

Beyond the Blocklist:

The Future of AI-Driven Brand Safety

Introduction

Since 2015, our SmartMatch™ brand safety framework has been powered by AI, now leveraging our LLM capabilities alongside continuous human review to maintain precision and relevance. SmartMatch™ delivers strong brand safety protection, while allowing clients additional controls like keyword lists. Keywords, however, are blunt tools, often excluding content without accounting for nuance or intent. By combining SmartMatch™ with publisher curation, we give brands an unprecedented degree of safety and customisation aligned with their individual risk tolerance.

This backdrop sets the stage for discussing how agentic AI could reshape brand safety. Traditional methods have been rigid, whereas emerging AI capabilities promise a more sophisticated understanding of context, emotion, and nuance. The interview explores whether agentic AI can truly interpret sentiment and make informed judgements at scale, where the line between automation and human oversight lies, and what barriers stand in the way of adoption.

Contributors



Patrick Hann has worked in the ad tech industry for nearly a decade, working as a programmatic lead, head of customer success, and during that time has been the lead buyer for two DSPs and worked at companies specialising in cookieless solutions such as contextual targeting and attention measurement. He is currently Ad Tech Manager at IAB UK.



David McMurtrie has had a long and varied career in media and technology, working at the forefront of some pivotal changes in the industry. He spent seventeen years at Google UK, where his role focused on driving digital transformation with major news publishers and rolling out Google's suite of publisher products. David has been on the advisory board of Smartology since 2024.



Marcus Keane possesses 25 years of experience in the publishing industry, with key roles at both The Financial Times and Euromoney. At the FT, he was responsible for the global engineering function and pioneered their subscription paywall. As Chief Product and Technology Officer at Smartology, his team redefined digital advertising, creating the SmartMatch Demand Side Platform, which utilises AI and Large Language Models for innovative, privacy-focused contextual targeting.

“AI offers a solution by evaluating the context of a word within an article or news headline, enabling the full monetisation of content.”



In what ways do you think agentic AI could change how we approach brand safety, particularly when it comes to understanding context, emotion, or nuance?

Patrick: I think with brand safety we’re moving away from keyword lists, which have been the way people have done it for a while, and now you have certain technologies that have quite a bit more nuance than that.

The AI layer, agentic or not, can add a kind of sentiment that wasn't there before. It can understand the full emotional tone of an article. We talk about semantics having that extra layer of interpreting a word, and this is the next stage to that, where you can actually understand the full emotional weight of something.

All of these things up to this point have been based on quite rigid models, and I think now with agentic AI, or even AI in general, you've got this increased ability to interpret and to make judgement calls. With all these things, because although there are a lot of people making quite big claims, the proof will be in the pudding. The user layer is still quite far away. There’s models being demoed, but in terms of public availability, it’s still not very accessible.

David: During Covid-19 many agencies implemented a blanket policy of blocking a very long list of keywords considered sensitive. Although the industry, led by the IAB, made significant strides in removing many of these blocks, a considerable number remain in place. This continues to prevent a substantial amount of revenue from reaching publishers. Ozone, a premium publisher platform, estimates that up to 40% of its ad requests are blocked.

The primary issue with such a blunt approach is the failure to consider the context in which the word is used. AI offers a solution by evaluating the context of a word within an article or news headline, enabling the full monetisation of content. In the long term, more nuanced distinctions will become possible. AI is already capable of writing in preferred styles suggesting the potential to match the tone of an advertisement or sponsored content with the editorial environment.

Marcus: Contextual nuance has been trained into SmartMatch™ brand safety AI models since 2016, long before the agentic AI term was popularised. SmartMatch™ brand safety is trained for 24 classifiers that define standard categories of content that meet the demands of even our most cautious clients.

