Ruggero Eugeni, "The Dividual Enunciation. Rethinking the Production of Images in the Age of Transformers", in Dario Mangano, Paolo Peverini (Eds.), *Semiotics in the Age of Artificial Intelligence*, Cham, Springer, 2026, pp. 151-175. https://doi.org/10.1007/978-3-032-29272-8_8

Abstract

The concept of enunciation has long held – and continues to hold – a central role in semiotic studies, serving both as a theoretical framework and as a tool for analysing texts, discourses, and signification processes. This paper seeks to apply the concept of enunciation to Generative Artificial Intelligence (GenAI), particularly Visual Generative Artificial Intelligence (VGenAI). My intent is twofold: on one hand, I intend to develop a deeper understanding of how algorithmic machines operate in the automated creation of images; on the other, I want to test the flexibility of the theoretical notion of enunciation and to identify the transformations that VGenAI introduces into this concept. My central hypothesis is as follows: traditionally, reflections on enunciation have emphasized the mutual *individuation* between utterances and their subjects, a kind of identification enacted through discursive production practices. In contrast, GenAI and VGenAI offer a different model based on operations of *dividuation* – that is, the arbitrary partitioning of both discourse and its subjects. This new scenario, on one hand, connects enunciative practices to a broader economic and political context; on the other, it prompts a reconsideration of the fundamental concepts of enunciation theory.

Chapter 8
The Dividual Enunciation. Rethinking the Production of Images in the Age of Transformers

Ruggero Eugeni

8.1 Introduction: Why (Not) Read this Chapter

The concept of enunciation has long held – and continues to hold – a central role in semiotic studies, serving both as a theoretical framework and as a tool for analysing texts, discourses, and signification processes. This paper seeks to apply the concept of enunciation to Generative Artificial Intelligence (GenAI), particularly Visual Generative Artificial Intelligence (VGenAI). My intent is twofold: on one hand, I intend to develop a deeper understanding of how algorithmic machines operate in the automated creation of images; on the other, I want to test the flexibility of the theoretical notion of enunciation and to identify the transformations that VGenAI introduces into this concept. My central hypothesis is as follows: traditionally, reflections on enunciation have emphasized the mutual *individuation* between utterances and their subjects, a kind of identification enacted through discursive production practices. In contrast, GenAI and VGenAI offer a different model based on operations of *dividuation* – that is, the arbitrary partitioning of both discourse and its subjects. This new scenario, on one hand, connects enunciative practices to a broader economic and political context; on the other, it prompts a reconsideration of the fundamental concepts of enunciation theory.

The paper is structured into three sections. In the first one, I introduce the concept of enunciation and outline its key developments within semiotic studies. Given the complexity of the topic, this section is inevitably brief but helpful for understanding the rest of the paper; readers already familiar with semiotic concepts may safely skip it. The second section addresses the ongoing semiotic discussion regarding the analysis of VGenAI through the tools of enunciation theory. I critically examine this debate, considering that recent generations of VGenAI employ specific models

R. Eugeni (✉)
Università Cattolica del Sacro Cuore, Milan, Italy
e-mail: ruggero.eugeni@unicatt.it

P. Peverini and D. Mangano (eds.), *Semiotics in the Age of Artificial Intelligence*,
Culture in Policy Making: The Symbolic Universes of Social Action,
https://doi.org/10.1007/978-3-032-29272-8_8

created for analysing and generating verbal texts – specifically, Transformer architectures – by connecting them with models designed more directly for image analysis and generation. A semiotic examination of these different procedures and their connections within specific algorithmic dispositives is therefore necessary. Additionally, I aim to expand this debate by addressing a topic that has received less attention until now: the analysis of user interactions with algorithmic machines occurring within query interfaces. Readers already familiar with these aspects and with VGenAI technology may safely skip this section. The third section situates the economic and political dimensions of the ongoing debate about the concepts of the dividual and dividuation. It illustrates the heuristic and theoretical benefits of applying these concepts to GenAI and VGenAI techniques and interpreting them in terms of enunciative and enunciational processes. Readers already acquainted with dividual theories (from Gilles Deleuze to Gerald Raunig, Michaela Ott, and others) can safely skip this final section.

In essence, readers who are already well versed in the semiotics of enunciation, familiar with the technical mechanisms of algorithmic image generation, possess a basic understanding of interactive interface analysis, and have kept up with the debate on the emergence of the dividual in control societies, may comfortably skip this chapter.

8.2 The Individual Enunciation

8.2.1 The Interpersonal Enunciation

French linguist Emile Benveniste introduced the concept of enunciation in its modern sense in a series of articles published between the 1950s and 1970s. For Benveniste, "The enunciation [*énonciation*] is the enactment of language through an individual act of use [that is,] the act of the speaker mobilizing the language on their behalf. The relation of the speaker to the language determines the linguistic characteristics of the utterance [*enoncé*]. It must be considered the act of the speaker, who takes the language as an instrument, and with reference to the linguistic features which mark this relation." (Benveniste, 2014, p. 141). Despite its apparent clarity, this definition opens a complex field in which we can identify at least two significant aspects. First, regarding the *operation* performed, there is the action of producing the utterance by the *enunciator*, and therefore his or her interaction with the repertoires and expressive resources available; this operation also includes the presentation of the utterance to a listener or interlocutor, the *enunciatee*. Second, regarding the *product* itself, the utterance contains particular personal, spatial, and temporal coordinates from the factual context in which it occurs. This incorporation is achieved through various linguistic forms, especially elements known as *deictics* or *shifters*. These linguistic elements can be personal (I/you), spatial (here/there), or temporal (a moment ago/in a moment), and gain meaning by referencing the personal, spatial, and temporal system

within which the utterance takes place. This incorporation serves a dual purpose: on one hand, it reflects or describes the situation of enunciation; on the other, it defines and constitutes it. In any case, the enunciator maintains a certain degree of flexibility here: he or she can openly display enunciational markers (regime of *Discourse*) or conceal them, as when producing an utterance "in the third person" (regime of *Story*).

For Benveniste, these two components – enunciative operations and enunciational forms of utterances – are closely interrelated. To understand this connection, it is crucial to recognize that Benveniste primarily focuses on actual uses of language within face-to-face dialogic interactions. Within this context, the general "characteristic of enunciation is *the accentuation of the discursive relationship to the partner*" (Benveniste, 2014, p. 145). In other words, enunciation serves as a means of *individuation* – or more precisely, *co-individuation* – of the partners involved in communicative interaction. By embedding the context of enunciation within the utterance, speakers collaboratively construct the social framework of their interaction, negotiating or reaffirming their roles and identities within the communicative scenario.

8.2.2 The Personal Enunciation

Benveniste (2014) concludes his seminal essay by observing that "we should […] distinguish the spoken enunciation and the written enunciation. The latter moves on two levels: the writer enunciates himself [sic] by writing and makes individuals enunciate themselves within his writing" (145). Indeed, semiotics and related disciplines such as narratology shifted their focus during the 1970s and 1980s toward texts, moving from face-to-face interactions to mediated productions like novels, paintings, and films. This new application led to a first turning point in enunciation studies: scholars utilized the concept of enunciation to reframe discussions of "point of view" in written, visual, and audiovisual texts. For example, Gerard Genette (1980) emphasizes the necessity of analysing novels not only in their narrative dimension [*récit*] – that is, as narrative utterances – but also in terms of narrating [*narration*], understood as an action performed by a narrator addressing a narratee. The focus on narrative enunciation suggests at least two directions of inquiry: on one hand, distinguishing and exploring the relationship between "the question who sees? and the question who speaks?" in narration (Genette, 1980, p. 186), and thus between the narrator's gaze (which Genette refers to as "mood") and voice. On the other hand, it concerns the narrator's relationship to the fictional world about which the narrative speaks (called "diegesis): in relation to this, the narrator (both as voice and as gaze/mood) can be internal (that is, one of the characters – an "intradiegetic" narrator) or external (an "extradiegetic" one). The narratee (the narrator's partner) can occupy the same positions. Various scholars have extended Genette's framework to other media, such as images (Dondero, 2020, pp. 15–48) or audiovisual media (Casetti, 1998), noting how the enunciator and enunciatee become abstract, invisible entities

positioned "upstream" of the text, while the various figures of narrators and narratees constitute respectively "the traces [or marks] of the enunciator and those of the enunciatee [endowed with] several levels of figurativization" (Casetti, 1998, p. 30).

