Compute Sponsoring

Toward a Category Definition for Advertising in Generative AI

September 2026

Author

Christian Leinberger

wavebird

Published on

Compute-Sponsoring.org

Abstract

This paper works toward a category definition for advertising in generative AI. A category definition in this sense develops a structure that holds from the problems that arise when advertising enters GenAI: a reference point, principles, and classifications proposed on that basis. Because GenAI produces content in the moment it could be altered by advertising, which is an integrity problem. But even if advertising does not alter the output of a GenAI system, a trust problem arises. This does not only affect users, but also advertisers who cannot prove that they did not influence the content. Since the content did not exist beforehand, addressing this at the level of content is not straightforward. The structural solution is to make the relationship between the GenAI system and advertising addressable through coupling advertising to a compute unit as a reference point. As advertising can draw on semantic sources the compute unit contains, a classification and notation are proposed. Situating the concept in the research literature remains open.

1. Introduction

Generative AI (GenAI) is spreading rapidly and will shape the digital landscape of the coming decades. A central paradigm shift of this technology lies in how digital value is created.

In the classic software paradigm, value is created primarily once during development and then scaled through distribution. In GenAI systems, digital value is not created solely through the one-time creation of software, but situationally through compute at the moment of use. Digital value creation becomes continuous.

Where compute becomes a constitutive part of usage, it simultaneously becomes a source of cost because it continuously binds resources. That this creates a monetization problem is widely recognised and discussed across the industry: systems must be limited, priced, or cross-subsidized. Advertising is a natural solution to the funding problem, but in the context of GenAI it creates a novel problem space, which is explored in the following.

This paper works toward a category definition for advertising in generative AI. It develops a structural analysis, works out the conditions that follow from it, and draws possible practical proposals from them. The analysis proceeds from the constitutive properties of GenAI rather than from empirical assumptions, while testing against practice and situating the work in the research literature remains open.

2. Advertising in GenAI is structurally different

In the digital space, advertising almost always appears in the context of content, whether on a website, before a video, or within a social media feed.

Until now, that content has been static in nature. It was created, uploaded, and then displayed, and advertising sat clearly apart from it. Personalisation did not change this: in classic digital advertising, content is tailored to the user, but through selection from an existing stock. With GenAI this structure changes. The content is generated in the moment and therefore there is no stock the content is pulled from.

This structural change brings a central problem with it: advertising could, for the first time, alter the content, because the content is generated in the moment and can therefore be influenced while it is being generated. If the substance of a GenAI output can be altered in this way, that is an **integrity problem**. Since GenAI can serve as a source of knowledge, it is an epistemic one.

But the risk is not only that advertising influences the GenAI output. Imagine that a particular brand appears naturally in a GenAI output, while that same brand is displayed beneath it as an advertisement by coincidence. Even if the advertising did not influence the output and was shown separately from it, a **trust problem** arises. As soon as doubt is raised about whether the advertising shaped the content, the integrity of the model is called into question. And the advertiser faces the same problem from the other side: it cannot demonstrate that it did not influence the content.

An obvious solution approach would be to examine the content for possible influence. But since the content was generated and did not exist beforehand, integrity cannot straightforwardly be established by that route, because no prior content exists against which it could be compared. The condition for solving these problems must therefore be structural in nature: it must be traceable at the level of the process whether a separation between GenAI and advertising was in place. This requires a clear and meaningful reference point for the advertising, in order to make the relationship between advertising and GenAI addressable.

3. Compute Sponsoring

The intuitive reading of the term Compute Sponsoring is that advertising sponsors the compute that makes GenAI systems available, which addresses the funding problem set out in chapter one. The term carries a second meaning:

Compute Sponsoring also means that a sponsorship exposure can be unambiguously attributed to an identifiable compute unit.

Sponsorship denotes a financial contribution that is coupled to an exposure. A sponsorship exposure is a clearly labeled event that can be measured independently. Classic brand sponsorship and classic digital advertising are typical manifestations of this category. For reasons of readability, “advertising” is used as an umbrella term for sponsorship exposures.

A compute unit denotes a unit that includes compute involved in a generation. What matters structurally is not the specific form, but that the unit is unambiguously determinable and used consistently within the system. Since advertising in GenAI appears most readily in the context of a single job, that is, one request from a user seeking a particular outcome, a job can serve as a stable and productive compute unit. A job in this sense consists of an intention, the generation process that responds to it including any intermediate step, and a result.

The reference point has to be the compute process itself. It is a process and not a result, and therefore the only thing about which a statement can be made independently of what was produced. And it is the process in which the content is generated, which is the reason the structural changes described above arise in the first place.

The term Compute Sponsoring therefore brings both the economic and structural argument together. Advertising that sponsors the compute is also coupled to the compute as a reference point.

4. Principles

Building on the definition, the following principles are proposed as characterising Compute Sponsoring.

Separation of GenAI and advertising

The GenAI model remains untouched by advertising. Advertising always takes place on a separate technical channel.

No input influence

Sponsor information is never injected into the model request and is not introduced into the prompt, the context window, or tool inputs.

No output influence

The content of the output is never influenced by advertising.

Independence from system-level decisions

Advertising never determines the selection of the model or the tools.

The principles above concern the separation of advertising from generation. A second question arises from the same structure regarding data usage. The dimensions the following principles refer to are set out in chapter five.

Data sovereignty

Sponsors never receive user-level or request-level semantic usage data. Sponsor reporting must not enable reconstruction of individual intentions, problems, or contexts.

Consent to the use of semantics

Semantic relevance targeting of ads based on user intention and/or model intelligence is tied to explicit consent.