Of course SmartMatch™ can apply a client’s own custom keyword exclusion lists, but these lists rarely reflect the advances that have been made in AI in recent years. Keyword lists tend to be extensive, redundant, and reflect a zero tolerance of advertisements appearing in unfavorable placements. Whilst keyword lists do still have a place in advertising, they are a blunt instrument with none of the contextual or emotional nuances of SmartMatch™ brand safety AI models, which means clients can miss placement opportunities, eliminating high quality inventory and increasing their CPM by competing for a smaller pool of competitive inventory. Personally, I’m very excited about the future of brand safety agentic AI in SmartMatch™, developing learning agents that improve over time and can more accurately reflect the risk profile of the individual advertiser, rather than today’s one size fits all approach.

Do you think we’re now approaching a level where AI will be able to handle the level of nuance that brands are looking for, or do you think that we’re still going to need human oversight?

Patrick: Until we reach the consensus that AI is superior in certain ways, or at least as good as us, human touch will be very important. Brand safety is probably the most important factor of campaign management where that human layer is needed.

From what I've gathered from talking to operational people on the subject of brand safety, there have also been attempts to automate it in the sense that, pre AI, people would use blocking tags, or various other methods to brand-safety-proof their campaigns. But actually, I think, everyone's in agreement that you do need to have human oversight in all cases. You need to have a plan for how you want to approach global events. There are different risk factors at play for different brands, and although you may be able to layer in some method of brand suitability scoring, you still need that human oversight where you’re deciding what is actually brand-safe and what isn’t. AI can think very quickly, but it doesn't necessarily know what you want. You still need to check your reports regularly, and you still need to discuss with your team what is suitable for your brand. So, a lot of the difficult parts of brand safety are not the bits that can be automated, and I think completely automated brand safety does have the potential to be dangerous.

Within the realms of something like brand safety, because it’s quite a human centric idea, I don't think AI necessarily improves on human ability. It will certainly speed up the processes and maybe bolster some of the brand safety levels, but I don't think it necessarily has a better idea of the nuances that make something brand unsafe.

David: Human oversight is going to continue to be important but I’m more optimistic than Patrick regarding AI’s future role.

Historically, human oversight has been responsible for broad keyword blocks and I believe AI can handle this more effectively. It’s a question of how risk averse we want to be. Blanket blocks have proven detrimental to the industry, often remaining in place for extended periods. Accepting a small amount of risk in exchange for AI’s enhanced contextual and tonal understanding would appear to be a favourable trade-off. Of course, this shift also suggests that the agency teams currently performing this work will likely see their roles automated.

Smartology's existing AI-driven content profiling will continue to advance and we’re working on adding additional levels of sophistication to this.

Marcus: Absolutely, and in time, I expect that manually maintained keyword lists will be wholly replaced by AI brand safety. But this transition will take some time due to fear of change, lack of adaptability or simply resource and time limitations. Smaller agencies and brands who are more nimble and more natural risk takers will inevitably lead the way on adoption, followed more slowly by the wider advertising community. Despite my optimism, there are real challenges in removing humans completely from the loop as tolerances vary. I remember one example where an advert appeared alongside an article about the impact of livestock on global warming. The article appeared on a respected business publishers website, but the eye-catching headline contained ‘burping cows’. Despite the content being a serious report on global warming and exactly the kind of content the advertiser wanted to reach, we were asked to block the content.

As you can see from this example, subjective nuances within individual advertiser risk appetite mean that, for the time being at least, I expect human oversight will continue to be a factor for many.

What are the main barriers to using agentic AI for nuanced brand suitability decisions today? And are they more technical, strategic or cultural?

Patrick: Good question. I suppose at the moment the barriers are cultural and strategic because culturally, although AI has a lot of information, it doesn't necessarily understand all of it in the context of how you want to deliver something. So, it kind of comes back to risk tolerance.

In terms of barriers to adoption, they’re more practical at the moment. Rather than the problem being that we don’t trust AI in this context, I think the problem is more that we haven’t had any exposure to practical examples of it.

David: While the technical capability exists and will become more readily available, it takes time to establish the necessary resources and business models to take full advantage of this.

The ad-tech industry is rapidly adapting to this challenge and learning how to best utilise the available technology. That said, large agency groups are often slow to adopt change, presenting significant strategic and cultural hurdles. The focus needs to shift from viewing this as a threat (to jobs, established practices, and revenue) to recognizing it as an opportunity (better jobs, more efficient practices, growing revenue).