The most thorough formalization of the concept can be found in the semiotics of Algirdas Julien Greimas. Greimas's theory is articulated as a "generative trajectory" of the text: "since every semiotic object can be defined according to its mode of production, we postulate that the components that enter into this process are linked together along a 'trajectory' which goes from the simplest to the most complex, from the most abstract to the most concrete" (Greimas & Courtés, 1982, p. 132). The deep level consists of *semiotic and narrative structures*, the intermediate level contains *discursive structures*, and the surface level comprises *textual structures* (which differ according to the substance of expression used for the final manifestation of the text). Enunciation intervenes during the transition from semiotic and narrative structures to discursive structures, performing two related functions. First, it is "the domain of mediation" between the *virtual* nature of the deep generative levels and their *actualization* at the surface level; thus, enunciation "is the place where semiotic competence is exercised." Second, it is simultaneously "the domain of the establishment of the subject (of the enunciation)" (Greimas & Courtés, 1982, p. 104). Establishing discourse entails introducing subjective, spatial, and temporal coordinates derived from the abstract subjects, place, and time where the discourse originates: "It is the projection (along with the processes that we bring together under the name of disengagement), outside [the] domain [of the 'ego hic et nunc'], both of the actants of the utterance and of the spatio-temporal coordinates, which constitutes the subject of the enunciation by everything that it is not. It is the rejection (along with the processes that we call engagement) of the same categories that are intended to hide the imaginary place of the enunciation, which confers upon the subject the illusory status of being." (Greimas & Courtés, 1982, p. 104). Two aspects here are particularly significant: first, enunciation relies on the combination of disengagement [*débrayage*] and engagement [*émbrayage*] processes – that is the disjunction or expulsion of the I–here–now respectively from the enunciation domain (disengagement) and from the utterance (engagement). Second, discoursivization simultaneously involves processes of actorialization (concerning subjects), temporalization, and spatialization.

The shift that has occurred within the conceptual map originally outlined by Benveniste, because of the increased attention given to written, visual, and audiovisual texts, is clear. On one hand, enunciation operations have become abstract instances presupposed by utterances – at most, traces of the original idea can be found in the concept of "writing", the combination and renewal of codes within texts (Metz, 1974, pp. 254–288). On the other hand, scholars now consider utterance as the scene of a communicative interaction involving entirely simulacral figures – actually, a "fictivization of enunciation" (Odin, 2022, p. 76). However, efforts to think enunciation as an individuating procedure have not vanished: instead, they now tend to manifest as a means of assigning the viewers a particular role (at the same time visual, narrative, and ideological) through their "positioning" (Rosen, 1986).

8.2.3 The Impersonal Enunciation

Starting in the 1990s, a second shift occurred in enunciation studies, emphasizing its "impersonal" nature. Two seminal texts mark this shift. The first is Metz (2015), programmatically titled *The impersonal enunciation.* Metz's intent is epistemological: he argues that film theory must not endorse or systematize the proliferation of imaginary, personal, and anthropomorphic subjects typically found in fictional texts (enunciators, enunciatees, narrators, narratees, etc.). Such an approach would incorrectly equate textual enunciation with face-to-face interaction, thus implying a symmetry between the two poles of communication that does not exist in film – or, by extension, in any form of textual enunciation. Instead, theory must highlight the unique situation of viewing a film, involving a viewer who cannot interact with the film and a film that, as it unfolds, can represent its act of presenting itself to the viewer. Metz concludes that "*enunciation* [is to be considered as] a process or … a function rather than an object… Positioned like a soliloquy, enunciation separates itself from interaction. Its deixis is simulated, and its principal markers are foldings [sic] of the text back on itself. *The film self-designates itself because there is only itself*… [In conclusion,] film enunciation is impersonal, textual, and metadiscursive and that it may comment or reflect on its own statement in a variety of ways" (Metz, 2015, p. 163, 173). The second seminal text is Bruno Latour's *Piccola filosofia dell'enunciazione* [*Tiny philosophy of enunciation*] (1998, further developed in Latour, 2013). Latour revisits Greimas's concept of enunciation but broadens its scope within an ontology addressing different "modes of existence" of entities. In this broad and pre-linguistic sense, Latour returns to the etymology of the term (*ex-nuncius*), redefining enunciation as an act of sending and emphasizing its qualities of transfer, transformation, substitution, mediation, and delegation. He also highlights how this continuous metamorphosis raises questions regarding the persistence of identity amid change and equally involves subjects and objects, creating hybrid entities such as *quasi-objects* and *quasi-subjects* (see Peverini, 2024, pp. 57–74).

These two contributions mark the shift from a static, regulated view of enunciation to a dynamic, processual conception. This new understanding appears on both fronts, central to the concept's development from the beginning: that of production and that of products. Terminology particularly distinguishes *enunciative* practices (production) from *enunciational* dynamics (products). On the first side, Jacques Fontanille (2006), pp. 183–209) proposes understanding enunciation as both an individual and collective operation, analysing "enunciative praxis," which "is […] especially concerned with the appearance and disappearance of utterances and semiotic forms in the field of discourse, or by the event constituted by the encounter between the utterance and the instance that takes charge of it… [In other terms,] enunciative praxis governs this presence of discursive entities in the field of discourse" (196). Essentially, enunciative practices – "praxis", in the terminology adopted by Fontanille in the referenced work – modifies the modes of existence of linguistic expressions. In an "ascending" trajectory, expressions move from *virtual* existence (within a linguistic system), through *actual* existence (their invocation in discourse not yet manifested),

to *real* existence (manifestation in discourse embedded in a particular substance of expression spoken, written, visual, etc.). Conversely, in a "descending" trajectory, expressions shift from *real* existence to *potential* existence (a reserve of expressions, motifs, and narrative structures available for new discourses) and back to *virtual* existence. Fontanille's model thus explains, in a narrow timeframe, discourse production from linguistic repertoires and the subsequent transformation of these repertoires (what Metz, 1974 termed "writing"), and, in a broader timeframe, the transformation of languages and discourses through the introduction or abandonment of linguistic expressions and forms within a "semiosphere".

On the complementary side of enunciational dynamics, Claudio Paolucci (2021, pp. 27–62) examines the relationship between linguistic practices and the constitution of subjectivity within updated cognitive semiotics. His conceptual framework stresses that subjectivity, understood as reflective self-awareness, emerges and continually evolves through a subject's continuous interactions with other subjects and objects, encompassing identification, disidentification, lies, hypotheses, and fictions. This approach acknowledges language's role in forming the subject (aligning with Benveniste's insight) but contends that this formation does not occur simply through the "I/you" versus "he/she" polarization (as Benveniste incorrectly suggested, for instance, by opposing discourse and story regimes). Instead, Paolucci describes a dynamic structured in two phases: first, the speaking subject establishes itself as an impersonal linguistic entity interacting with other subjects or objects whose personhood is similarly suspended (a condition which he defines delocutive status, illeity, union of person and non-person: e.g., the impersonal "it" in expressions like "it rains"). In the second phase, the subject identifies, within this fluid network of roles and presences, those (subjects or objects) who "speak (and act) for him" – according to Latour's idea of *ex-nuncius*: "Inside every discourse there are semiotic entities which cannot be reduced to the 'I'. For this reason, Benveniste's theory is unsatisfactory: it is radically subject-centric and ignores the other enunciating instances reverberating within every language act… [In this sense,] the illeity of the 'he[/she]' (third person) is an impersonal event which opens up the intersubjectivity of the 'I-you' ([first and] second person) which, in turn, makes possible the position of the subject (first person) … From the starting point of these illeities, the construction of a subjectivity involves a long and tortuous journey." (Paolucci, 2021, pp. 45–46, 57). In a sense, Paolucci's model echoes Greimas's but reverses it: engagement (*émbrayage*: suspending the identification of enunciational persons) precedes rather than follows disengagement (*débrayage*: (re)-identifying subject roles).

In conclusion, recent developments in enunciation theory emphasize the impersonal nature of enunciational operations, highlighting their procedural and dynamic character. Nevertheless, in my opinion, they do not fundamentally challenge the fundamental principle underlying enunciation: the mechanism of mutual individuation between produced discourses and producing subjects. Furthermore, the "impersonal" approach multiplies the relationships between both human and non-human enunciative agents and effectively eliminates the distinction between subjects and objects in the processes of mutual identification carried out through enunciation processes.

8.3 The Delegated Enunciation

8.3.1 The Semiotic Debate on VGenAI

Many scholars have recently applied the conceptual tools of enunciation theory to the procedures and products of various VGenAI platforms – such as Dall-E, Midjourney, Imagen, and others – particularly within the International Seminar on Semiotics coordinated by Maria Giulia Dondero and Juan Alonso Aldama (Dondero et al., 2025). I will briefly summarize the primary directions of this debate by referencing several positions expressed. Although the discussion touches upon, and sometimes intersects with, classic themes in the broader conversation around algorithmic images (such as the issue of creativity or data-driven analytical methods applied to large image corpora), its essential core revolves around the question of whether – and in what sense – the production of images via VGenAI can be considered enunciative operations. Two different viewpoints emerge in this context. On one hand, some scholars reject this possibility, primarily because they maintain that VGenAIs lack autonomous intentionality. The most precise articulation of this position is in Marion Colas-Blais (2025). According to her, VGenAI platforms construct a machinic enunciative sequence (*séquence énonciative machinique*) that connects human subjects (present at the beginning and end of the sequence) with mechanical agents (in the middle). This connection follows a logic of disengagement (*débrayage*), since it involves delegating productive agency from human subjects to mechanical agents and distributing it among them. However, these mechanical agents lack complete enunciative competence because they do not possess autonomous intentionality. In other words, duty, knowledge, and ability are separated from intentional will; therefore, the machine can co-create but cannot co-enunciate.

