; being distinctive was. This is a correlation study, and should be read as such.

F3 Generic AI pulls work towards the average

In a controlled experiment published in Science Advances, writers given AI-generated ideas produced stories rated more creative, better written and more enjoyable, with the largest gains among less creative writers [8]. But AI-assisted stories were also more similar to each other than stories written by people alone [8].

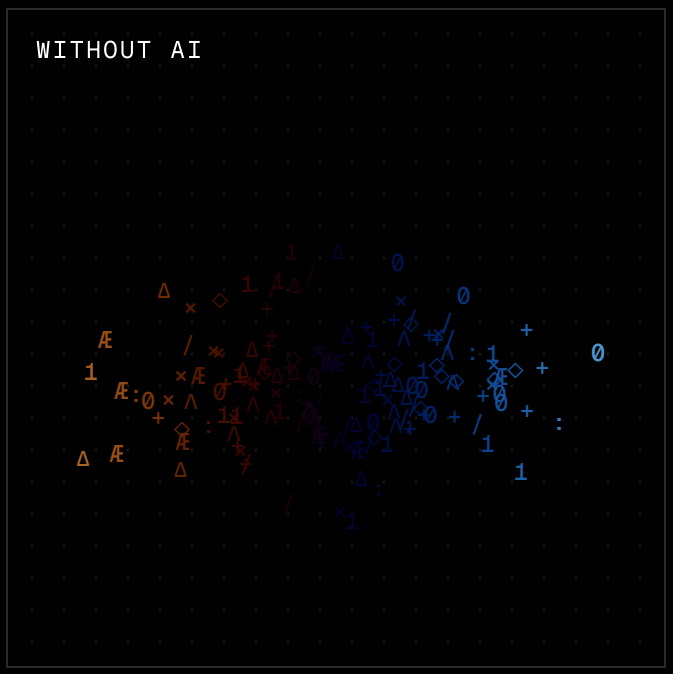
The authors describe this as a social dilemma: each individual is better off, but collectively a narrower range of novel work is produced [8]. For a brand competing on distinctiveness, the collective effect is the one that matters.

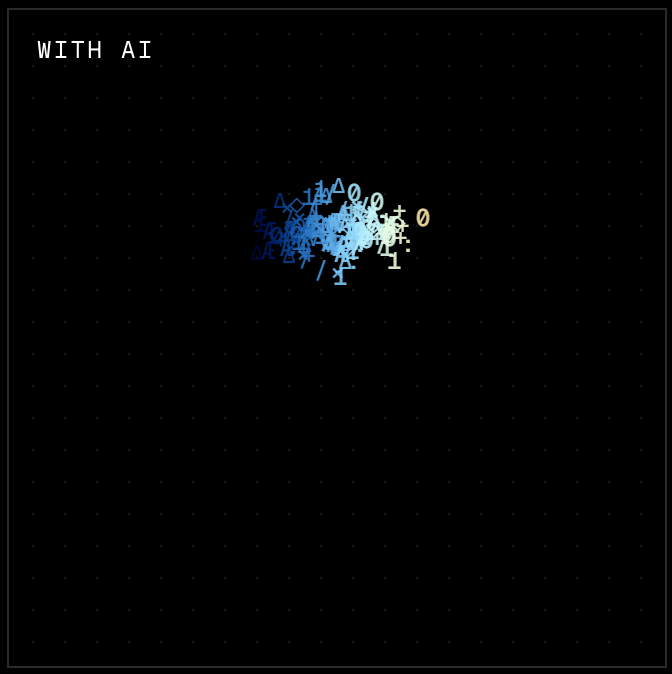


INFOGRAPHIC · F3 GENERIC AI PULLS WORK TOWARDS THE AVERAGE

Better alone, more alike together

Each point is one piece of work. With AI, individual pieces score higher, but the cloud of ideas tightens around the same centre: a narrower range of novel work overall.





Higher

rated creativity, writing quality and enjoyment, largest for less creative writers

Narrower

stories more similar to each other than those written by people alone

ILLUSTRATIVE, NOT PLOTTED DATA · FINDINGS FROM DOSHI & HAUSER, SCIENCE ADVANCES (2024) [8]

F4 Human judgement decides whether AI helps or harms

The largest field experiment on AI and knowledge work, involving 758 BCG consultants, found that AI's value depends entirely on the task [9]. On tasks within AI's capabilities, consultants completed more work, faster and at higher quality. On a task deliberately set outside those capabilities, they were significantly less likely to get the right answer [9].

			OUTSIDE FRONTIER
+12.2%	25.1%	40%	−19 pts
more tasks completed	faster	higher quality	likelihood of a correct answer

BCG CONSULTANTS WITH AI VS WITHOUT · SOURCE: DELL'ACQUA ET AL. (2023) [9]

The researchers call this the jagged frontier: AI is strong at some tasks and weak at others of similar apparent difficulty, with no obvious signal of which is which [9]. Knowing where that line falls is a matter of expert judgement.