Marcus: I think it’s a bit of each.

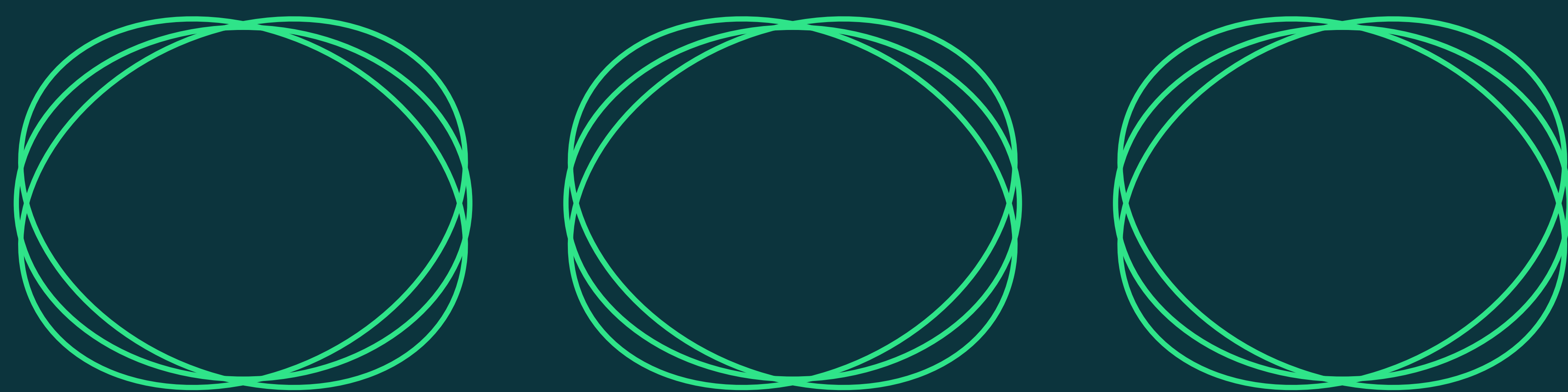
Technically speaking the advancements over the last few years and importantly the exposure to and adoption of AI by the individual, means we’re in a sweet spot of wonder at what AI can do and expectations of what will do in the future, which is great. But brand safety models especially need continuous retraining if we are to meet end user expectations consistently.

I see the advertising industry starting to deliver in terms of the growing number of platforms already leveraging AI for brand safety, yet campaigns still employ keyword blocklists.

This will variously reflect scepticism, an absence of strategic focus on AI and a lack of skilled resources to undertake the necessary due diligence, planning and testing of the capabilities of an exclusively AI approach. Workloads on the teams behind planning and delivering campaigns mean they simply don’t have the time necessary to undertake proper adoption.

From a strategic perspective, AI is on the radar of most leadership teams and speed of adoption will reflect the culture in the business and its ability to resource that adoption.

“Advancements over the last few years and importantly the exposure to and adoption of AI by the individual, means we’re in a sweet spot of wonder at what AI can do and expectations of what will do in the future”



Do you think that the growing belief that AI can understand context and brand nuance is moving faster than what is realistically possible, especially when in a high stake environment, like on a live campaign?

Patrick: My view is that the landscape is more sceptical of AI’s capabilities when it comes to context and brand nuance at the moment, and will be at least until it becomes more widely tested.

The use cases I’ve seen from agentic AI so far have been more related to campaign management. People are starting with really functional applications. Stuff that takes quite a lot of time to set up, and that’s more mindless, in a sense. Setting up line items, et cetera.

When it comes to something like brand safety, it’s much more thought out, so I think understanding nuance will be the next step after the functional elements.

David: It’s up to media and ad-tech companies working together to demonstrate what is possible and this can be done, initially in low risk environments. There is no reason why AI shouldn’t be applied to live campaigns to help make contextual decisions; in fact, this is already being done. We’re not quite at the point of real-time personalised advertising being applied at scale but it’s a major discussion point within the industry and there is no doubt that it’s on its way.