On the other hand, other scholars do not doubt that VGenAIs, despite lacking autonomous intentionality, perform enunciation – or rather, co-enunciation. For example, D'Armenio et al. (2024) clearly state that "generative AIs are enunciational machines – for the simple fact that they produce visual or verbal utterances [… More specifically,] we will define these artificial intelligences as follows: co-enunciating machines, devoid of intentionality and initiative, which nevertheless produce visual utterances in collaboration with a human operator and on the basis of highly structured and reconfigurable archives." In this perspective, Maria Giulia Dondero et al. (2025) addresses the question of the enunciative practices carried out by and through VGenAIs using Jacques Fontanille's semiotics (2006), which we examined earlier. According to this framework, the repertoires of forms that VGenAIs utilize for their productions should be considered objects existing in a *virtual* mode. More specifically, virtual space in this context appears as a "latent space." This term originally refers, in a technical sense, to the domain in which algorithms perform data manipulation operations destined to become images. However, Antonio Somaini has theoretically expanded this concept, defining it as "the abstract space within which complex, multi-dimensional data structures (such as images, texts, and sounds…) are represented in a more simplified, lower-dimensional form, in order to be processed

through different, mathematical operations" (Somaini, 2025, p. 21; see also Somaini, 2023, p. 77). For Dondero, "today, the realm of the virtualized… has been concretized in databases and is available for our manipulation (as analysts and as generators of new images from old ones)." (Dondero, 2025, p. 114). Thus, latent space, understood as a database (or a set of databases), constitutes a new form of operational archive. Image generation from such an archive (Dondero particularly examines the architecture of Diffusion Models) involves a phase of *actualization* (the production of an algorithmic image model based on user requests expressed through verbal prompts or visual examples), followed by a phase of *realization* (the actual visualization of the pictures). In this manner, the ascending trajectory of Fontanille's enunciational practices is mirrored in VGenAI operations; the opposite descending trajectory (less considered by Dondero) would involve the training processes of VGenAI (i.e., the creation of latent spaces – databases). Nevertheless, the notion of style remains central to this process: Dondero distinguishes between "regional style," defined by the specific visual characteristics of a given platform (e.g., pictorial vs. photographic), and "global style," determined by features specific to images from particular historical eras (e.g., Renaissance vs. Cubist, etc.).

In summary, the debate: (a) leaves open the question of whether VGenAI productions can be considered forms of enunciation; (b) analyses latent space as an operational digital archive of visual forms; (c) does not thoroughly examine the training processes of VGenAI; and (d) emphasizes enunciative practices rather than the enunciational dynamics resulting from interactions with VGenAI. I will discuss these points further in sections 3.3 and 3.4, beginning with a brief presentation of the current technical state of VGenAI in 3.2.

8.3.2 The Evolution of VGenAI from GANs to Foundation Models

In 2022, Imagen, Google's visual generation tool, underwent a significant overhaul in the transition from the second to the third generation. As the authors explain:

> Imagen builds [now] on the power of large Transformer language models in understanding text and hinges on the strength of Diffusion Models in high-fidelity image generation. Our key discovery is that generic large language models (e.g., T5), pretrained on text-only corpora, are surprisingly effective at encoding text for image synthesis: increasing the size of the language model in Imagen boosts both sample fidelity and image-text alignment much more than increasing the size of the image Diffusion Model. (Saharia et al., 2022, p. 1)

Imagen 3 thus jointly employs a Transformer-type language model (T5) and a visual generation model from the Diffusion Models family. The increased power of Imagen 3 does not stem from improvements in the Diffusion Model responsible for image creation; rather, it arises from enhancements in the Transformer itself – a stronger underlying language model results in better-quality images. This occurs because,

in the concatenation between the two components, the Transformer not only understands the linguistic prompt but also guides and controls the visual production process carried out by the Diffusion Model.

To fully grasp the novelty and implications of this process, it is helpful to briefly review the development of algorithms utilized by VGenAI (Bie et al., 2025; Li et al., 2025). Image generation using deep neural networks traces back to the introduction of Convolutional Neural Networks (ConvNets or CNNs) in the late 1980s and early 1990s, designed by Yann LeCun for automated image recognition. The concept behind CNNs was to replicate the human brain's visual processing: beginning with an analysis of minimal image elements (angles, lines, contours, etc.), ConvNets progress through the neural network layers to analyze increasingly larger image portions, producing progressively more general, complete, and abstract descriptions. Building upon ConvNets, Ian Goodfellow introduced Generative Adversarial Networks (GANs) in 2014. In this architecture, two convolutional networks compete against each other using the same corpus of training images. The first network (the generator) constructs potential images, progressively composing them by reversing the decomposition technique to create images similar to those in the source corpus. The second network (the discriminator) attempts to determine the nature of the image. Created to train convolutional networks in a semi-supervised manner, GANs demonstrated that neural networks could generate images that are not present in the original training corpus. The widespread adoption of ConvNets and GANs marked an initial period of social and artistic hype of VGenAI. Examples include the spread of DeepDream (developed by Alexander Mordvintsev between 2014 and 2015), a program utilizing a ConvNet named Inception to produce powerful and unsettling pareidolia effects; or Portrait of Edmond de Belamy, an artwork created by the Paris-based collective Obvious using a GAN, which sold for $432,500 at a Christie's auction in 2018.

A few years later, in 2020, Jonathan Ho and colleagues introduced another algorithmic architecture for image generation: Diffusion Models (DMs). These models train an algorithm to recognize an initial image, progressively adding "noise" until the image is entirely unrecognizable, and subsequently removing this noise step-by-step until returning to the original image. In practice, similar to GANs, DMs can generate original images not contained in the original training corpus, thus converting algorithmic errors into creative opportunities; moreover, this process of generating original images can be conditioned or guided using text-to-image prompts. Compared to ConvNets and GANs, DMs produce more accurate and detailed images; nevertheless, these two processes are not mutually exclusive: a specific ConvNet architecture, called U-Net, is commonly employed to guide the denoising and image-generation procedures in DMs.

In the meantime, a pivotal development was occurring in another field of artificial intelligence – language data processing (Large Language Models, LLM; Natural Language Processing, NLP). In 2017, a group of Google engineers introduced the Transformer architecture for the "understanding" (encoding) and "generation" (decoding) of verbal statements. The operational principles behind Transformers build on a series of ideas that are not entirely new:

> The roots of the model lie in the 1950s when two big ideas converged: Charles E. Osgood's idea […] to use a point in three-dimensional space to represent the connotation of a word, and the proposal by linguists such as Martin Joos, Zelig Harris, and [Simon] Firth to define the meaning of a word by its distribution in language use, meaning its neighbouring words or grammatical environments. Their idea was that two words that occur in very similar distributions (whose neighbouring words are similar) have similar meanings. (Jurafsky & Martin, 2025, p. 105)

In essence, according to the vector distribution semantics approach, it is possible to represent the meaning of any linguistic term mathematically and geometrically as a series of vectors indicating proximity or distance from all other terms. This proximity or distance does not stem from how speakers' linguistic knowledge is structured, but rather from the concrete recurrences and co-occurrences of individual terms within the corpora of linguistic productions used to train the algorithms – that is, to construct the vector space.

However, Transformers do not merely replicate the procedures of earlier LLMs; they represent a significant leap forward. Three main reasons account for this advancement. The first and most obvious is the introduction of attention, self-attention, and cross-attention mechanisms, which enable the algorithm to identify key terms. It considers their co-occurrence and position within the utterance analysed or produced. It is no coincidence that the paper introducing this architecture stated, paraphrasing the Beatles, "Attention is all you need" (Vaswani et al., 2023). The second reason is that algorithmic processes no longer apply to entire words, but rather to parts of words: techniques such as Byte Pair Encoding (BPE) or WordPiece fragment the continuous flow of discourse into fixed-length segments called "tokens"; it is these tokens, not whole lexemes, that are inserted into the vector space (an operation known as "embedding"). Finally, the enormous amount of verbal material available on the web constitutes an unprecedented training ground, which (somewhat surprisingly) results in an exponential increase in effectiveness. In 2018, researchers at the OpenAI laboratory introduced the Generative Pre-trained Transformer (GPT) model, the first in a series capable of leveraging increasingly extensive training datasets.

