



Augmented Intelligence in Creative testing - AI as co-pilot for human values and decision making



Andy Santegoeds
- Senior Solution Consultant MetrixLab, Effie Juror Netherlands and Lector at Bcekestijn Business School



Björn Heijligers
- Head of Data Science MetrixLab, Alumnus Cosmology - Sterrewacht Leiden



Claudia Gelbe - Managing Director MetrixLab Germany, Effie Juror Germany and Lector for Market Intelligence at DHBW

Judging creative work - like advertising - is often considered as **'noisy' decision making**. Experienced professionals tend to have high confidence in the accuracy of their own judgments, and they also have high regard for their colleagues' intelligence. Kahneman et al. (2016) come with directions for combating this 'noise': The application of rules or algorithms.

Algorithms, often "upgraded" to "Artificial Intelligence" in marketing texts, have long entered our daily lives. Technologies like NLP and computer vision algorithms are now deployed in advertising testing too. AI seemingly promises a world where all actions and decisions can be made fully automated. In our opinion algorithms are not sophisticated yet to test ads

and place them in a real-world media environment - without human interference. We believe in Augmented Intelligence, where human actionability excels, grounded in data and statistics, backed with machine learning and guided by human values.

Keywords:
Artificial Intelligence. Augmented Intelligence. Prediction. Creativity. Marketing effectiveness. Marketing impact. ACT Instant. MetrixLab. Sense making.

A version of this article was originally published in German in Transfer - Zeitschrift für Kommunikation und Markenmanagement (<https://transfer-zeitschrift.net>).

Abstract

1 Introduction

Advertising is all around us. Whether it's on your TV-screen, watching a football match or movie; on your smartphone when checking the latest news or weather forecast, at roadside billboards on your way to the supermarket or the gym, through the sound system of your car... And with the latest technology more and more data are collected on what people are doing and where

they are, facilitating targeting of ads and messages by predicting their potential next move.

This proliferation of advertising across media, formats and devices offers a plethora of opportunities for marketers for bringing the brand under the attention of consumers. At the same time marketers are now creating more ads for more media than ever before.

Laurens Vreekamp (2022) introduces a **"New Synthetic Time"** in his book The Art of AI. He foresees synthetic media (like the Metaverse), as the next phase in media evolution, from analogue and electronic to the phase which we are about to leave: digital media. Synthetic creators and perhaps synthetic walled gardens bring us on the verge of even more data on ever more

advertising formats.

Already in 1999 tech futurist Kevin Kelly expressed his concerns: **“The only factor becoming scarce in a world of abundance is human attention.”** This fight for eyeballs puts pressure on generating strong, creative ads. It also puts pressure on marketers, since

judging creative work isn't an easy task.

Judging creative work - like advertising - is often considered as **‘noisy’ decision making**. Experienced professionals tend to have high confidence in the accuracy of their own judgments, and they also have high regard for

their colleagues' intelligence. This combination inevitably leads to an overestimation of agreement.” Kahneman also comes with directions for combatting noise in decision making: The application of rules or algorithms.

2. The hype of AI

2.1 The application of rules or algorithms as popular synonym of Artificial Intelligence

Algorithms, often “upgraded” to “Artificial Intelligence” in marketing texts, have long entered our daily life, for many consumers without even being conscious of it (Urbach & Hofmann, 2022). The popular

usage of the word Artificial Intelligence refers to anything that uses a computer algorithm to ‘automagically’ create a predictive outcome.

Such automation, as using speech recognition to enable a computer to place a call, to using facial recognition to help identify missing children, to using

computer vision to enable the use of autonomous driving cars are all around us. The list is endless. The same technologies as speech recognition, natural language processing, facial recognition and computer vision algorithms are more and more deployed in market research and in particular in advertising testing too.

2.2 Deploying algorithms and machine learning in creative testing

Many algorithms are performing spotless tasks (think of converting Speech to text or counting words or scenes in an ad, or the time a logo is visible). At the same time, for essential understanding of the impact of advertising not all features can be captured by algorithms (yet) and human intervention is needed.

Imagine this ad, where you see two people running on a boulevard, both dressed in runners outfit, talking to each other, except for one runner the head is replaced by a naked foot. Obviously, the algorithm is not trained with a dataset that contains a ‘foot-head’. The machine learning prediction of attention differs from the eye tracking in a lab situation (EEG) and falsely underestimates the potential of the ad. This could lead to rejection of a potentially strong ad.