A follow-up field experiment with 791 professionals at Procter & Gamble found that individuals using AI matched the performance of two-person teams without it [10]. AI can extend what a skilled person does, which is precisely why the skill of the person matters.

F5 Audiences notice generic AI output

NielsenIQ's neuroscience research found that consumers readily identified AI-generated ads and responded with more negative sentiment, weaker memory activation and lower intent to act than for conventionally made ads [13]. AI ads were consistently rated more annoying, boring and confusing, and even high-quality AI output left a weaker impression [14].



This finding needs qualifying, and Section 4 does so. Coca-Cola's AI-made holiday ads scored the maximum rating in System1's effectiveness testing [23], but respondents were not told AI was involved, and System1 itself said the scores would likely have been lower had they known [24].

The evidence points to a consistent conclusion: audiences reward strong, distinctive ideas, and penalise AI when it becomes visible as a substitute for them.

Supporting evidence on expertise

Research on designers suggests expertise changes how AI is used rather than whether it helps. In one comparative study, AI mainly helped novice designers generate ideas, while for experienced designers it contributed more to the quality and refinement of the final work [15]. A controlled experiment comparing professional artists with laypeople using image-generation tools found that expertise produced better results, though the margin was modest [16].



SECTION

04

What the market has already shown: six cases

The research predicts a pattern. Public results from six brands confirm it: AI produced exceptional outcomes when it executed a strong human idea under creative direction, and produced reversals when it replaced judgement.

CASE	YEAR	WHAT AI DID	HUMAN JUDGEMENT	OUTCOME
Heinz, “A.I. Ketchup” (Rethink)	2022	Generated images from the prompt “ketchup”	The insight that AI would draw Heinz, and the campaign built on it	850m+ earned impressions, worth over 2,500% of media spend [18]; later reported at 1.15bn with a 38% engagement lift [19]
Cadbury, “Shah Rukh Khan-My-Ad” (Ogilvy Mumbai, Rephrase.ai)	2021	Synthesised a film star's face and voice for thousands of personalised ads	The idea of lending Cadbury's ambassador to small shops	130,000 ads created, consideration +7.3%, sales +35% [20]; Cannes Creative Effectiveness Grand Prix 2023 [21]

CASE	YEAR	WHAT AI DID	HUMAN JUDGEMENT	OUTCOME
Nutella, “Unica” (Ogilvy Italy)	2017	Algorithm generated 7 million unique jar labels	Designers curated the patterns, palettes and rules the algorithm combined	All 7 million jars sold out within a month [22]
Coca-Cola, “Holidays Are Coming” (Secret Level, Silverside AI, WPP Open X)	2024–25	Recreated the 1995 Christmas ad with generative video	A distinctive asset created by humans 30 years earlier; around 100 people selecting from 70,000 clips [25]	Maximum 5.9 System1 score, but sustained backlash from the creative community [23]
Google, “Dear Sydney”	2024	Featured as the product: AI writing a child's fan letter	Missing: nobody questioned whether the idea should run	Tested well pre-air; pulled from Olympics coverage within days [27]
Klarna, AI customer service	2024–25	Replaced the work of around 700 agents	Removed, then reinstated for complex cases	CEO admitted lower quality and began rehiring humans in 2025 [29]; hybrid model now claims USD 60m saved [30]



INFOGRAPHIC · SIX CASES, 2017-2025

Where AI served an idea, and where it replaced one

ABOVE THE LINE · AI EXECUTED A HUMAN IDEA BELOW · AI REPLACED JUDGEMENT

2017 · WORKED

Nutella

7 million unique jars, sold out within a month

2021 · WORKED

Cadbury

130,000 personalised ads; sales +35%

2022 · WORKED

Heinz

850m+ earned impressions, 2,500% of media spend

2024-25 · MIXED

Coca-Cola

Top System1 score; sustained creative backlash



2024 · REVERSED

Google

“Dear Sydney” pulled within days of airing

2024-25 · REVERSED

Klarna

Quality fell; humans rehired for complex work

SOURCES: [18]-[30] · POSITIONS BY YEAR

Heinz: AI as the proof of a human insight

Heinz's agency, Rethink, had already run a campaign asking people to draw ketchup, and found they drew Heinz. In 2022 the team asked DALL·E 2 the same question, and it did too, even for prompts like “ketchup tarot card” [18]. The campaign generated more than 850 million earned impressions, worth over 2,500% of its media investment, and went on to win major creative awards [18].

The AI produced the images, but the idea came from a human understanding of what the brand owned. Without the insight, the same tool would have produced generic pictures of condiments.

Cadbury: generative technology in service of a generous idea

During the pandemic, Ogilvy Mumbai used machine learning from Rephrase.ai to recreate Shah Rukh Khan's face and voice so that thousands of small Indian shops could have the country's biggest star advertise their store by name [20]. The campaign produced 130,000 unique videos, lifted consideration by 7.3% and sales by 35% [20]. It won the Titanium Lion in 2022 and the Creative Effectiveness Grand Prix in 2023 [21].