Ultimately, the success of Transformers derives from two operations of reduction and two of implementation. The two reduction operations are, first, the fragmentation of verbal or visual discourse flow into elementary components – already performed by other VGenAI architectures, particularly GANs; second, the elimination of any semantic extra-linguistic reference from these components, retaining only their pure materiality. This latter point is particularly significant, as it connects Transformers to the broader epistemic and technological project of automatic information processing, which entails a semantic deprivation of languages and discourses, and was inaugurated by the advent of the mathematical theory of communication in the 1940s (Geoghegan, 2023). As for the two implementation operations, the first is that these reduction processes allow the construction of a broader and more complex network of relationships between components based on the principle of co-occurrence. Secondly, the resulting algorithmic machine, when trained on extensive discourse corpora, becomes capable of comprehending and generating discourse with

extraordinary competence – thus making it a perfect delegated enunciation machine. We will examine the critical implications of these aspects in greater detail later.

So far, we have observed Transformers operating primarily in the domain of verbal language. However, in recent years, they have also demonstrated effectiveness in understanding and generating multimodal discourse – not only verbal texts, but also iconic, musical, and audiovisual ones. In this context, these models are referred to as Multimodal Large Language Models, Large Vision-Language Models, or even Foundation Models. In some cases, these models directly apply the principle of correlation between segments extracted from images (now termed "patches" rather than tokens – or "sound frames" in the case of audio clips), based on their probability of co-occurrence. These models are known as Autoregressive Transformers (such as DALL-E 1 by OpenAI, 2021, or Parti by Google, 2022). However, beginning around 2022–2023, the dominant trend has shifted toward combining a Transformer with a Diffusion Model, as this architecture proves significantly faster and more reliable. This phenomenon brings us back to the initial point of this paragraph: the architecture of Imagen 3, introduced, as we have already noted, in 2022. This is not an isolated instance: other platforms such as Stable Diffusion, DALL-E 3 (both introduced in 2023), and likely Midjourney (introduced in 2022 but whose architecture has not been publicly disclosed) adopt similar solutions. The subsequent development of Imagen 4 (2025) also adopts and further refines this structure. In essence, Transformers in these cases do not simply decode the user's instruction prompt but reformulate it into a highly detailed description, occasionally presented as a low-resolution image. Furthermore, from this point onward, they strongly guide and constrain and condition the Cascading Diffusion Models – the mechanism that produces images by generating versions at progressively higher resolutions (upsampling). Consequently, the action of Transformers is a driving force that permeates the entire image-generation process. In short, Transformers write the script (in accordance with the instructions encoded in the user's prompt) and ensure it is followed, while Diffusion Models shoot the movie.

8.3.3 The Many Faces of Latent Space

The situation presented above allows me to revisit and discuss the main aspects of the semiotic debate concerning the enunciation of and within algorithmic machines, as described in sect. 8.3.1. I will begin with point (a), which addresses whether it is appropriate to speak of "enunciation" in these cases. In my view, Marion Colas-Blais's position (2025) is not entirely sustainable. Indeed, while this claim remains debatable and would require further elaboration, I suggest that generative AI demonstrates not only operational competence, and thus a form of functional agency, but also a goal-directed orientation that may be described as volition-like; indeed, it produces structured verbal, visual, or auditory outputs that conform to content- and style-related constraints articulated, to varying degrees of explicitness, by the user in the prompt. At the same time, however, I am also not entirely convinced by the concept

of co-enunciation proposed by D'Armenio et al. (2024), as it positions human and mechanical subjects on the same ontological and operational level, thereby risking a return to the "anthropomorphic" enunciation criticized by Metz (2015). Instead, I find it helpful to explore further another point raised by Colas-Blais (2025). In transitioning from the first to the second phase of the machine concatenation, the human user *delegates* the enunciative function to the machine. Through a disengagement (*débrayage*), the machine is assigned the enunciative task according to instructions from the human subject, so becoming their "*ex-nuncius*". Therefore, I believe it is appropriate (following Latour's idea) to speak of *delegated enunciation*. Whether this delegation occurs through a discursive form of dialogue and collaboration between the user and the platform is a separate issue, relating to the underexplored area of interface interactions, which I will address in the next section.

Next, I will jointly address points (b) and (c) of the debate. These, as previously stated, pertain to the form of latent space and the processes involved in its construction during the training phase, as well as its use in generative processes within VGenAI. Initially, I observe that the currently dominant collaborative model between Transformers and Diffusion Models effectively illustrates the application of Fontanille's theory (2006) to VGenAI, as proposed by Dondero et al. (2025). Beginning from the *virtual* condition of latent space, elements are first *actualized* by Transformers and subsequently *realized* by Diffusion Models. Conversely, as Dondero suggests, machine training involves transforming *real* expressions into *potential* elements (segments of visual or verbal discourse of varying lengths) and then *virtualizing* them by embedding them into latent space. However, based on this framework, the current state of the art, as described previously, requires a fundamental clarification: *not all latent spaces are structured identically.* Consequently, not all processes of algorithm training and verbal or visual discourse generation operate similarly. Specifically, *the organization of latent space and generative procedures differ substantially between Transformers and Diffusion Models*. As a result, contemporary VGenAI systems use and connect two distinct artificial "neural circuits," with one (that of Transformers) dominating the other. I intend to demonstrate that *the claims made by the authors cited in sect. 8.3.1 regarding latent space configuration are valid for GANs and Diffusion Models, but do not apply to Transformers.*

As previously noted following Somaini (2025), a latent space, in general terms, is a mathematical domain within which input information – learned by a neural network through machine learning – is stored and structured in the compressed form of variables that are not immediately observable. In this way, the data can be manipulated more easily to generate new outputs. The components responsible for this transformation-reduction are called encoders or (variational) autoencoders. Thus, latent spaces act as a "bottleneck" between the input data processed by the encoder and the output data generated or reconstructed by the decoder. Latent spaces first emerged within VGenAI concerning GANs (Carter & Nielsen, 2017). In this context, individual points in space correspond to minimal variations of the same image; the actualization process thus involves traversing the *space* to locate the image or image component most suited to the prompt. Diffusion Models differ slightly: here, the latent space consists of varying levels of noise that progressively blur the image (while

remaining potentially reconfigurable). The enunciative practice, therefore, involves *temporal* traversal and the selection of variants (Liu et al., 2019). However, in both cases, machine training associates images or image components with descriptive tags. Consequently, the latent space structure in GANs and Diffusion Models is "biplanar," since each visual element, however minimal, corresponds to a semantic description. In this context, I find it appropriate to compare latent space to an archive as an operational space for manipulation and reconfiguration, as suggested by D'Armenio et al. (2024) and others.

However, the situation changes radically with the introduction of Transformers. Here, the transition from the *real* to the *potential* and subsequently *virtual* condition is markedly different. Autoencoders first fragment the discourse used for learning into distinct particles devoid of independent meaning (tokens or patches: for convenience, we will refer only to the first ones from here on). Then, the algorithm inserts these tokens into latent space (embedding procedure) through three interconnected processes: evaluating the co-occurrence of tokens in the same sentence (attention, self-attention, cross-attention); assessing the probability of co-occurrence between each token and others; and translating this evaluation into values expressed by vectors indicating the proximity or distance among tokens based on co-occurrence probabilities. As training progresses, these vectors become increasingly precise reflections of the training corpus, displaying continuous dynamism. The generative enunciative practice now involves progressively aggregating tokens – under continuous attention-based monitoring – that probabilistically co-occur, guided by the user's prompt with its specific instructions, constraints, and conditions (Mickus et al., 2022). Consequently, Transformers' latent vector space is no longer biplanar but "monoplanar" (Paolucci, 2025): it comprises meaningful elements (tokens or patches) whose values do not depend on any semantic reference but solely on the probability of co-occurrence with other similar components. Therefore, I find the archive analogy proposed by D'Armenio and Dondero less fitting in this case, primarily due to the fragmentary and a-semantic nature of the archive's components. Additionally, this process effectively explains VGenAI's tendency to produce stylistically coherent images, as noted by Dondero et al. (2025), and its ability to handle what cognitive science refers to as frames (or schemas, or scripts) – that is, "a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party" (Minsky, 1975, p. 211). However, this capability does not imply semantic comprehension by Transformers in a traditional sense; instead, in both formal coherence (style) and content consistency (frames) – as well as in their combination (Pinotti, forthcoming) – Transformers operate according to a logic of monoplanar combination of discursive segments.

8.3.4 The Performance of Incompetence

I now turn to point (d) identified at the conclusion of sect. 8.3.1. As previously noted, the semiotic debate on VGenAIs has primarily concentrated on enunciative

practices, while tending to overlook the specific enunciational forms that emerge from user interactions with VGenAI systems within interfaces – though there are some notable exceptions, such as Reyes (2024). In what follows, I aim to partially address this gap by analyzing a typical exchange between a user and a generative platform.

As is widely recognized, VGenAIs – and GenAI systems more broadly – can be accessed through various types of interfaces, including, for instance, traditional search engines. However, the format that has most significantly contributed to their widespread adoption is chat simulation, an interaction model with historical roots extending back at least to Turing's machine (Natale, 2021). It is precisely this simulated conversation that unfolds when querying a GenAI system such as ChatGPT or the related VGenAI platform DALL·E.