Reliable coding of advertising material is not the only challenge for an AI system. In order to translate insights into actions, an AI system requires the definition of a so called “Objective Function”. An algorithm that predefines the goal the AI system should achieve in terms of unambiguous and measurable quantities that are combined to create a final numerical value.

This means that in developing a traditional AI system, the creators need to anticipate any possible solution the AI system might find and tell ahead of time whether that action is desirable or not.

In real world situations it is almost impossible to define this Objective Function ahead of time in a way that prevents all possible unintended side effects or strategies. This challenge is called the **“Alignment problem of AI”**. Brian Christian writes “The problem, of course, with a system that can, in theory, learn just about anything from

a set of examples is that it finds itself, then, at the mercy of the examples from which it's taught”

For advertising testing the Objective Function is as big a challenge as the deployment of algorithms. Many algorithms are performing spotless tasks (think of converting Speech to text or counting words or scenes), but for essential understanding of the impact of advertising not all features can be captured by algorithms (yet) and human intervention is needed. No algorithm can judge rightly how many messages (explicit or implicit) are embedded in a story, which is genuinely considered to be a decisive factor for driving impact.

AI seemingly promises a world where all actions and decisions could be made fully automated. We argue that this vision will continue to be tantalizingly close, yet just out of reach for some time to come.

2.3 AI-supported decision-making and meaning-finding

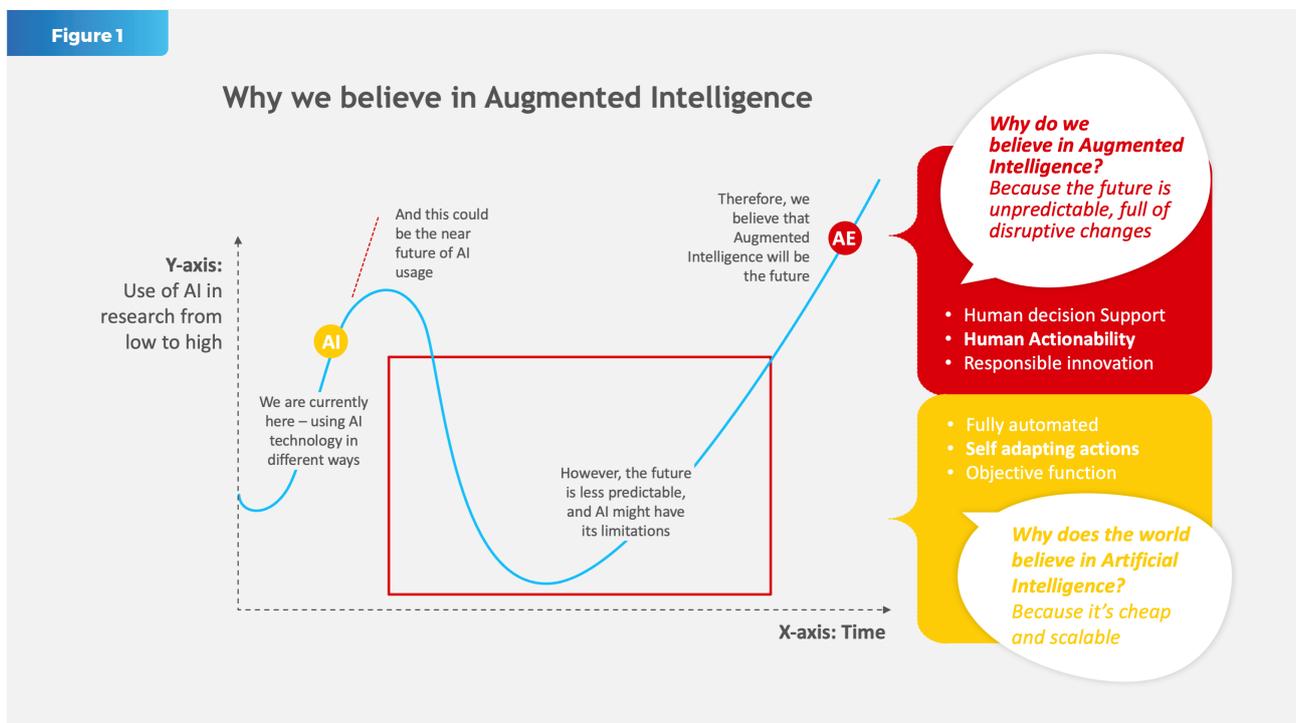
In the future, companies that design AI systems with the human in the loop will be at a strategic advantage in dealing with radical changes to the domain of marketing resulting from e.g. constant development of new advertising formats such as TikToks, the Metaverse and NFTs to disruptive new technologies

to the evolution and to the core values of our world economy. So, it's about developing and deploying an expert system where AI and humans work together. **Figure 1** shows this process.