The creative team designed the idea around what the technology could do. The technology did not generate the idea.

Nutella: designers set the constraints, the machine scaled them

In 2017, before generative AI, Ogilvy Italy built an algorithm that combined dozens of designer-created patterns and thousands of colour combinations into 7 million unique Nutella labels, each with an authentication code [22]. Every jar sold within a month [22].

This is the model for artist-led AI in miniature. Humans defined the visual language and the rules; the machine supplied variety at a scale no studio could. The result was distinctive precisely because the constraints were designed by people who understood the brand.



INFOGRAPHIC · BY THE NUMBERS

Three wins, one pattern

Human idea first, technology second, people with taste deciding what ships.

HEINZ · “A.I. KETCHUP” · 2022

850m+

earned impressions at launch [18]

1.15bn

impressions in later summaries [19]

2,500%

earned value vs media spend [18]

CADBURY · “SHAH RUKH KHAN-MY-AD” · 2021

130,000

unique personalised ads [20]

+7.3%

consideration [20]

+35%

sales [20]

NUTELLA · “UNICA” · 2017

7m

unique jar labels from designer rules [22]

1 month

to sell out every jar [22]

0

generative AI: an algorithm, curated by designers



Coca-Cola: the asset did the work, not the technology

Coca-Cola's AI-generated remakes of its 1995 “Holidays Are Coming” ad drew strong criticism from creative professionals in both 2024 and 2025, yet scored the maximum 5.9 in System1's effectiveness testing [23]. Three details matter. First, respondents were not told AI was involved, and System1 said the reaction would likely have been more negative had they known [24]. Second, the ad's power came from distinctive assets, the red trucks and the melody, created by human teams three decades earlier. Third, the 2025 version required around 100 people to select from 70,000 generated clips [25], and System1's own analysts warned that AI-driven novelty wears out fast, noting that Google's Gemini ad fell from 4.4 to 3.7 stars in a year [26].

The case is often cited as proof that consumers do not care about AI. The fuller reading is that consumers care about strong ideas and distinctive assets, and that AI can reproduce them well when human judgement guides every selection.

Google: pre-testing is not judgement

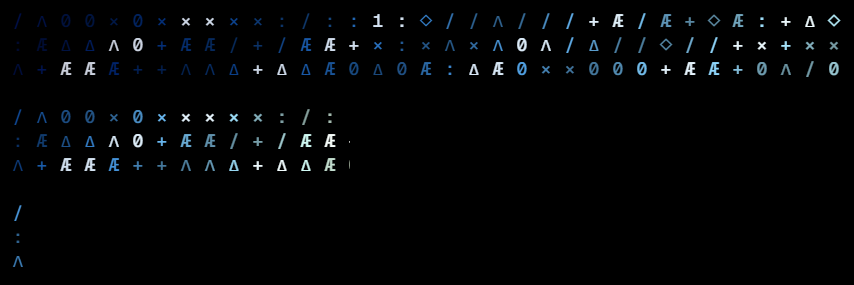
Google's “Dear Sydney” ad showed a father asking Gemini to write his daughter's letter to an Olympic athlete. It tested well before airing, scoring a strong 4.4 in System1's model [28], and was pulled from Olympics coverage within days after widespread criticism that it replaced a child's own words with a machine's [27].

No production flaw caused the failure. The problem was the idea, and no testing framework flagged it. That is the kind of judgement a creatively literate team supplies before an ad is made.



INFOGRAPHIC · COCA-COLA, 2025

Selection is the work



GENERATED CLIPS	70,000
PEOPLE SELECTING	~100
FINAL FILM	1

LOG SCALE · SOURCE: SYSTEM1 (2025) [25]

INFOGRAPHIC · SYSTEM1 STAR RATINGS, MAXIMUM 5.9

Testing measures response, not wisdom



5.9

Coca-Cola holiday ad, AI-made, respondents not told [23][24]



4.4

Google “Dear Sydney” pre-air. Pulled within days [27][28]



4.4

Google Gemini ad at launch [26]



3.7

Same Gemini ad one year later: novelty wears out [26]



Klarna: the efficiency trap, at scale

Klarna is not a creative case, but it is the most public natural experiment in replacing human judgement with AI. In 2024 the company said its AI assistant did the work of 700 agents. In May 2025 its chief executive said the company had focused too much on efficiency and cost, that quality had fallen, and that it was recruiting humans again so customers could always reach a person [29]. By late 2025 Klarna described a hybrid model in which AI handles volume and people handle complexity, and still reported USD 60 million in savings [30].

The pattern is the one this paper predicts. AI is valuable at scale. The value survives only when humans keep the judgement-heavy work.



WHAT THE CASES HAVE IN COMMON

In every successful case, three things were true: a human team framed the idea; the technology was chosen to serve it; and people with taste decided what was good enough to release. In both reversals, at least one of those was missing.

That is the distinction between AI as a tool for creative people and AI as a substitute for them.