Marcus: Imagine an event similar to the Eyjafjallajökull eruption in 2010 happening now, causing significant travel disruptions, evacuations and economic impact. Within the context of advertising you would expect that brand safety training would prevent ads being served into stories covering this disaster. But what about travel articles written before the event that name the volcano for sight seeing? Brand safety should be expected to understand the context around volcanic eruptions and associated impact of the same. But at the immediate time of the event, it would be too soon to expect the name of the volcano to be included in brand safety training models. Even

so, with no other disaster related context in a destination focussed article, if an advertiser had a zero tolerance for the volcano, then a high sensitivity would be required for blocking. And fifteen years after the event would a blanket block on any mention of the volcano still be appropriate?

This is where a dialogue between advertiser and platform is crucial at managing expectations and understanding what and how quickly your partner can respond, by understanding your partner’s model retraining frequency and approach to live events.

Do you think there's a barrier to improving AI capabilities if we don't trust it enough to use it, and thus train it?

Patrick: There's a theory that after a certain point AI will pick up your bad habits, for example. So it can be a circular thing. If you keep feeding it information, it will eventually run out of its own ability to process, and then it won't be any better than a human. It might be quicker, but it might have a human-like margin of error.

David: No, if anything the opposite is true. In the pursuit of first-mover advantage, we spend too little time testing and fully understanding the potential and limitations of the technology.

AI has the potential to disrupt the advertising industry even more profoundly than with the advent of programmatic advertising. We are only now, a decade after its widespread adoption, beginning to understand the shortcomings of programmatic advertising, especially for premium publishers. Therefore, as we rush to embrace AI models, it's crucial to avoid the pitfall of industry-wide mass adoption before fully understanding their limitations.

While the largest companies will continue to invest billions in developing and training their models, the risk is that the rest of the industry will struggle to keep up and lack the resources to compete. I'd like to see greater consensus across the industry on shaping the future of AI, establishing guidelines for its training and adoption, particularly concerning data usage.

Marcus: I don't perceive a barrier to improving AI capabilities, and it is incumbent on companies like Smartology to keep striving for better results and demonstrating to clients the benefits of AI over other methods.