At the time of writing (July 2025), the platform's homepage featured a minimalist design dominated by white space, with the platform's name prominently displayed. Upon logging in and being recognized by the system, this text is replaced with the prompt "What can I help with?", while previous queries appear in a sidebar to the left, and the model version currently in use is displayed at the top. At the centre is a prompt field inviting input, marked by the phrase "Ask anything." Beneath this field, additional affordances offer preformatted interaction options such as "Surprise me," "Brainstorm," or "Analyze images." (When logged in, these are replaced by a more generalized "tools" menu.) To test the system, I submit a query relevant to this very article: "Summarize the current debate on the semiotics of enunciation." Upon entering the request and initiating the prompt, the interface scrolls rapidly downward, shifting my query into a speech-bubble-like box on the right, outlined in light grey. This repositioning opens a new white space in the centre of the screen, where a pulsating black dot appears, soon followed by the message "Searching the web." After approximately 3 s, small icons referencing the consulted websites appear on the left. A second later, the message "Here's a refined summary of the current debate on the semiotics of enunciation" is displayed. (This entire intermediate stage is bypassed if one is logged in.) Next, the answer is rapidly generated on-screen, presented one line at a time, and organized into bullet points. When logged in, however, the output is rendered differently – via a rapid, letter-by-letter animation. It is worth noting that the responses differ significantly between logged-out and logged-in modes, as the latter incorporates data on the user's prior interactions, which are stored by the system.

Should I now request a conceptual diagram or image (a feature available only when logged in), the marginalization process resumes: my request is again moved to a grey balloon on the right, the black dot reappears, and the message "getting started" is shown. Within about 2 s, a light grey box appears, bearing an icon of two tiny, stylized mountains with alternating sun and moon motifs, accompanied by a circular progress indicator. After roughly 15 s, the icon disappears, and the box's background darkens, revealing indistinct shapes as if seen through frosted glass. A few seconds later, a diagonal highlight passes across the box. Approximately 22 s after the box darkens, the top edge of the generated image begins to appear; simultaneously, "getting started" is replaced by "Creating image. May take a moment." The image

renders progressively from top to bottom and is fully displayed within 24 s. Beneath the image appears the message: "Here is the diagram you requested. Download high-resolution image." The latter part is underlined in blue and initiates the download when clicked.

This basic description sets the stage for four observations. First, what we observe here is a paradigmatic case of "enunciated enunciation": the screen displays not only the inscription of actions but their ongoing performance by both human and machine actors. These actions are narrated in real time, while also offering the user a set of affordances – additional actions they may take, which are in turn displayed and narrativized.

Second, this interaction entails a clearly defined distribution of roles. Here, the tools of the semiotics of enunciation prove particularly helpful – especially those developed by Greimas, who, as you may recall, defines enunciation as the process of bringing deep structures into discourse (discursivization) through three operations: the definition of subjects (actorialization), spaces (spatialization), and times (temporalization). In this case, we see an actorialization of the two participants in the exchange: a requester/addressee and a responder/addresser/donor. These are configured according to two thematic roles – namely, the incompetent (who does not know and cannot do) and the competent (who knows, can act, and crucially, can enunciate). The distinction between these roles is underscored through both spatialization and temporalization strategies. Spatially, the interface marginalizes the user's utterance (via the side-positioned balloon), clearing space for the central, expansive stage on which the VGenAI's triumphal enunciative performance unfolds. Temporally, the respondent's activity is imbued with suspense, anticipation, and dynamism, while the user's contribution recedes into an indistinct, temporally anonymous background. Consider the symbolic potency of the pulsating dot – an index of silent thought or computation; the energetic sequence of animated writing that dramatizes the system's performance; the temporal metaphor of alternating day and night as the image is rendered; and the slow emergence of visibility through the veil of the darkened screen. In short, this is a carefully choreographed performance of VGenAI's enunciative prowess in response to the user's modal incompetence – despite any actual expertise possessed by the user at an empirical level.

Third, what had until now been described in "impersonal" terms (following Paolucci, 2021) is now assigned to concrete agents. The user comes to recognize the requester/addressee as their proxy, mainly because their sensorimotor activity – moving the mouse, typing – is mirrored on screen. Despite the user's awkwardness, latency, and spatial marginalization, this perceptual alignment creates identification. In contrast, the partner – the VGenAI – emerges as swift, fluent, and rhetorically dominant. This distribution and recognition of roles has a significant consequence that connects back to earlier points. As Treleani (2019, p. 219) has observed, "digital interfaces shape our belief systems through a discursive framing of content to which they give access." In this case, the interface also shapes the user's experience of action. If we accept that, at the level of enunciative practices, user prompts correspond to a delegation of discursive agency to the GenAI system, it is equally crucial to observe that this delegation is not presented as such. Rather, the interface frames

it as an interaction in which the human subject acknowledges their lack of competence and entrusts themselves to the machine's hypercompetence – establishing the system as a trustworthy surrogate, whose labour is made legible through temporally and spatially dramatized discourse.

Fourth and finally, these observations may seem to contradict my earlier discussion: there, I analysed *real* enunciative practices; here, I have examined *imaginary* enunciative dynamics. But the distinction is not so stark. Consider the different responses generated by ChatGPT when I am logged in versus when I am not. In both cases, the system applies the same procedures: it decomposes user utterances into tokens, patches, or sound frames; embeds them into latent vector spaces shaped by training; and extracts features that can enhance the personalization of responses. Until October 2023, these utterances were also used for training the models; today, that only happens with the user's consent (via the "Chat history & training" setting). Still, the resulting data are stored in short- and long-term memory banks, the management of which is currently a significant area of research (see Pan et al., 2025). In sum, when a human user delegates enunciative labour to the algorithmic machine through interface interaction, they place their utterances at the service of machine enunciation – submitting them to processes of fragmentation and data extraction.

8.4 The Dividual Enunciation

8.4.1 A Quick Recap

Before proceeding to my concluding remarks, it is helpful to summarize the findings presented in Sect. 8.3. There, we examined how enunciation operates within the domain of Visual Generative Artificial Intelligence (VGenAI) on two distinct but interrelated levels: enunciative practices and enunciational dynamics – the same two dimensions that have shaped the theoretical articulation of the concept from the outset (see Sect. 8.2).

On the one hand, in terms of enunciative practices, I proposed interpreting the algorithmic generation of images (and, by extension, of any type of discourse) as a case of enunciation *delegated* by human operators to a sequence of algorithmic machines. In analysing the internal structure of this chain, I observed that architectures derived from Large Language Models – most notably, Transformers – now play a central role in the actual generation of images, even in processes primarily associated with Diffusion Models. More specifically, Transformers enable the virtual dimension of enunciation, while Diffusion Models bring this into the actual. The operational mode and logic of Transformers thus become particularly relevant. Drawing from a longstanding tradition in computational linguistics (itself rooted in the development of mathematical communication theory), Transformers enact a dual reduction of discursive utterances: first, by fragmenting them into segments – whether verbal, visual, or auditory – of arbitrary length; and second, by stripping them of any inherent

semantic content. The criteria for assigning value to these segments in latent space, as well as those governing their retrieval and recombination during the generation of new utterances, follow exclusively the logic of co-occurrence probability (whether adjacency or intra-sentential), as guided and constrained by the user's input (e.g., stylistic parameters).

On the other hand, at the level of enunciational dynamics and forms, my analysis of GenAI interfaces as experienced by users has revealed how actorialization, spatialization, and temporalization function to disguise the process of delegated enunciation. Instead, these operations construct an enunciative scene in which the machine's hypercompetence is staged against the user's incompetence, producing a distinct dramatic tension. I also noted that the two levels – practices and dynamics – are strictly connected: this connection is not only due to the fact that the dynamics of the interface structure and narrativize the phases of actualization and realization of the machine's enunciative performance; but it also refers to the fact that the user, by participating in the dramaturgical script proposed by the interface and translating their delegation into requests, provides the machine with new discursive materials. These are then subject to datafication, processing, and various downstream uses – ranging from personalized response generation to further model training. In other words, the role-playing encouraged by the interface through its enunciational design incorporates the user's utterances into the broader enunciative concatenation that defines the functioning of algorithmic machines.

In this concluding section, I suggest interpreting these processes through the conceptual framework of the "dividual" and "dividuation." I argue that bringing the semiotic debate on VGenAI into dialogue with selected theoretical perspectives on dividuality produces a twofold contribution. First, it clarifies the procedures and operative logic underlying algorithmic image generation. Second, it provides enunciation theory with new analytical leverage for addressing some of the limitations and challenges that currently shape its disciplinary discussion.