We believe in a situation in which human actionability excels, grounded in data and statistics, backed with machine learning and guided by human values. Augmented Intelligence provides

therefore a way forward that is anti-fragile, value driven and future proof.

Stan Sthanunathan, Executive Vice-President of Unilever's CMI organization put this combination into a simple formula: **"Augmented Intelligence equals Artificial Intelligence plus Human Intelligence"**.



3 Deploying algorithms and machine learning in creative testing

This is possible by creating an opportunity to move away from fully automated black box AI systems to a hybrid form, where AI proactively and creatively prepares both insights and options for the human decision maker. We apply this **Augmented Intelligence framework** also for testing advertising.

The ultimate goal was to develop an ad pre-testing system that

could predict ad performance based on machine learning algorithms and predictive modelling – without any survey, so no respondents were involved. This hybrid approach – we call it ACT Instant – offers marketers speed and granularity from algorithms and contextual judgement by trained humans. When pretesting, more datapoints can be presented faster, without comprising the soul of creativity

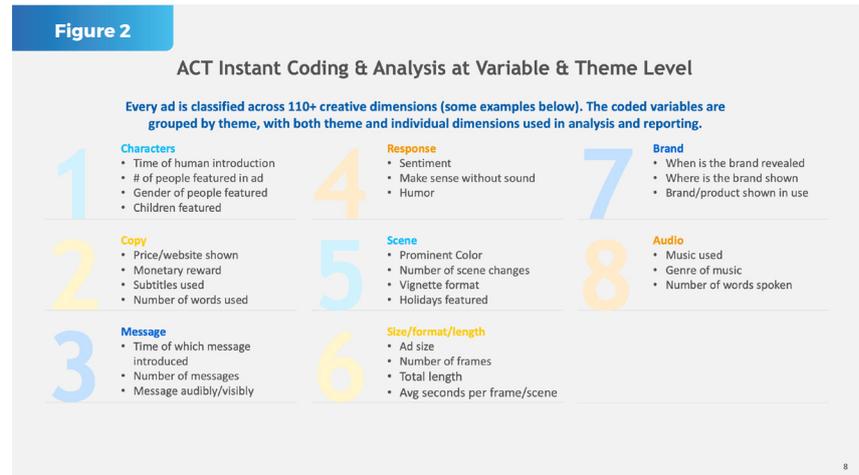
and adaptable to an everchanging media landscape.

The whole process can be divided in three phases. (1) It starts with coding of an on 110+ creative elements, (2) next these datapoints are inputted into a predictive model, (3) finally a human advertising expert interprets the results and writes recommendations for the client.

3.1 Coding of an ad

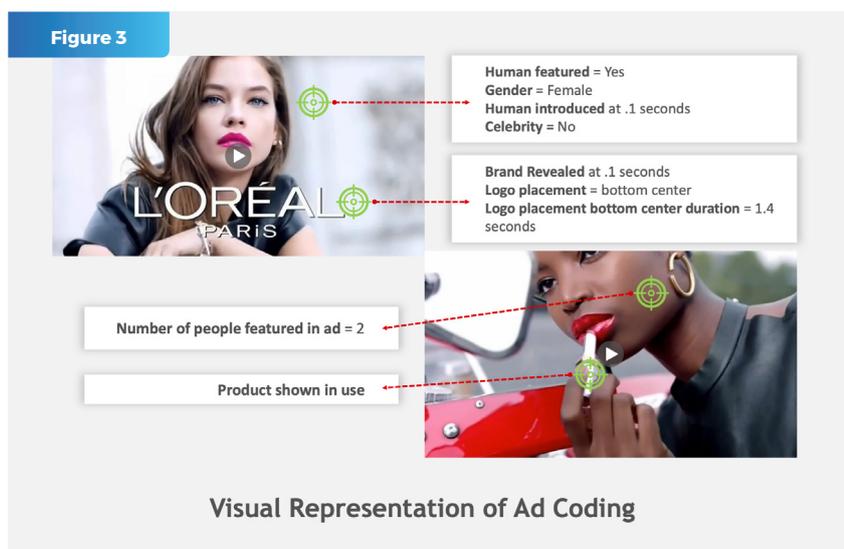
A combination of AI algorithms and human advertising specialists code ads across 110+ creative elements. Based on our vast experience in copy testing and the latest findings in advertising research, we have selected the elements that need to be coded to produce a prediction of the potential power of an ad. A script runs a series of predefined algorithms to collect datapoints on eg. identifying when the brand is introduced, whether the ad features people, music played etc. Algorithms produce these datapoints flawlessly and faster than humans. In building this solution it became clear that human touch is needed for some things computers can't yet accurately code. **70% of the coding of ads is AI-driven via automated algorithms, but 30% is human.** Seasoned researchers are trained to code a list of selected creative elements to ensure accuracy for variables with more subjectivity (e.g. humor, emotion, number of messages, brand integrated in the story etc).