Consent to persistence of semantics

Without explicit consent, semantics are not accumulated across sessions or condensed into a profile-like history.

Transparency of semantic relevance targeting

Users must be clearly informed when semantic relevance targeting is used for ad selection and delivery.

5. Classification of semantic relevance targeting

Where advertising is coupled to a compute unit, it can also draw on what that unit contains. A compute unit, taken as a job, consists of an input, a generation (including whatever intermediate steps it involves), and an output, and it is bounded in time. Both of these properties open a dimension of their own: what the advertising draws on, and how long it may hold it.

A compute unit has exactly two places where semantics arise: what the user brings to it, and what the GenAI system produces. Any semantic signal drawn from a compute unit stems from one of these two, or from their combination.

Semantic Source

Semantic Source describes the possible sources for semantic signals used in semantic relevance targeting in the context of a compute process.

S0: No semantics

No semantics are used for relevance targeting.

S1: Intention (user semantics)

Relevance targeting is based on the semantics of the user intention. In practice, this usually derives from the prompt.

S2: Model intelligence (model semantics)

Relevance targeting is based on the model-derived semantic signals. In practice, this usually derives from intermediate results or the model output.

S3: Combined source (intention + model intelligence)

Relevance targeting is based on semantic signals from both user intention (S1) and model intelligence (S2).

A compute unit is also bounded in time. Semantics either do not outlast it, outlast it within a session, or outlast the session itself.

Semantic Persistence

Semantic Persistence describes how long semantic data are stored and used for semantic relevance targeting.

P0: No persistence

Semantics are not stored. Signals are purely transient.

P1: Session persistence

Semantics remain usable within a session, without being carried across session boundaries.

P2: Cross-session persistence

Semantics are stored and accumulated across sessions and used for semantic relevance targeting.

6. Compute Sponsoring profiles

The source and persistence of the use of usage-based semantics create a matrix from which different Compute Sponsoring profiles can be constructed:

	P0	**P1**	**P2**
S0	Compute Sponsoring Neutral	Not applicable	Not applicable
S1	Compute Sponsoring Situational	Compute Sponsoring Situational	Compute Sponsoring Profiling
S2	Compute Sponsoring Situational	Compute Sponsoring Situational	Compute Sponsoring Profiling
S3	Compute Sponsoring Situational	Compute Sponsoring Situational	Compute Sponsoring Profiling

For S0, “Semantic Persistence” beyond P0 is not applicable, since no semantic signals are used.

Profile 1: Compute Sponsoring Neutral (CS-N)

Compute Sponsoring Neutral describes Compute Sponsoring without any semantic relevance targeting (S0). CS-N is the least invasive form of Compute Sponsoring.

Profile 2: Compute Sponsoring Situational (CS-S)

Compute Sponsoring Situational describes semantic relevance of advertising in the moment, without accumulating semantic data beyond a session. Semantics are used to make advertising situationally more appropriate, either from intent (S1), from model intelligence (S2), or from intention and model intelligence (S3), but only transiently (P0) or within a session (P1). CS-S is the zone in which relevance can be used without generating profile-forming data.

Profile 3: Compute Sponsoring Profiling (CS-P)

Compute Sponsoring Profiling describes systems in which semantic signals are accumulated across sessions (P2). Whether the source is intent (S1), model intelligence (S2), or both (S3) is

secondary; what matters is the storage and use of semantics beyond specific usage situations. As soon as semantic data are stored across sessions, profile-specific advertising can emerge.

7. Notation: description of systems

The distinctions above can be stated. The following notation is one way of doing so: a system names its profile and its position on both axes.

	P0	**P1**	**P2**
S0	CS-N (S0/P0)	Not applicable	Not applicable
S1	CS-S (S1/P0)	CS-S (S1/P1)	CS-P (S1/P2)
S2	CS-S (S2/P0)	CS-S (S2/P1)	CS-P (S2/P2)
S3	CS-S (S3/P0)	CS-S (S3/P1)	CS-P (S3/P2)

The notation only applies where the advertising is coupled to a compute unit and the principles hold. The axis position then says which form of Compute Sponsoring a system uses.

Example

A system has the following properties: the user submits a prompt. While the model processes the request, topic-specific ads are displayed based on the prompt. The prompt's semantic content is not stored and therefore cannot be used for semantic relevance targeting of ads on the user's next request, even within the same session.

This system is described as: **CS-S (S1/P0).**

8. Conclusion

Compute Sponsoring responds to a new structure. In GenAI systems, content is produced in the moment, which gives rise to the structural possibility that advertising influences it, and with that to a problem of trust in the integrity of the system. The advertiser faces the same problem, as it cannot show that it did not influence the content. This cannot easily be addressed by examining the content, which did not exist beforehand, but at the level of the process that

produced it. Coupling advertising to a compute unit is what makes the relationship between the two addressable. From this follow principles on the separation of advertising and GenAI, and further principles on data usage. With Semantic Source and Semantic Persistence, the paper proposes a structural classification, and with the notation a way of describing where a system stands.

Note

A first version of this paper appeared in February 2026 as *Compute Sponsoring: A Category Definition for Advertising in Generative AI*, co-authored with Mario von Bassen. It was written as a white paper addressed to industry practice. The present version takes a more academic form: it makes the underlying argument explicit and removes passages referring to technical specification. It was written by the present author alone. AI assistance was used for language and formulation.

No claim to novelty is made for the individual concepts it develops. The argument here is structural and does not rest on prior findings and the paper therefore carries no references. Situating the category in the research literature is the next step. Where that situating shows the structural assumptions do not hold, the category is to be refined accordingly.