8.4.2 From Individuation to Dividuation, from Dividuation to Quasi-Individuation

The concepts of dividual and dividuation – along with many related compound terms that I will briefly reference – are currently at the centre of a complex and multifaceted debate, which unfolds along four principal trajectories. First, there is the philosophical genealogy of the term *dividuum*, with particular attention to historical questions such as the medieval theological debates surrounding the Trinity and the modern philosophical discourse on personal identity (Raunig, 2016, pp. 25–70; Botti et al., 2024). Second, the concept finds anthropological application, especially in the analysis of religious or magical systems that emphasize fragmentation and distribution of the subject. These systems give rise to models of identity that diverge significantly from the Western ideal of the autonomous individual (Linkebach & Muslow, 2019;

Rohatynskyj, 2015). Third, the concept of dividuation has influenced contemporary aesthetic theory, particularly concerning artists and art practices that foreground themes such as molecularity, repetition, and dispersion (Bodini, 2024; Ott, 2018, pp. 229–248). Finally – and this is the dimension on which I will concentrate – the term has gained relevance in discussions concerning the economic and political dimensions of contemporary advanced capitalism and its technological apparatuses.

At the origin of this economic and political interpretive trend lies a short but seminal essay by Gilles Deleuze, *Postscript on the Societies of Control*, first published in English in 1992 (Deleuze, 1992). In this text, Deleuze observes that the model of *disciplinary* societies described by Michel Foucault – which dominated the eighteenth and nineteenth centuries – was gradually supplanted in the twentieth century by a new configuration: the societies of *control*. In disciplinary societies, subjectivation was achieved through the allocation of fixed roles within enclosed institutional spaces – the family, school, military barracks, factory, hospital, or prison. The emergence of the control society eroded these rigid structures, since social agents are now ostensibly "free" to move within open systems; however, subjectivation has not disappeared, but rather has mutated: it now operates through the tracking and modulation of individual trajectories via the tools of digital information technology: "Enclosures are *molds*, distinct castings, but controls are a *modulation*, like a self-deforming cast that will continuously change from one moment to the next, or like a sieve whose mesh will transmute from point to point" (1992, p. 4). In this framework, while disciplinary societies aimed to produce an *individual* – that is, a discrete, indivisible unit within a social mass – control societies give rise to the *dividual*: a decomposable and recombinable component, divisible into units of data stored, processed, and reassembled across multiple digital archives.

> Disciplinary societies have two poles: the signature that designates the individual, and the number or administrative numbering that indicates his or her position within a mass […] In control societies, on the other hand, what is essential is no longer either a signature or a number, but a code […] The numerical language of control is made up of codes that mark access to information, or reject it. We no longer find ourselves dealing with the mass/individual pair. Individuals have become "dividuals," and masses, samples, data, markets, or "banks." (5)

Deleuze only partially develops his insights; moreover, other instances complicate a comprehensive interpretation of the essay in question in his work, where the term "dividual" is used with partially divergent meanings (Bastidas, 2023; Ott, 2018, pp. 123–143). Nonetheless, scholars have returned to this short but seminal text in a variety of disciplinary contexts. I will focus here on three in particular.

The first and perhaps most unexpected area of renewed interest in Deleuze's, 1992 essay emerges within marketing and business communication studies, which have since increasingly adopted a critical stance toward capitalism. In the early 2000s, it became evident that the rise of big data economies, data-driven marketing, and "profiling machines" (Elmer, 2004) was profoundly transforming sales strategies. As Zwick and Denegri-Knott observe, "new forms of database marketing are … customer production processes that rely on the exploitation of the multitude of

consumer life […] More specifically, […] database marketers collapse the production–consumption dichotomy by *manufacturing customers as commodities*" (2009, p. 321). In this context, marketing had to reconceptualize the idea of consumer, shifting from demographic segmentation – based on variables like gender, race, age, and socioeconomic status – toward a data-centric logic. In this sense, consumers now present themselves as data clusters or "dividuals": that is, as fragmented data assemblages that can be exposed, dissected, and resegmented into new marketable categories (Cluley & Brown, 2015).

This reconceptualization has prompted significant theoretical developments. For instance, many scholars have highlighted the role of marketing in both social control and the dividuation of subjects: "While interpretative marketing may typically allude to a relatively unified subject on the level of representational meaning, the market increasingly operates according to the future potentials of amassing aggregates of dividual data and unceasing connective fragmentation that produces a subject always in the progress of becoming: searching relentlessly for an essence, a fixity, long since a present absence" (Hietanen et al., 2022, p. 174). In the same vein, a growing body of work has challenged the traditionally optimistic tone of consumer culture studies. One such strand, known as terminal marketing, responds to a prevailing sense of cultural stagnation and apolitical apathy: "an atmosphere of apolitical apathy where the future has increasingly been 'cancelled' and all that remains is a carnivalesque consumer culture that has resigned itself to extinction, even if on the semiotic surface it is increasingly ethical and ecological" (Ahlberg et al., 2022, p. 667; see also Hjelm, 2025).

A second domain in which Deleuze's, 1992 insights have been revisited lies in the economic and anthropological analysis of contemporary financial capitalism. A foundational text here is a paper by Arjun Appadurai, later included in one of his books (Appadurai, 2015, pp. 101–124). Appadurai examines the 2007–2008 financial crisis, locating its origins in the breakdown of contractual trust – previously, a key mechanism for managing risk in economic, social, and political relations. At the heart of this crisis lies the widespread use of financial derivatives, particularly mortgage-backed securities: "the failure of the derivatives market (especially in the domain of housing mortgages) is primarily about failed promises […], a type of failure that was neither occasional nor ad hoc but became systematic and contagious, thus bringing the entire financial market to the brink of disaster" (2). This collapse stems from the joint operation of two technologies – financial and digital. On the financial side, derivatives manipulate existing contracts (mainly mortgages) by breaking them into divisible, recombinable components: "The bizarre materiality of the mortgage-backed American house is that while its visible material form is relatively fixed, bounded, and indivisible, its financial form, the mortgage, has now been structured to be endlessly divisible, recombinable, saleable, and leverageable for financial speculators, in a manner that is both mysterious and toxic" (61).

On the digital side, these instruments rely on the datafication of consumers; as Deleuze anticipated, this involves a logic of partition and multiplication: "[The] new forms of data gathering and analysis […] atomize, partition, qualify, and quantify the individual to make highly particular features of the individual subject or actor more

important than the person as a whole" (109). Two shared principles underpin this joint operation. First, both finance and information technology function according to a logic of dividuation: they fragment a given subject or object into discrete elements that can be recombined into unstable aggregates. Second, this is a predatory form of dividuation, one that extracts value from the original subjects or objects, multiplies it through segmentation and recomposition, and reallocates it elsewhere: "the multiple derivative instruments that were developed by slicing and dicing individual mortgages so as to generate profit for financial institutions exemplify this predatory logic and have the effect of making dividualized actors incapable of any concerted critique, resistance, or reform with regard to these predatory logics" (110).

The third area of Deleuze's revival concerns the intersection with communication technologies. Particularly noteworthy is the work of Carbone and Lingua (2023), who argue that processes of dividuation begin even before full-fledged datafication, originating instead in the immediate entanglement of human bodies with the dispositives of machinic capitalism (see also Raunig, 2016, pp. 109–114). Devices such as smartwatches, smartglasses, biometric tattoos, and neural implants – alongside more everyday tools like smartphones and screens – form direct connections with bodily organs (eyes, skin, hands). These organs are thereby functionally separated from the unified body through dividuation and reconfigured into human-technological assemblages. In this sense, Carbone and Lingua revisit the concept of prosthesis – applying it to both organic and technological components – and redefine it as "quasi-prosthesis." The prefix "quasi" serves a double purpose. First, it denotes the temporary and contingent mode of existence of the assemblage components: "the prefix 'quasi' is […] intended to indicate not a constant property of the organs but their temporary mode of existence" (137). Second, it highlights how human-technical couplings (a key concept derived from Simondon, 1980) redistribute agency, granting the so-called passive object pole a distinctive role within the assemblage: "[in particular,] the use of the prefix 'quasi'… express[es] the peculiar agency exercised, within specific relationships, by the pole that is traditionally considered passive in that dualism: the object pole" (143).

Notably, the authors stress that dividuation is always paired with recomposition into new "individual" assemblages: "we believe that the relationship with digital technologies today requires us to think in terms of a constant integration and negotiation between dividual and individual, rather than starting from their preconceived opposition" (172). On this point, they align with other commentators, such as Bruno and Rodríguez (2023), who describe "a complex dividual-individual composition, focusing on biotechnologies, digital culture, and financial capitalism" (27). In my view, however, referring to these hybrid assemblages as "individuals" or invoking "individuality" is potentially misleading, since it risks obscuring the irreversibility of dividuation. Alternative terms like "con-dividual" (Raunig, 2016) or "con-dividuation" (Ott, 2018) have been proposed to designate collective formations that not only recognize but also mobilize the dividual condition – through solidarity and networking – as a response to the exploitative logics of market and finance. Along similar lines, I propose that the configurations described by Carbone

and Lingua might better be called *trans-dividual* or *co-dividual* – or more precisely, following the authors' suggestion, *quasi-individual*.