INFOGRAPHIC · KLARNA, 2024-2025

The efficiency trap, in three moves

2024

AI does the work of 700 agents

Speed and cost metrics improve [29]

May 2025

Quality falls; humans rehired

CEO: too much focus on efficiency and cost [29]

Late 2025

Hybrid: AI for volume, people for complexity

Still reports USD 60m saved [30]

INFOGRAPHIC · WHAT THE SUCCESSES SHARE

Three conditions, every time



01

A human team frames the idea

Heinz asked what ketchup looks like before generating anything

→



02

Technology chosen to serve it

Cadbury designed the idea around what synthesis could do

→



03

Taste decides what ships

Nutella's designers set the rules the algorithm combined

In both reversals, at least one of the three was missing.



SECTION

05

How artists build AI differently

Æ designs AI systems, agents and workflows so that creative judgement is part of the architecture, not a final check. The principles below draw on the research in Section 3 and the cases in Section 4.

Draft design principles

PRINCIPLE	IN PRACTICE	EVIDENCE IT RESPONDS TO
Frame before you generate	Workflows begin with problem definition and a point of view, not a prompt	Heinz and Cadbury succeeded on the idea, not the tool [18][20]; value depends on how collaboration is structured [12]
Design for divergence	Systems deliberately widen the range of options rather than converging on the most likely one	AI assistance narrows collective novelty [8]
Build in critique	Agents surface weaknesses and alternatives, prompting review rather than acceptance	Over-confidence in AI reduces critical thinking [11]; Google's ad passed testing but failed judgement [27]
Know the frontier	Each workflow maps which tasks AI does well and where human judgement takes over	Performance falls on tasks outside AI capability [9]; Klarna reinstated humans for complex work [29]
Protect the brand's distinctiveness	Outputs are tuned to a brand's own voice and visual language, not model defaults	Distinctiveness drives disproportionate returns [3]; Nutella and Coca-Cola worked because humans owned the assets [22][23]



SECTION

06

Value, risk and brand safety

The business case rests on three sources of value and three risks that artist-led AI is designed to reduce.

WHERE THE VALUE COMES FROM

Productivity on the right tasks. AI raises speed and output substantially on tasks within its capabilities [9], and can let one skilled person match the output of a small team [10]. Coca-Cola's 2025 holiday film was made by around 100 contributors working from generated footage rather than a full live-action production [25]. The gain is real but conditional on task selection.

Creative quality, which drives sales. Because creative is the largest single driver of advertising sales effect [1][2], even modest improvements in creative quality can outweigh savings from cheaper production. Cadbury's AI-enabled campaign delivered a 35% sales uplift [20]; the technology was cheap relative to that result.

Distinctiveness, which the market rewards disproportionately. The returns to standing out are concentrated at the top [3]. Tools that preserve a brand's distinctiveness protect the most valuable part of its creative output, as Nutella's designer-constrained algorithm did [22].

THE RISKS ARTIST-LED AI REDUCES

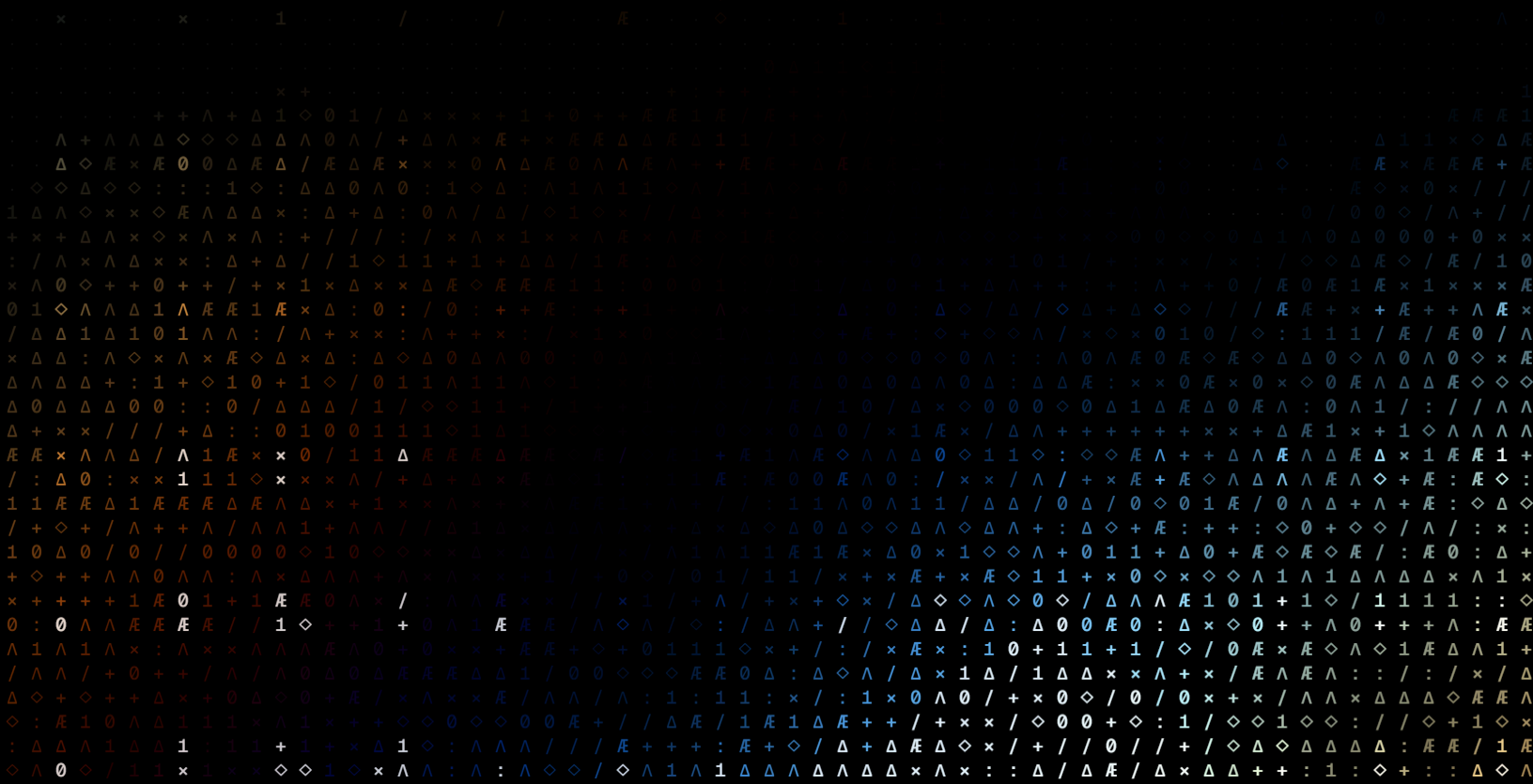
Brand dilution. Generic model outputs converge [8], eroding the distinctive assets brands spend years building.

