→ AI

# Rise of the Multi-Armed Bandit Algorithms

**Customer-centric industries like media, retail, banking, insurance, and healthcare rely on digital advertising to acquire new customers and boost customer engagement. As audiences grow and evolve, campaign goals shift, prompting companies to develop innovative approaches to better reach their target audiences and maximise results.**

Programmatic advertising campaigns utilise multiple channels to deliver personalised messages to specific audience segments. However, optimising campaigns across a vast array of channels to maximise outcomes is a real challenge.

The manual process of running, analysing, and adjusting campaigns – despite the use of reporting tools – often proves inefficient.

To address this, many companies employ A/B testing to determine the best-performing variant and maximise returns.

In this approach, different versions are tested equally over a set period, with the best-performing option chosen for the remainder of the campaign. This creates a clear division: learning (testing) and earning (using the winning option).

However, two issues arise:

1. During the learning phase, campaigns are not optimised for maximum returns, as low-performing options are used equally with high-performing ones.
2. Market and consumer factors can shift quickly, making results from earlier tests less relevant. For instance, user behaviour before an event like Black Friday may differ significantly during the event itself.

The ideal solution is a system that continuously learns and optimises campaigns in real-time, leveraging the most current data.

This challenge is known as the exploration-exploitation dilemma: Balancing the need to explore new strategies with the need to exploit the best-known option to maximise long-term results.

This is where multi-armed bandit (MAB) algorithms come into play.

These algorithms are designed to handle this balance effectively and optimise outcomes dynamically.

## What is a multi-armed bandit algorithm?

The multi-armed bandit problem and its algorithms are a subset of reinforcement learning, which itself falls under the broader umbrella of artificial intelligence (AI).

Reinforcement learning focuses on continuous learning from an environment through trial-and-error, where feedback signals strengthen correct actions and diminish incorrect ones.

Multi-armed bandit algorithms are particularly effective at balancing exploration (testing new options) and exploitation (leveraging known options) to maximise total rewards, offering relatively simple yet efficient solutions.

The problem gets its name from the classic one-armed bandit (slot machine) in casinos. Imagine a row of slot machines, each with an unknown probability of payouts. Your goal is to maximise your total earnings over time. While it may be tempting to pull the lever on what seems to be the most lucrative machine, the inherent uncertainty means you occasionally need to test the others to confirm or adjust your strategy.

Piotr Banaszczyk
CEO
CLEARCODE
PART OF QINSHIFT

## Applications of multi-armed bandit algorithms in programmatic advertising

In the programmatic advertising industry, campaigns require strategic decisions on who to target, which creative to use, how to allocate budgets, and how much to bid. Simultaneously, we aim to maximise key outcomes, such as clicks and conversions.