In conclusion, the debate outlined here opens to a general model of dividuation within the framework of an expanded political economy – that is, an account of how material and immaterial resources are managed both in terms of their subsistence and logistics and under the perspective of their ownership shifts within shared techno-social environments (Eugeni, forthcoming a, forthcoming b). In this view, dividuation can be considered as a technical procedure for extracting value from a given resource (e.g., a person's behaviour, consumption patterns, or loan contracts). The process entails two phases. The first, following Raunig, 2022, pp. 55–64), can be called "dissemblage" or "dividuation": it involves fragmenting a coherent unit ("individual") into a freely divisible surface ("dividual"), from which arbitrary segments can be extracted and recombined in a way that prevents any restoration of the original unit. Following up on our examples, this applies to social data, financial contracts, and so on. The second phase, which I propose to call "reassemblage" (after Casetti, 2015, pp. 67–98) or "quasi-individuation" (drawing on and reinterpreting as I said Carbone & Lingua, 2023), involves reassembling these segments into provisional composites with an appearance of coherence – whether for targeted advertising or for constructing marketable financial products. For these operations to be profitable, the market value of the quasi-individuals must exceed the value of the initial resources. It is important to note that, in light of the preceding discussion, reassemblage does not produce true individual entities; rather, it yields aggregates of dividual units whose apparent individuality is an effect of meaning. From this perspective, we may state that every reassemblage is, in fact, a form of resemblance.

While keeping in mind these two phases of dividuation (dissemblage) and quasi-individuation (reassemblage), I will henceforth refer to the machines responsible for performing these operations simply as "dividuation dispositives". Dividuation dispositives typically function autonomously once activated by those who design and own them. However, in many cases – such as wearable technologies – human users must actively participate, feeding data into the system while deriving individual benefits. This user participation ensures the smooth operation of the device and increases the value of its outputs. Moreover, the dividuation device not only extracts and multiplies value but also redistributes it. Not by chance, as Raunig notes (2016, pp. 25–38), the earliest Latin uses of "*dividuum facere*" in Plautus and Terence refer to the political-economic partitioning of resources, happily resolving narrative tensions through final redistributions. In the same vein, today's dividuation dispositives allocate a small portion of benefits to users (who buy products, wear devices, or take out loans), while reserving a far larger share for those who manage and profit from the infrastructure of datafication – data brokers, marketers, and financial institutions.

8.4.3 The Dividuation of Discourse

Readers who have followed the discussion this far will likely have inferred the direction of my argument. I propose *viewing the work of GenAI and VGenAI* – summarized in Sect. 8.4.1 – *as an application of the techniques of dividuation and quasi-individuation* discussed in Sect. 8.4.2 *to a specific economic-political resource: discourse.* In other words, if we consider the corpus of discourses (or "documents") circulating within a society as a stock of economic assets (Oakley, 2022; Ferraris, 2022; Eugeni, forthcoming b), we may argue that the processes of valorisation through dividuation and quasi-individuation analysed above with regard to goods and financial products are equally applicable to discursive formations; from this perspective, *GenAI can be ultimately understood as the dispositive that channels discourses into the operational logics of dissemblage and reassemblage that characterize algorithmic capitalism.*

From this conceptual vantage point, it becomes possible to reconsider the semiotic debate on enunciation that has emerged in relation to GenAI and VGenAI. In particular, such a perspective enables us to reinterpret the observations made in Sect. 8.4.1. The training phase of GenAI and VGenAI corresponds, within Fontanille's (2006) model of enunciative practices, to the descending phase in which *real* utterances are transformed into *potential*, and then *virtual*, forms of existence. We can now frame this transition – particularly in the case of Transformer architectures – as a *dividuation or dissemblage of discourse*: *potentialization* appears as the segmentation of the discursive continuum into discrete tokens or visual patches, while *virtualization* entails embedding these segments into a vector space by assigning them multi-vector values within the latent space. Conversely, the production of discourse by users corresponds to the ascending phase of Fontanille's model: the transition from *virtual* to *actual* and then to *real* existence. Here, too, we can reinterpret the process through the lens of dividuation techniques: tokens and patches move from an abstract state to a specific aggregation (*actualization* carried out by Transformers), which subsequently *materializes* into an image that should be considered as a quasi-individual reassemblage (realization performed by Diffusion Models). In delegating their enunciative agency to algorithmic machines via their enunciative gestures, human operators become integrated into the dispositive itself, placing their utterances in the service of new processes of dividuation and quasi-individuation.

I propose to speak in these cases of a *dividual enunciation.* As is evident, such a conception requires a radical rethinking of the idea of enunciation. Indeed, in Sect. 8.2, I emphasized that the concept of enunciation has historically developed in close relation to processes of *individuation*, particularly to the reciprocal individuation of utterances and the subjects who produce and interpret them (enunciators and enunciatees). The analysis presented in this paper – focused on the enunciative and enunciational operations of VGenAI – confronts us with an alternative procedure, one that entails both the dividuation of discourses and utterances and the simultaneous dividuation of the subjects responsible for their emergence (with all the associated issues, including those related to copyright). This situation opens two possible lines

of interpretation. One option is to accept, with Ott (2018, p. 8), that "the concept 'individual' has never been adequate," and to view GenAI and VGenAI as simply forcing us to confront the true nature of our enunciative behaviour – an idea that seems to be supported by Paolucci (2025). The other path is to propose that algorithmic machines are inaugurating a new type of enunciative practice and a new set of enunciational dynamics, grounded not in individuation but in dividuation. In this light, the growing reliance on enunciative delegation to generative systems may carry profound consequences for our understanding of discursive production – challenging its traditional framing as a shared interpretive practice through which subjects construct meaning and make sense of the world.

References

Ahlberg, O., Coffin, J., & Hietanen, J. (2022). Bleak signs of our times: Descent into 'Terminal Marketing'. *Marketing Theory, 22*(4), 667–688. https://doi.org/10.1177/14705931221095604

Appadurai, A. (2015). *Banking on words. The failure of language in the age of derivative finance.* The University of Chicago Press. https://doi.org/10.7208/9780226318806

Bastidas, D. (2023). *The dividual remainder. For a Deleuzian history of dividuality (Bergson, Spinoza, Simondon).* Paper presented at the 15th International Deleuze and Guattari Studies Conference and Camp, Ian Buchanan, Belgrade, Serbia. https://theses.hal.science/SPH/hal-04195057v1

Benveniste, E. (2014). The formal apparatus of enunciation. In J. Angermuller, D. Maingueneau, & R. Wodak (Eds.), *The discourse studies reader. Main currents in theory and analysis* (pp. 140–145). John Benjamins Publishing Company. https://doi.org/10.1075/z.184.32ben (Original work published 1970)

Bie, F., et al. (2025). RenAIssance: A survey into AI text-to-image generation in the era of large model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 47*(3), 2212–2231. https://doi.org/10.1109/TPAMI.2024.3522305

Bodini, J. (2024). Dividuation d'artiste. Déconstruction echo-logiqued'un mode de subjectivation modern. In G. Caignard, C. Orlando, & M. Roche (Eds.), *Explorer le "dividuel" / Exploring the "dividual" / Esplorare il "dividuale"*. Special section of *Chiasmi International, 26*, 51–76. https://doi.org/10.5840/chiasmi2024268.

Botti, C., Quarantotto, D., Stimilli, E., & Valente, L. (2024). *Condividuale. Genealogie di un concetto mancato.* Quodlibet.

Bruno, F., & Rodríguez, P. M. (2023). The dividual: Digital practices and biotechnologies. *Theory, Culture & Society, 39*(3), 27–50. https://doi.org/10.1177/02632764211029356

Carbone, M., & Lingua, G. (2023). *Toward an anthropology of screens. Showing and hiding, exposing and protecting*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-30816-1

Carter, S., & Nielsen, M. (2017, December 4). Using artificial intelligence to augment human intelligence. *Distill*. https://doi.org/10.23915/distill.00009

Casetti, F. (1998). *Inside the gaze: The fiction film and its spectator* (N. Andrew & C. O'Brien, Trans.). Indiana University Press. (Original work published 1986)

Casetti, F. (2015). *The Lumière galaxy. Seven key words for the cinema to come.* Columbia University Press. https://doi.org/10.7312/columbia/9780231172431.001.0001

Cluley, R., & Brown, S. D. (2015). The dividualised consumer: Sketching the new mask of the consumer. *Journal of Marketing Management, 31*(1–2), 107–122. https://doi.org/10.1080/0267257X.2014.958518

Colas-Blais, M. (2025). La machine crée, mais énonce-t-elle? Le computationnel et le digital mis en débat. In Dondero, Aldama & Leone, pp. 147–187. https://doi.org/10.1515/sem-2024-0188.