Audience rejection. Consumers identify and respond negatively to visibly AI-generated ads [13] [14], and to ads that position AI as a replacement for human expression [27]. Poorly judged AI content can cost brand equity rather than save money.

Silent quality loss. On tasks outside AI's capabilities, confident but wrong output reduces performance [9]. Klarna's efficiency metrics improved while service quality fell, and the company only reversed course after customers noticed [29]. Workflows that build in expert review catch these losses before they reach clients.

Training data and consent

Buyers and investors increasingly ask how AI tools were built and whose work shaped them. Questions about training data, consent and artist compensation remain unresolved legally and reputationally. An AI company built by artists is well placed to answer them credibly, and to turn a common objection into a point of difference.



INFOGRAPHIC · THE BUSINESS CASE

Three sources of value, three risks reduced



VALUE

+40%

Productivity. Higher quality on tasks inside AI's frontier [9]

1 ≈ 2

One person with AI matched a two-person team without it [10]

+35%

Creative quality. Cadbury sales from an AI-served idea [20]

+56 pp

Distinctiveness. Shareholder return growth, top-quartile design [3]



RISK

Converge

Brand dilution. Generic outputs narrow towards the same centre [8]

Reject

Audience rejection. Visible AI ads draw negative sentiment and weaker memory [13][14]

−19 pts

Silent quality loss. Correct answers drop outside the frontier [9]



SECTION

07

Objections answered

A credible case has to meet the strongest counter-arguments directly. These are the ones buyers and investors are most likely to raise.

“Coca-Cola's AI ads scored top marks. Consumers clearly don't care.”

The scores are real [23]. But respondents did not know AI was involved, and System1 said disclosure would likely have lowered the result [24]. More importantly, what they were scoring was a distinctive asset built by human creative teams in 1995 and reproduced faithfully by around 100 people curating 70,000 clips [25]. The case shows that AI can execute a strong existing idea well under close direction. It does not show that AI can originate one, and System1's own analysts warn that AI-led novelty decays quickly [26].

“Research shows AI helps novices most, so why pay for experts?”

This is well supported. AI gave the largest creative gains to less creative writers [8], and one pair of experiments in fiction and songwriting found the smallest gains for domain experts, though reviewers have questioned aspects of its design [17]. A study of professional artists found their advantage over laypeople using AI tools was real but modest [16].