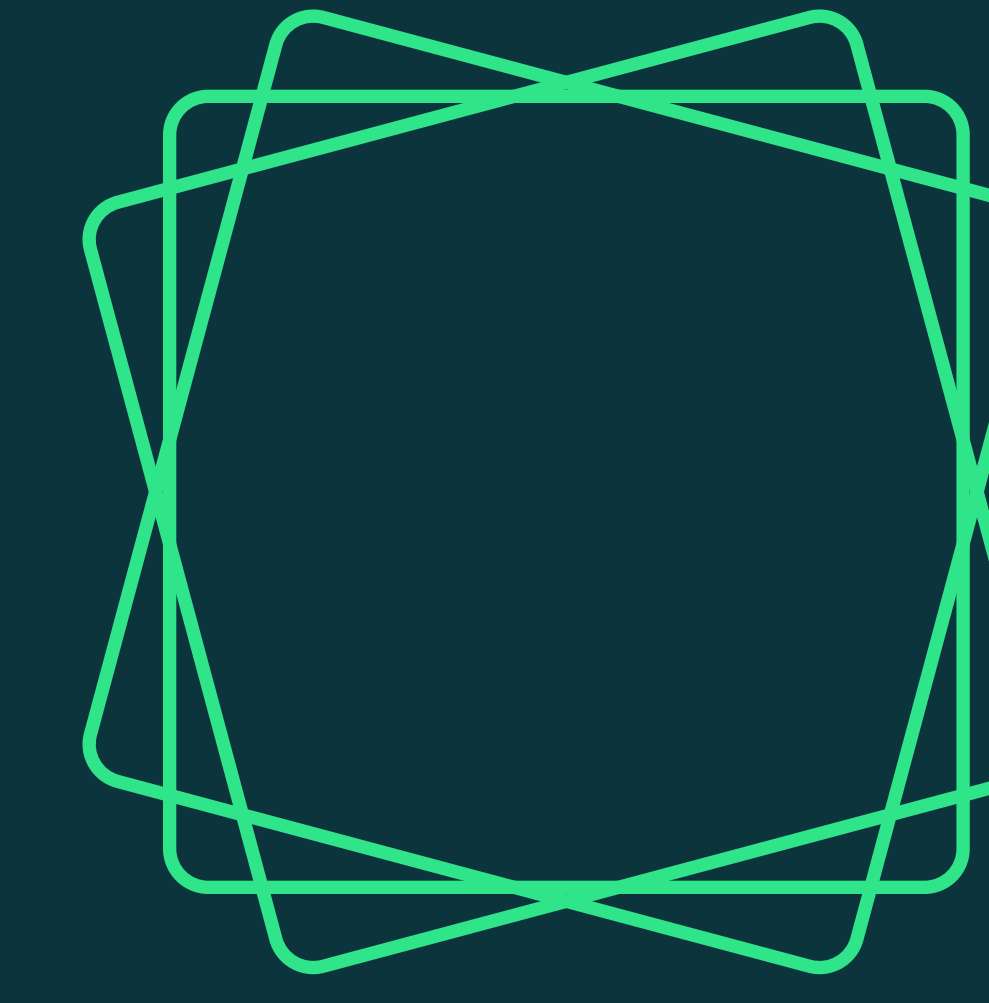
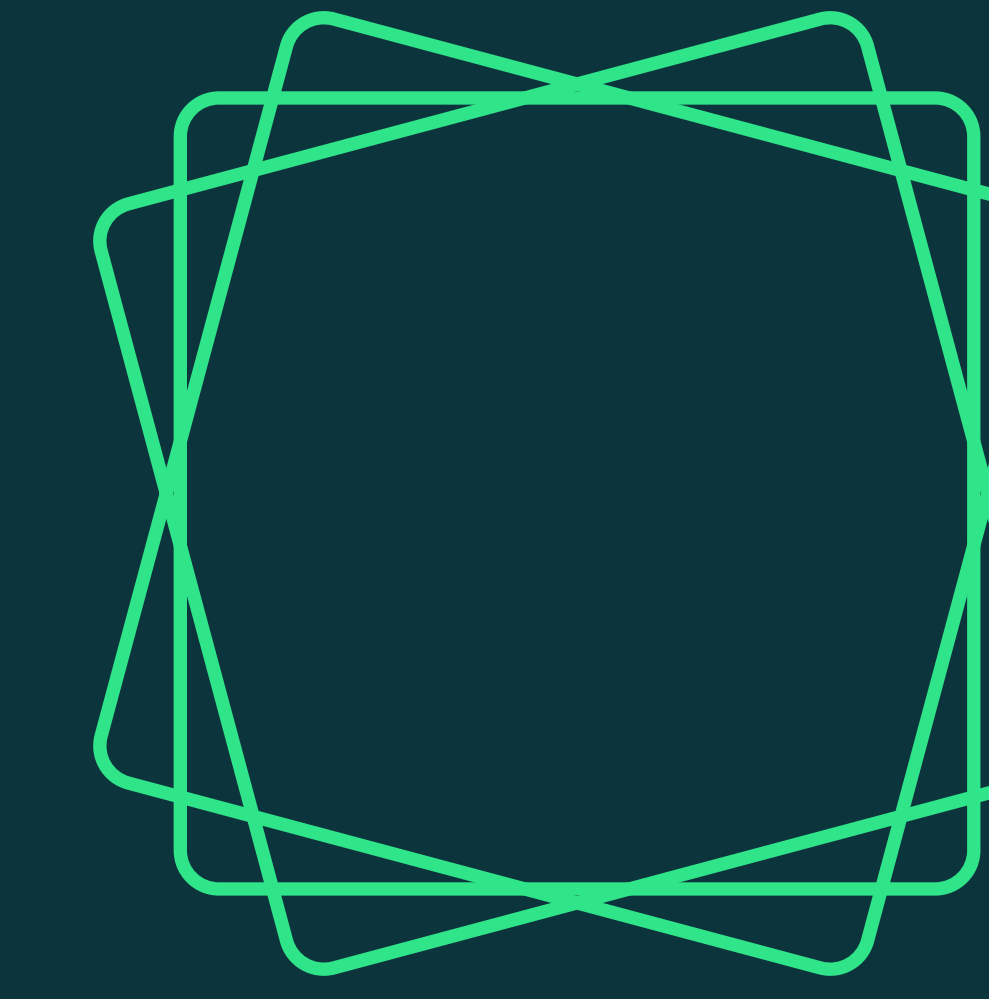
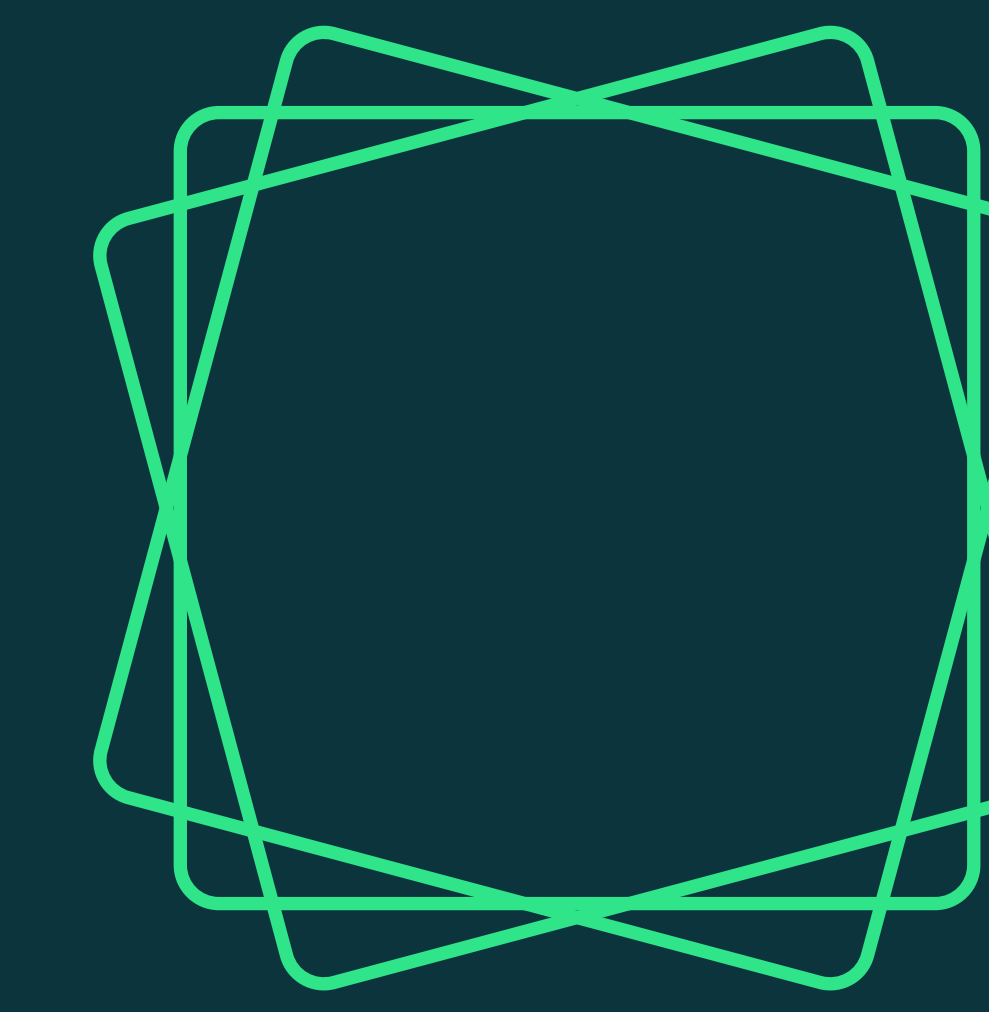
SmartMatch™ applies AI brand safety to all campaigns and has a feedback loop, so like us it is constantly learning and improving. Campaigns tend to focus on impressions, viewability, clicks and other statistics around delivered ads rather than how effective the brand safety tools were - what we didn't deliver - or the opportunity costs associated with that. Addressing this is something I'm delighted to see on our roadmap as it will lead to better understanding and the conversations around this data will inevitably lead to better brand safety.