D'Armenio, E., Deliège A., & Dondero, M. G. (2024). Semiotics of machinic co-enunciation. About generative models (Midjourney and DALL·E). *Signata. Annales des sémiotiques/Annals of Semiotics*, *15*. https://doi.org/10.4000/127x4.

Deleuze, G. (1992, October). *Postscript on the societies of control* (Vol. 59, pp. 3–7). JSTOR. http://www.jstor.org/stable/778828 (Original work published 1990)

Dondero, M. G. (2020). *The language of images. The forms and the forces.* Springer. https://doi.org/10.1007/978-3-030-52620-7

Dondero, M. G. (2025). Semiotics of artificial intelligence: Enunciative praxis in image analysis and generation. *Semiotica, 2025*(262), 111–146. https://doi.org/10.1515/sem-2024-0195

Dondero, M. G., Aldama, J. A., & Leone, M. (Eds.). (2025). Aspects of AI semiotics: Enunciation, agency, and creativity. *Special Issue of Semiotica, 262*. https://doi.org/10.1515/sem-2025-0014

Elmer, G. (2004). *Profiling machines. Mapping the personal information economy*. The MIT Press. https://doi.org/10.7551/mitpress/5614.001.0001

Eugeni, R. (Forthcoming b). *Algorithmic images. The new political economy of light*. Routledge.

Eugeni, R. (2026a). Algorithmic economies of light. The SARS-Cov-2 micrographs as algo-images. In P. Conte, A. C. Dalmasso, M. G. Dondero, & A. Pinotti (Eds.), *Algomedia: The image at the time of artificial intelligence* (pp. 117–129). Springer. https://doi.org/10.1007/978-3-032-08726-3_7

Ferraris, M. (2022). *Doc-humanity* (S. De Sanctis, Trans.). Mohr Siebeck. https://doi.org/10.1628/978-3-16-161667-9 (Original work published 2021)

Fontanille, J. (2006). *Semiotics of discourse* (H. Bostic, Trans.). Peter Lang. https://doi.org/10.3726/978-1-4539-0954-6/10 (Original work published 1998)

Genette, G. (1980). *Narrative discourse. An essay in method* (J. E. Lewin, Trans.). Cornell University Press. (Original work published 1972)

Geoghegan, B. D. (2023). *Code. From information theory to French theory*. Duke University Press. https://doi.org/10.2307/j.ctv3006zjr

Greimas, A. J., & Courtés, J. (1982). *Semiotics and language. An analytical dictionary* (L. Crist et al., Trans.). Indiana University Press. (Original work published 1979)

Hietanen, J., Ahlberg, O., & Botez, A. (2022). The 'dividual' in semiocapitalist consumer culture. *Journal of Marketing Management, 38*(1–2), 165–181. https://doi.org/10.1080/0267257X.2022.2036519

Hjelm, M. (2025). Phenomenology of a dividual. *Consumption Markets & Culture*, 1–15. https://doi.org/10.1080/10253866.2025.2502385

Jurafsky, D., & Martin, J. H. (2025, January 12). *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition with language models. Third edition draft*. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Latour, B. (1998). Piccola filosofia dell'enunciazione. In P. Basso & L. Corrain (Eds.), *Eloquio del senso. Dialoghi semiotici per Paolo Fabbri* (pp. 71–94). Costa & Nolan.

Latour, B. (2013). *An inquiry into modes of existence. An anthropology of the moderns* (C. Porter, Trans.). Harvard University Press. (Original work published 2012)

Li, J., Zhang, C., Zhu, W., & Ren, Y. (2025). A comprehensive survey of image generation models based on deep learning. *Annals of Data Science, 12*(1), 141–170. https://doi.org/10.1007/s40745-024-00544-1

Linkenbach, A., & Mulsow, M. (2019). Introduction: The dividual self. In M. Fuchs et al. (Eds.), *Religious individualisation. Historical dimensions and comparative perspectives* (2 vols., pp. 323–343). De Gruyter. https://doi.org/10.1515/9783110580853-015.

Liu, Y., Jun, E., Li, Q., & Heer, J. (2019). Latent space cartography: Visual analysis of vector space embeddings. *Computer Graphics Forum, 38*(3), 67–79. https://doi.org/10.1111/cgf.13672

Metz, C. (1974). *Language and Cinema* (D. J. Umiker-Sebeok Trans.). Mouton. https://doi.org/10.1515/9783110816044 (Original work published 1971)

Metz, C. (2015). *Impersonal enunciation, or the place of film* (C. Deane, Trans.). Columbia University Press. https://doi.org/10.7312/columbia/9780231173674.001.0001 (Original work published 1991)

Mickus, T., Paperno, D., & Constant, M. (2022). How to dissect a muppet: The structure of transformer embedding spaces. *Transactions of the Association for Computational Linguistics, 10*, 981–996. https://doi.org/10.1162/tacl.a.00501

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). McGraw-Hill. (Original work published 1974)

Natale, S. (2021). *Deceitful media artificial intelligence and social life after the turing test*. Oxford University Press. https://doi.org/10.1093/oso/9780190080365.001.0001

Oakley, T. (2022). Semiotics in economics and finance. In J. Pelkey & S. Walsh Matthews (Eds.), *Bloomsbury semiotics Vol. 2 semiotics in the natural and technical sciences* (pp. 239–257). Bloomsbury. https://doi.org/10.5040/9781350139350.ch-10

Odin, R. (2022). *The spaces of communication* (Ciaran O'Faolain Trans.). Amsterdam University Press. https://doi.org/10.5117/9789462987142 (Original work published 2011)

Ott, M. (2018). *Dividuations. Theories of participation* (A. Kirkland Trans.). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-72014-2

Pan, Z., et al. (2025). On memory construction and retrieval for personalized conversational agents. Preprint retrieved from *arXiv*:2502.05589v3. https://doi.org/10.48550/arXiv.2502.05589

Paolucci, C. (2021). *Cognitive semiotics Integrating signs, minds, meaning and cognition*. Springer. https://doi.org/10.1007/978-3-030-42986-7

Paolucci, C. (2025). The myth of meaning: Generative AI as language-endowed machines and the machinic essence of the human being. *Semiotica, 2025*(262), 5–23. https://doi.org/10.1515/sem-2024-0204

Peverini, P. (2024). *Bruno Latour in the semiotic turn. An inquiry into the networks of meaning*. Springer. https://doi.org/10.1007/978-3-031-57178-7

Pinotti, A. (2026). Artificial intelligence, VR, and the virtual. In P. Conte, A. C. Dalmasso, M. G. Dondero, & A. Pinotti (Eds.), *Algomedia: The image at the time of artificial intelligence* (pp. 83–103). Springer. https://doi.org/10.1007/978-3-032-08726-3_5

Raunig, G. (2016). *Dividuum. Machinic capitalism and molecular revolution*, Vol. 1 (Aileen Derieg Trans.). Semiotext(e). (Original work published 2015)

Raunig, G. (2022). *Dissemblage: Machinic captialism and molecular revolution*. Minor Compositions. (Original work published 2021)

Reyes, E. (2024). Characterizing AI in media software: An interdisciplinary approach to user interfaces. In V. Burgio & V. Manchia (Eds.), *Interfacce. Forme dell'accesso e dispositivi d'intermediazione*, Special issue of *Carte Semiotiche. Rivista Internazionale di Semiotica e Teoria dell'Immagine*. Annali 11, 34–47.

Rohatynskyj, M. (2015). Empowering the dividual. *Anthropological Theory, 15*(3), 317–337. https://doi.org/10.1177/1463499615570919

Rosen, P. (Ed.). (1986). *Narrative, apparatus, ideology. A film theory reader*. Columbia University Press.

Saharia C., et al. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Preprint. https://doi.org/10.48550/arXiv.2205.11487

Simondon, G. (1980). *On the mode of existence of technical objects* (Ninian Mellamphy, Trans.). University of Western Ontario. (Original work published 1958)

Somaini, A. (2023). Algorithmic images: Artificial intelligence and visual culture. *Grey Room, 93*, 74–115. https://doi.org/10.1162/grey_a_00383

Somaini, A. (2025). A theory of latent spaces. In P. Conte, A. C. Dalmasso, M. G. Dondero, & A. Pinotti (Eds.), *The world through AI. Exploring latent spaces* (pp. 21–51). Jeu de Pome – JBE Books.

Treleani, M. (2019). The degree zero of digital interfaces: A semiotics of audiovisual archives online. *Semiotica, 241*, 219–235. https://doi.org/10.1515/sem-2019-0043

Vaswani, A., et al., (2023). *Attention is all you need*. Version 7. https://doi.org/10.48550/arXiv.1706.03762 (Original work published 2017)

wick, D., & Denegri Knott, J. (2009). Manufacturing customers. The database as new means of production. *Journal of Consumer Culture, 9*(2), 221–247. https://doi.org/10.1177/1469540509104375