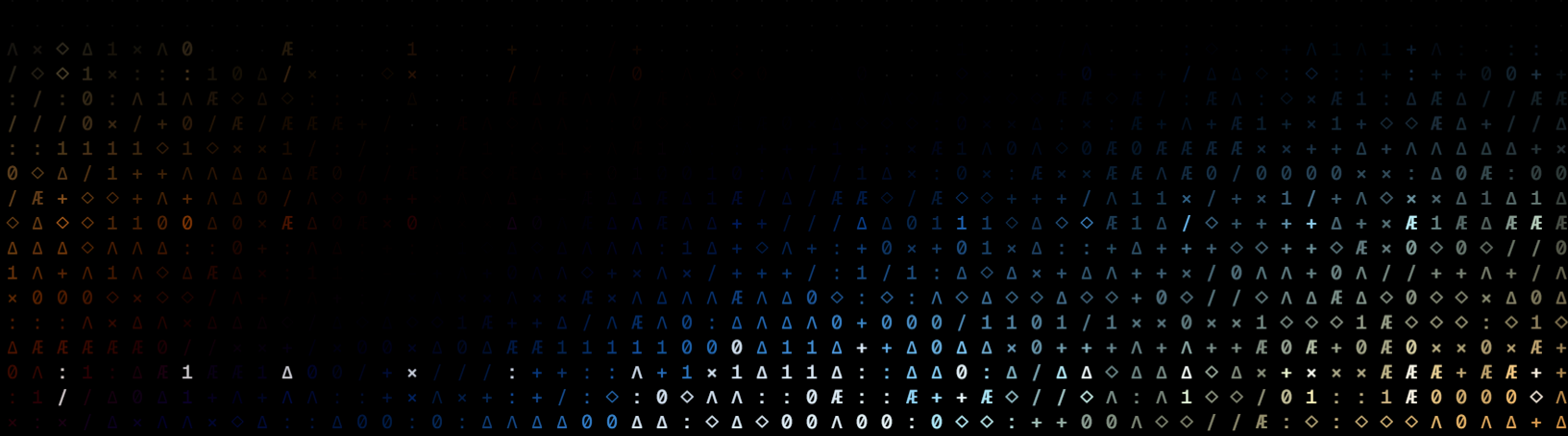
This strengthens the case rather than weakening it. If AI lifts everyone towards the same level, that level becomes the new minimum and stops differentiating. Competitive advantage moves to what lies above it: framing, originality and discernment. Those are the capacities Æ builds into its systems, and they are what separated Heinz and Cadbury from the campaigns nobody remembers.

“Why not just use off-the-shelf AI tools?”

Off-the-shelf tools are built for the average user and optimised for the most probable output. That is exactly what produces convergence [8]. Their value is real on suitable tasks [9], but they carry no knowledge of where a particular brand's frontier lies or what would make its work distinctive. The WFA found that two-thirds of major advertisers are still struggling to find new ways of working with AI [7]; the tool alone has not solved that.

“Isn't this just a correlation between good design and good companies?”

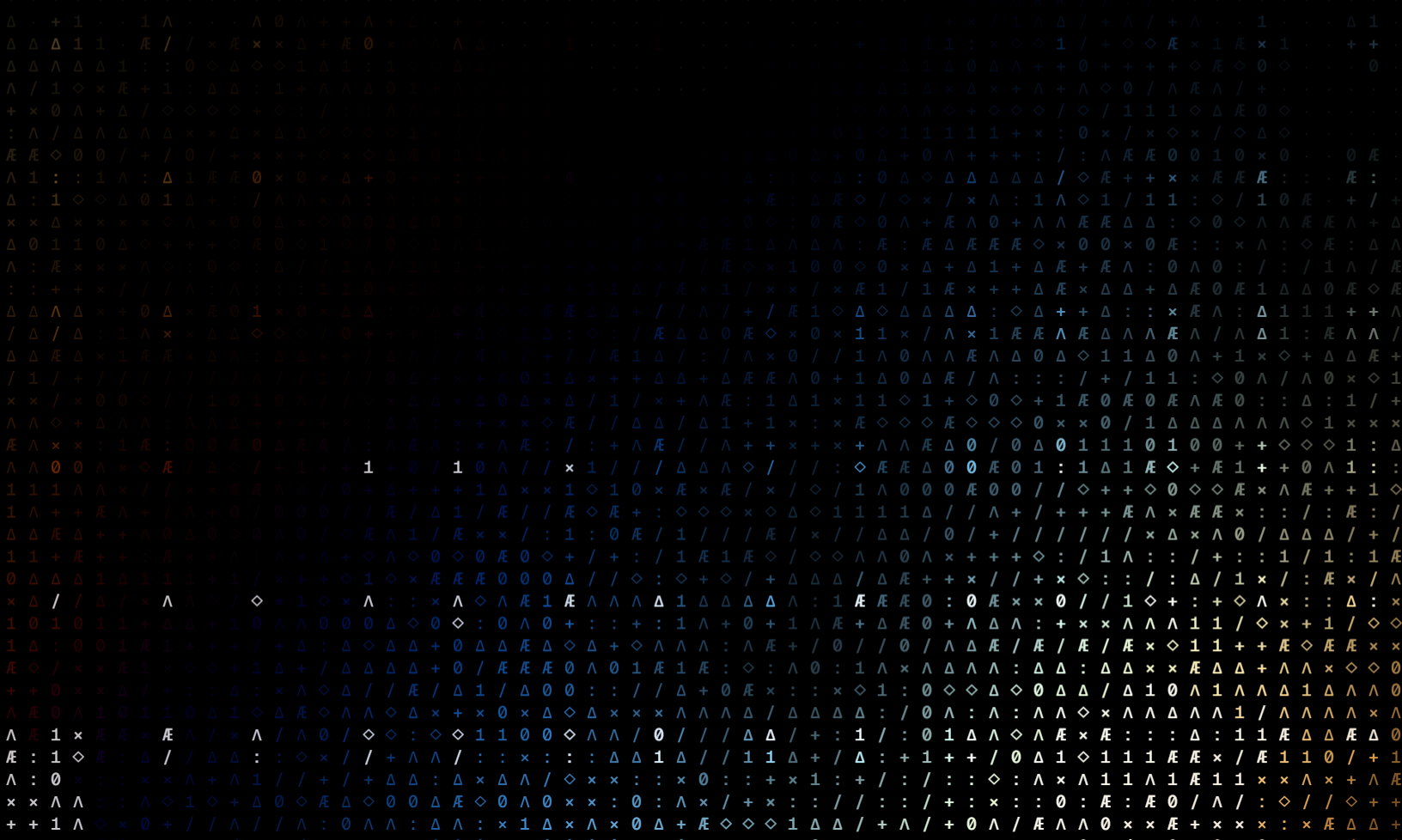
Partly, yes. The McKinsey design research is correlational [3]. But the experimental evidence in this paper, from the BCG, Procter & Gamble and Science Advances studies [8][9][10], is causal. Together they show both that distinctiveness is rewarded and that the way AI is used changes the output.