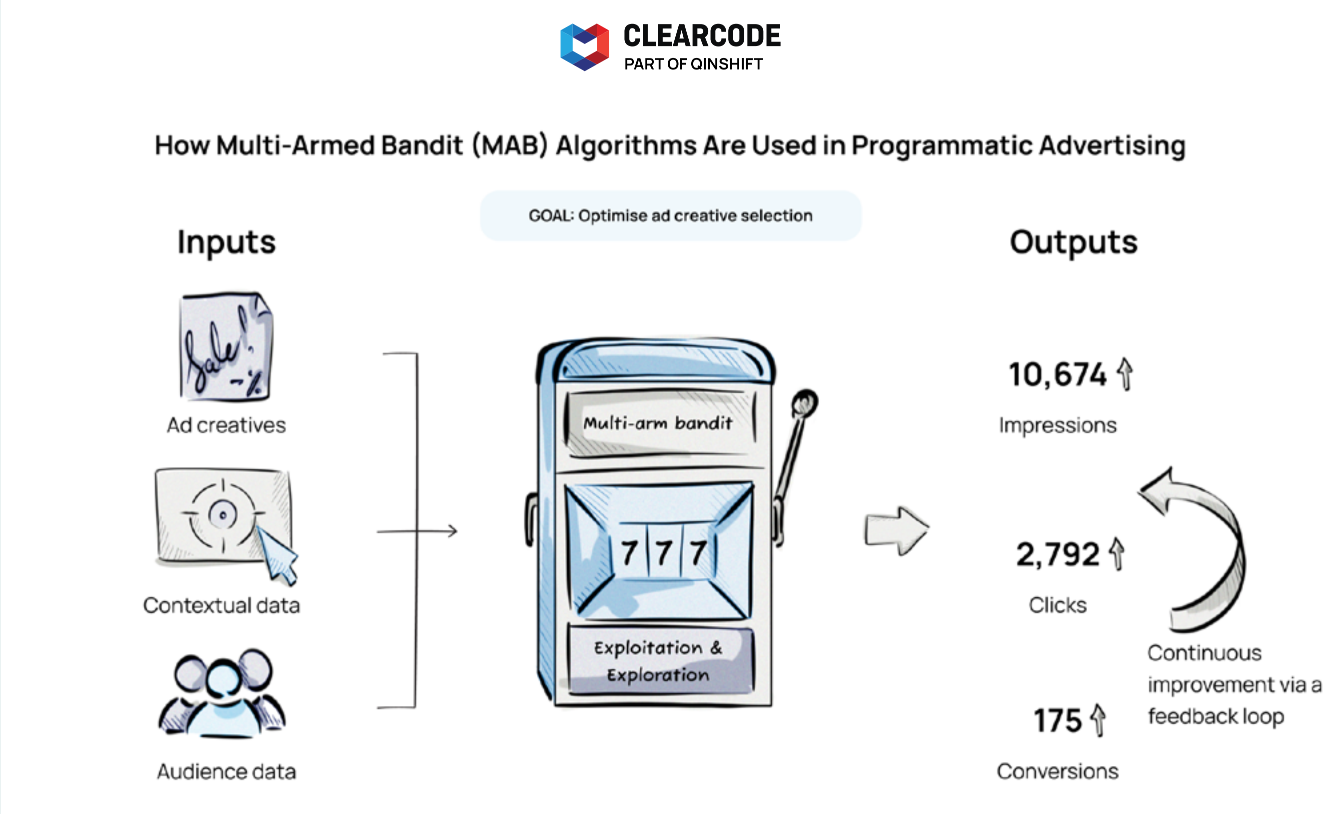
There are numerous multi-armed bandit algorithms, from basic ones like epsilon-greedy, UCB, and Thompson Sampling, to more advanced variants, such as contextual bandits, Bayesian bandits, and adversarial bandits, as well as those using deep neural networks.

However, no single solution fits all cases, regardless of how advanced the technology may seem.

To effectively leverage multi-armed bandit algorithms in programmatic advertising, we need to model our specific use case accordingly. This involves defining actions, rewards, and contexts:

**Actions:** A set of options from which we must choose. For example, given several creatives for each request, we need to select one to serve.

**Reward:** Feedback we observe in response to our chosen action. For instance, we might observe a click event after displaying a creative.




**Context:** Information about the environment and available options, used to optimise the decision-making process. For example, one creative might generate more clicks when shown to a specific audience, while another performs better with a different audience.

The spectrum of implementations ranges from simple to highly complex systems where multi-armed bandit algorithms are just the tip of the iceberg.

Often, the most effective approach is to start simple and expand gradually.

## Popular use cases of multi-armed bandits

Several leading companies have publicly shared their successful applications of bandit algorithms:

- Yahoo! News used contextual bandits to select promoted stories, achieving a 12.5% increase in clicks over context-free approaches.
- Wayfair applies bandits to optimise advertising and marketing throughout the customer lifecycle.
- Meizes utilises multi-armed bandit algorithms to automatically allocate campaign budgets across different social media platforms (Facebook, Instagram, etc.) and optimise bid prices.
- Lyft reported USD$30m (£24m) in annual savings by using multi-armed bandit algorithms for budget allocation in acquisition campaigns.
- Netflix leverages multi-armed bandit algorithms to decide which artwork to show to a user to maximise likelihood of engagement.

Multi-armed bandit algorithms provide an effective way of optimising decisions, particularly in environments with numerous potential actions and inherent uncertainty – such as advertising campaigns and customer engagement driven by preferences, trends, and events.

One advantage of multi-armed bandit algorithms is their scalability; businesses can start with a targeted application and gradually expand as needed, adapting dynamically to changing market conditions.

These algorithms have been well-researched and widely used across industries for over a decade, though they haven't received the same level of hype as recent technologies like Generative AI.

However, they are highly complementary to generative models, enhancing acquisition strategies and customer experiences.

As we move forward into 2025 and beyond, the hype around AI will fade and be replaced with a focus on utilising and implementing proven solutions to deliver real business value. Multi-armed bandit algorithms are one example of this