For example, it is crucial for the success of an advertisement when the main message is conveyed.



Since this is not always done via language or supertext, but also via the story itself, AI quickly reaches its limits. Imagine an OTC (Over the counter-medication) commercial for a headache medicine and the so-called 'relief-moment' (when a smile spreads across the protagonist's face) and you, as a human, immediately know that no subtitle is needed at this point to show that the headache is gone immediately. A machine (or AI) must first learn this, which of course can be represented differently depending on the creation. The assessment of the conveyance of the message can therefore lead to false statements based purely on AI. Another example is

recognizing product usage, which of course can differ from other situations in the spot. Does driving a car represent the product or is it "just" a by-product in a vignette film? Does the spot still make sense without sound? And should one classify a celebrity in a local advertising campaign? For some of these questions we tried to use algorithms like celebrity status, by looking at the number of followers on social platforms, but for many variables it has not yet been possible to achieve a sufficiently high level of accuracy. **Figure 2** shows an overview of the coding process, in **Figure 3** the ad coding is visualized using an example of a lipstick advertisement from L'Oréal.



We developed a proprietary approach for coding that utilizes existing APIs combined with customized processing in order to maximize accuracy. The advantages of using APIs are: Fast processing times, cost savings, reliability and ease of use. The entire automated coding is thus checked and, if necessary, adjusted by an analyst before the first process step is completed.

In the longer term, we will continue to look for ways to automate more variables. Over time, we expect that with more data points and machine learning, more creative elements of automation can be added. We don't see manual coding going away entirely, at least not any time soon.

3.2 Predictive modelling

Once coding is completed, we have more than 300 datapoints for each test ad. All coded variables are fed into a database and the predictive model is run. Machine learning PROC Adaptive Regression Modelling is used to develop performance results

3.3 Analysis and interpretation of the data

The results from the predictive model and the diagnostics are then put into a scorecard report and analyzed by one of our advertising analysts. A key result is the Overall Performance Index of an advertising asset, which indicates the strength of the creation on a scale from 0 to 100.

The algorithm classifies the potential of the ad on the basis of the ads that have already been

3.4 Validation

ACT Instant models were based on our existing validated solutions for copy testing, which takes into account Breakthrough and Response from standard, validated survey-based approaches. What we did here is take N=90 recent ads, globally across TV, Social Video and Online Video, tested using our survey-based pre-testing approaches ACT Copy. We also ran the same ads via ACT Instant and compared the overall performance values across the methods. See **Figure 4**. Results for individual formats (TV vs Social Video or Online Video) and across regions are all very much alike. Specifically, one sees e.g. that all survey-based tested commercials in the top 40 percent area are also classified there based on AI-based predictions (38.7 percent vs. 39.0 percent). The overall prediction within ACT Instant aligns with standard approaches.

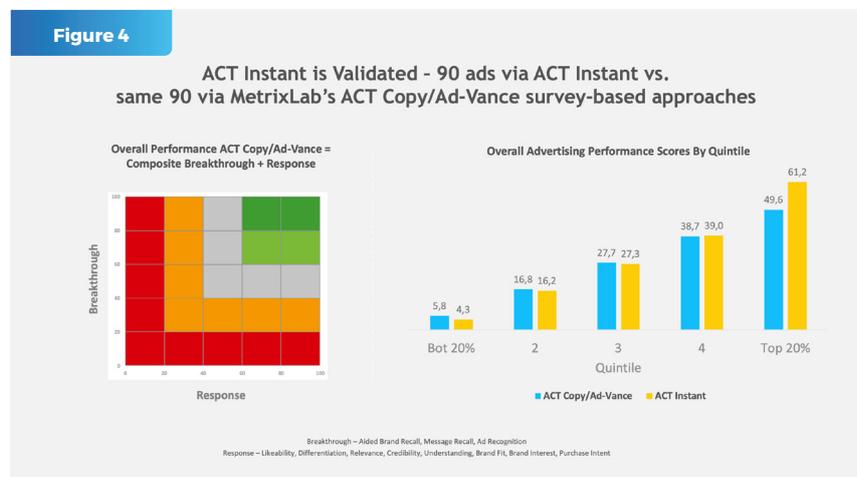
Since its launch in 2019 we have tested thousands of ads