“AI is improving so fast that human judgement won't matter soon.”

The frontier moves, but it remains uneven [9]. More capable models still default to the most probable output, so the homogenising pressure does not disappear as quality rises. Google's ad failed on judgement, not production quality [27], and Klarna's reversal came after its AI had improved, not before [29].

The judgement about what is worth making, and for whom, remains a human responsibility.



SECTION

08

What this means for you

MARKETING LEADERS

Creative quality is your largest lever on advertising sales effect [1], and generic AI output puts it at risk [8][13]. Only three in ten of your peers say their teams consistently deliver creative excellence [6]. The question to ask of any AI investment is whether it makes your brand more distinctive, as it did for Heinz and Cadbury, or merely faster.

CREATIVE AGENCIES

When clients can generate competent work themselves, agencies compete on what AI cannot supply alone: framing, taste and critical judgement. The most awarded AI-enabled work of the past five years came from agencies that designed the idea first and chose the technology second [18][20][22]. Artist-designed tools let agencies scale that judgement rather than dilute it.

PROCUREMENT AND OPERATIONS

Productivity gains from AI are real but conditional on the task [9]. Klarna's experience shows that speed and cost metrics can improve while quality falls [29]. Evaluate AI suppliers on measured outcomes such as revision rounds, time to approval and audience response, not on generation speed alone, and ask how each tool was built and trained.

INVESTORS

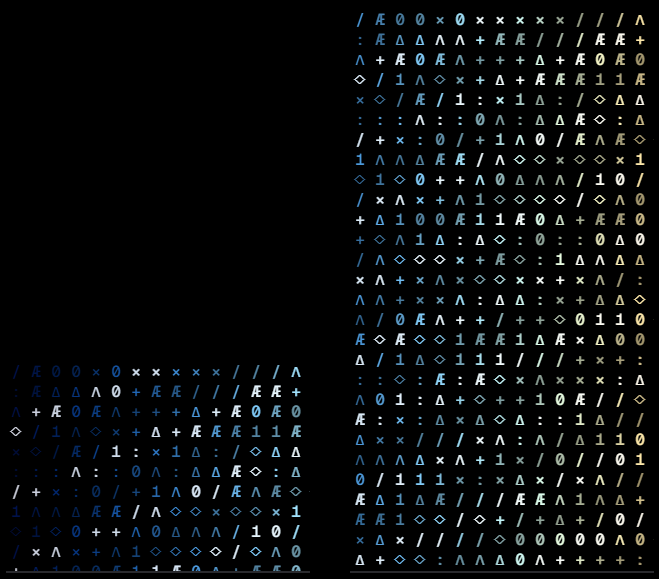
Generative AI in marketing is forecast to grow at close to 30% a year to 2030 [32], though estimates vary widely by definition. As foundation models commoditise, defensible value moves to the layer that shapes their output. Two-thirds of major advertisers say they have not yet found a way of working with AI that improves their creative [6][7]; that gap is the market. Creative judgement embedded in product design is difficult to replicate.



INFOGRAPHIC · FOR INVESTORS

A growing market with a judgement gap

GENERATIVE AI IN MARKETING, USD



6.6bn

2026

~18bn

2030

Close to 30% a year

FORECASTS VARY BY DEFINITION · [32]

MAJOR ADVERTISERS STILL WITHOUT A BETTER WAY OF WORKING



~2 in 3

have not found a way of working with AI that improves their creative. That gap is the market [6] [7].

SECTION 09

Conclusion

The evidence points one way. Creative quality still drives commercial results [1][2][4], the market still rewards the distinctive disproportionately [3], and audiences still notice when work lacks deliberate direction [13][27]. AI has changed how fast competent work can be made. It has not changed what makes work worth making.