How do you think we can stay critical and transparent about what AI is actually capable of delivering at the moment?

Patrick: I don't think that the user layer is there yet. We start off in this conceptual way, and then it gets built practically, it's available through APIs, but it's not accessible.

When we see people demoing things practically we can get a real grasp on capabilities, but until an AI model becomes mainstream, it's still very abstract for most people.

David: As with every exciting new development, we have to embrace the potential while maintaining a healthy scepticism for some of the wilder claims.



“AI has the potential to disrupt the advertising industry even more profoundly than with the advent of programmatic advertising.”

My advice is to start with a clear focus on your own business, its USPs and what AI can do to enhance these. Keep initial projects small, manageable and, crucially, measurable. Avoid investing funds and resources without a clear understanding of your expected return. Most importantly, recruit the right people to lead these smaller, targeted projects, who can also serve as advisors for the more ambitious “blue sky” initiatives.

Marcus: As with any new technology you’ll have your early adopters whose use and feedback develop that technology ready for wider adoption.

It’s fair to say that it won’t be perfect and so businesses must be up front about their appetite for risk. If you have a zero tolerance where even one miss in a million is too much, then a purely AI approach to brand safety probably isn’t right for you...yet. But if some small percentage of false negatives is acceptable and you have a process for identifying, reporting and addressing those ‘misses’ in partnership with your brand safety provider, then you have a platform for transparency that will allow you to make the right decisions for your campaign.

How much of what's being promised around AI do you think is actually new, and how much is a rebranding of existing capabilities?

Patrick: That’s the crux of a lot of the AI conversations to an extent, because a lot of this conversation is sort of framed around the fact that we're at this big tipping point.

When it comes to AI versus machine learning, machine learning has really fuelled programmatic so far. It's not a human making decisions with bidding algorithms, for example. A human is feeding in information, but it’s machine learning that is making the bids, all within microseconds.

AI is making the decisions, in the sense that it's not just following a very specific template like a bidding algorithm would, but acting in a dynamic way. AI is still following rules, but it will have more of its own level of thinking, and I think that’s what’s new and different.

Overall, the rulebook stays the same, but the speed at which the rulebook is read, interpreted or rolled out is different.

David: In the past, LLMs and machine learning were used to develop specific programmes and technologies but stayed within the domain of the product and engineering teams. While some of the technology has been around for some time,

the adoption of agentic and generative AI at scale has unlocked unprecedented possibilities for individuals and businesses.

The ability to write content, develop plans and analyse data using AI is now readily accessible to everyone. Although expertise remains crucial for meaningful application, we are undoubtedly at a pivotal moment, as Patrick observed.

Based on industry discussions, I am optimistic that this will result in improved decision-making and better outcomes for advertisers.

Marcus: Given the phenomenal investments in AI, naturally there's a lot of hype and repackaging to sell the AI story. But there’s no denying that the advancements made in the last couple of years have been exceptional.

I’m really excited about the improvements in tooling in particular which is significantly lowering the cost and accessibility for AI development and deployment.

Key Takeaways: Summarising the Brand Safety Discussion

The consensus is clear: the future of brand safety lies beyond the blunt instrument of keyword lists, moving towards the nuanced understanding offered by agentic AI and advanced machine learning. While AI offers the potential to interpret the full emotional weight and context of content, thereby recovering lost revenue for publishers, human oversight remains vital for setting a brand's unique risk tolerance, especially around sensitive global events. The main barriers to full adoption are currently strategic and cultural; namely, skepticism and the slow pace of change in larger organisations, rather than technical capability. As the ad tech industry navigates this transition, the focus will shift from simple content exclusion to a more dynamic, AI-informed approach, setting the stage for future discussions on how these same agentic capabilities will soon revolutionise campaign management and functional automation.