and predictive scoring. We chose this method to account for the non-linearity and interaction of data. The models were built using MetrixLab's database results of 16.000+ ads from survey-based pre-testing studies in order to establish ad performance baselines relevant to format and

analyzed. The Index therefore corresponds to a quintile value, where results are grouped in five, equally sized sets. TV commercials (or other formats like Online Videos) with a performance in the top 40 percent of our database for the two core parameters of Breakthrough and Response are considered strong performers and thus offer the potential for a disproportionately good advertising effect (see matrix in **Figure 4**).

market. In general, this is done to predict impact which features metrics like brand and message recall and ad recognition as well as metrics like sympathy, differentiation, relevance, credibility, understandability, brand fit, brand interest, and purchase intent.

Then we go one level further to identify specific variables that are influencing overall performance. We also provide a comparison against best practices which acts as a checklist to identify for every test ad, how many of the Best Practices for that given media are being met. This is not an automated data dump, this is a fully analyzed report by an experienced advertising analyst supported by an evolving and self-learning AI - Data Butler.



via ACT Instant. ACT Instant can be a very valuable option in a number of different situations such as when speed is of critical importance, for an immediate go-no-go decision. **However, there are some situations where we would not recommend an AI-based advertising test and would instead recommend a different solution within our suite.** For example, situations where the ads being evaluated are very similar to each other with only minimal differences. The AI-

based algorithms will most likely produce similar datapoints, and the model will predict a similar performance. Also situations where you want to get an in-depth understanding of how an ad performs across two different target or consumer groups, or situations where there is a need to understand a qualitative audience reaction to things such as humor, or message or scene relevance? These are all examples where a consumer-based survey approach is more appropriate.

4 Conclusion and outlook

In an ever-changing world, marketing organizations can up their game with Augmented Decision Making, where algorithms support human decision making with speed and accuracy. A wide range of algorithms has proven to be a valuable addition to the toolbox of market research and adds valuable new opportunities for copy testing TV commercials and other creative assets. Many algorithms are

fast, flawless, easy to use, attractively priced, and accessible. Developments are going fast, offering new opportunities like predicting sentiment of the music score of an ad. At the same time, **human advertising experts are - still - indispensable for judging critical creative elements that are more subjective,** like humor and messaging.

Management -Takeaway

The hybrid deployment of AI-supported and human coding, the modelling as well as the AI-supported analysis and the hybrid interpretation of the results are well suited for pretesting advertising assets.



Literature

Christian, B. (2020). The Alignment Problem: Machine Learning and Human Values. New York: Norton & Company.

Heath, C., & Heath, D. (2006). The Curse of Knowledge. Harvard Business Review (Dezember 2006). Abruf von <https://hbr.org/2006/12/the-curse-of-knowledge>.

Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Learning, how to Overcome the High, Hidden Cost of Inconsistent Decision Making. Harvard Business Review, 94(10), 38-49. Abruf von

<https://rapaport.com/wpcontent/uploads/2020/09/20171001-AR-HBR-Noise.pdf>.

Kelly, K. (1999). New Rules for the New Economy: 10 Radical Strategies for a Connected World. London: Penguin Books.

Sthanunathan, S., & Friedman, L. (11.12.2018). A Conversation with Stan Sthanunathan of Unilever on His New Book. Abruf von <https://www.greenbook.org/mr/market-research-leaders/a-conversation-with-stan-sthanunathan-of-unilever-on-his-new-book/>.

Urbach, N., Hofmann, P. (2022). Warum wir nicht nur von Künstlicher Intelligenz sprechen sollten. transfer-Zeitschrift für Kommunikation und Markenmanagement, 68(3), 10-13.

Vreekamp, L. (2022). The Art of AI. Een praktische introductie in machine learning voor mediamakers. Culemborg, NL: VANDUUREN MEDIA.