The brands that have won with AI so far, Heinz, Cadbury and Nutella among them, did not ask the technology for ideas. They brought ideas to it, and put people with taste in charge of what came out. The brands that stumbled, and the company that rehired the people it had replaced, learned the same lesson in reverse.



The organisations that gain most from AI will not be those that generate the most. They will be those whose AI is shaped by people who know what is worth making. That is the premise on which Æ is built.



Notes on sources and method

This paper prioritises peer-reviewed research and pre-registered field experiments, supported by large industry meta-analyses and publicly documented campaign results. Every quantitative claim is referenced below.

Commercial interests. McKinsey sells design services; Nielsen, NCSolutions, NIQ and System1 sell measurement; the IPA, WARC and LIONS represent the advertising industry. Their findings are widely cited but are not independent of those interests.

Campaign results. Figures for Heinz, Cadbury and Nutella come from agency, awards and trade-press sources and have not been independently audited. Heinz's earned impressions are reported as 850 million in the original entry and 1.15 billion in later summaries; both are cited.

Correlation and causation. The design and sales-driver studies [1][2][3][4][5] are correlational. The experiments [8][9][10] and the survey [11] are the basis for causal claims. The six cases in Section 4 are illustrative, not experimental.

Market forecasts [32] vary widely between research firms depending on how the market is defined. Treat them as indicative. The IAB adoption figures [31] are cited via secondary reporting and should be confirmed against the original.

Pre-publication check. Figures are taken from abstracts, publisher pages and reputable summaries. Before release, verify each against the full text and confirm the latest version of working papers.

Homogenisation chart. The Science Advances study [8] is described in prose rather than charted, because its similarity measures need to be taken from the full paper's figures.



References

1 Nielsen (2017). When it comes to advertising effectiveness, what is key?

2 NCSolutions (2022). The Five Keys to Advertising Effectiveness, updated meta-study. Summarised by MarketingCharts.

3 Sheppard, B., Kouyoumjian, G., Sarrazin, H. and Dore, F. (2018). The business value of design. McKinsey Quarterly.

4 Field, P. (2019). The Crisis in Creative Effectiveness. IPA.

5 Field, P. Creativity and effectiveness. WARC, on the IPA/Gunn Report analysis.

6 WFA and LIONS (2026). Clients and Creativity 2026.

7 MediaPost (2026). AI Creative Excellence: All Talk, Little Action, WFA Finds.

8 Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28).

9 Dell'Acqua, F., McFowland III, E., Mollick, E. et al. (2023). Navigating the Jagged Technological Frontier. Harvard Business School Working Paper 24-013.

10 Dell'Acqua, F., Ayoubi, C., Lifshitz, H. et al. (2026). The Cybernetic Teammate. Organization Science.

11 Lee, H.-P., Sarkar, A., Tankelevitch, L. et al. (2025). The Impact of Generative AI on Critical Thinking. Proceedings of CHI 2025.

12 Randazzo, S., Lifshitz, H., Kellogg, K. C. et al. Cyborgs, Centaurs and Self-Automators. HBS Working Paper 26-036. Summarised by Lakhani.

13 NielsenIQ (2024). NIQ Research Uncovers Hidden Consumer Attitudes Toward AI-Generated Ads.

14 Marketing Dive (2024). Consumers call AI-generated video ads annoying, confusing, per NIQ.

15 (2025). Exploring creativity in human–AI co-creation: a comparative study across design experience. Frontiers in Computer Science.

16 (2025). Expertise elevates AI usage: experimental evidence comparing laypeople and professional artists. arXiv 2501.12374.

17 Huang, M., Jin, M. and Li, N. (2024). Augmenting Minds or Automating Skills. arXiv 2412.03963.

18 Rethink for Heinz (2022). A.I. Ketchup, campaign entry. Ads of the World.

19 Zenodo (2025). Heinz AI Ketchup campaign summary.

20 WARC. Cadbury Celebrations: Shah Rukh Khan-My-Ad, case study.

21 Contagious (2023). Cannes Lions: Creative Effectiveness winners 2023.

22 shots (2017). Ogilvy Italy's 7 Million Unique Nutella Jars.

23 Adweek (2024). How Coca-Cola's AI Holiday Ad Went From Praise To Rage.

24 MediaPost (2024). Not Everyone Hated That AI-Generated Coca-Cola Holiday Ad.

25 System1 (2025). AI or No AI, Coke Gets the Christmas Love.

26 MediaPost (2025). What Marketers Can Learn From Backlash To Coke AI Ads.

27 The Hollywood Reporter (2024). Google Pulls “Dear Sydney” Gemini AI Ad From Olympics Coverage.

28 Fortune (2024). Google Olympics advertisement Gemini AI reactions.

29 Fortune (2025). Klarna plans to hire humans again.

30 CX Dive (2025). Klarna says its AI agent is doing the work of 853 employees.

31 Implicator (2025). Coca-Cola debuts AI holiday ads, citing IAB and Attest figures. Secondary source; verify against originals.

32 Research and Markets (2026). Generative AI in Marketing Market Report.