AI & SOCIETY
https://doi.org/10.1007/s00146-026-02948-4

REVIEW

Not a statistical accident: framing bias as a semiotic property of image datasets

Simone Fabbrizzi[1] · Symeon Papadopoulos[2] · Eirini Ntoutsi[3] · Ioannis Kompatsiaris[2]

Received: 7 November 2025 / Accepted: 16 February 2026

Abstract
Despite being a well-studied problem, bias continues to affect modern Computer Vision (CV) datasets, models and systems, including Generative AI models that have been found to replicate and amplify harmful biases and stereotypes. In this work, we focus on *framing bias*: a type of bias that arises from the way images convey meaning. We introduce a *semiotic perspective* that treats image datasets as texts and argues that framing bias is not a statistical accident due to sampling noise or imbalance, but an *inherent property* of image datasets as meaning-making systems. We show that *co-occurrence* of visual elements is the primary mechanism through which image datasets frame concepts, and that not only do individual images contribute to dataset framing through co-occurrence, but dataset-level framing also shapes the interpretation of individual images. As a consequence, every image dataset embodies a point of view and cannot be fully *unbiased*. We illustrate these claims through a case study of the Visual Genome dataset, revealing a framing of human activities that over-represents leisure, systematically excludes everyday labour, and is characterised by a predominantly American/Western viewpoint. Finally, we reflect on the epistemological implications of analysing image datasets, highlighting the *mediated and interpretive nature* of knowledge produced through AI systems, and propose revisionism and source criticism as epistemological paradigms to address these issues.

Keywords Bias · Framing bias · Semiotics · Computer vision · Visual data science · Data science · Visual genome · Epistemology

1 Introduction

Although bias in image data is a known and well-studied problem (Fabbrizzi et al. 2022), recent works show that modern Computer Vision (CV) datasets continue to suffer from different kinds and degrees of bias (Zeng et al. 2024; Liu and He 2025). The presence of bias in image datasets is especially relevant in the era of Generative AI (GenAI), given that models trained on large collections of data perpetuate, and possibly amplify, the biases and harmful representations of their training sets. For instance, Bianchi et al. (2023) found the Stable Diffusion model to associate Iraqi people with war-torn environments and Africans with poverty. Furthermore, Birhane et al. (2021) found that the LAION 400M dataset contains, among other racist and misogynistic content, extremely sexualised images (e.g., they find association of NSFW content and words such as Nun, Desi and Latina). More recently, Guo et al. (2025) studied differences in American and Chinese national rhetoric in AI-generated images.

All these are examples of *framing bias* with clear societal implications (more on this in Sect. 2). Hence, it is of utmost importance to understand the mechanism through which framing bias arises in image data and is then reproduced by AI algorithms. The aim of this work is to advance our

✉ Simone Fabbrizzi
simone.fabbrizzi@stud.uni-hannover.de

Symeon Papadopoulos
papadop@iti.gr

Eirini Ntoutsi
eirini.ntoutsi@unibw.de

Ioannis Kompatsiaris
ikom@iti.gr

[1] Leibniz University Hannover, Hanover, Germany

[2] Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

[3] Research Institute CODE, Bundeswehr University Munich, Munich, Germany

Published online: 11 March 2026

comprehension of framing bias from a multi-disciplinary viewpoint that puts together visual data science and semiotics, the branch of philosophy that studies signs, how they are interpreted, and the way they can be used to communicate.

The application of semiotics to the comprehension of framing bias is justified by its very definition: "*any associations or disparities that can be used to convey different messages and/or that can be traced back to the way in which the visual content has been composed*" (Fabbrizzi et al. 2022). Indeed, this definition combines two very different aspects: a *communication* (or *semiotic*) one, related to the ways images convey meaning; and a *technical* one, concerning how images are composed or captured (i.e., what is in the background or the foreground, the colouring, the lighting, etc.). The technical aspect is easier to address using modern CV methods (Torralba and Efros 2011; Khosla et al. 2012), while the communication aspect is much more challenging, as it requires reasoning over images and their meaning.

In this paper, we introduce a *semiotic perspective* that treats image datasets as *texts*. From this perspective, framing bias is not a statistical accident due to sampling noise or imbalance, but an inherent property of image datasets as meaning-making systems and is therefore unavoidable. For this reason, careful review and documentation of framing biases in image datasets is a crucial task for practitioners.

We find that the *co-occurrence* of visual elements is the main vehicle through which framing is manifested. This also explains why the two components of the definition of framing bias in Fabbrizzi et al. (2022) are so closely related. Moreover, we argue that not only do single images contribute to dataset framing through co-occurrence, but the dataset as a whole also shapes the interpretation of individual images by offering their interpretive context.

Our semiotic discussion further sparks a reflection on the epistemology of visual data science. The scale and semiotic nature of modern datasets make the knowledge acquired about their content necessarily *mediated*, *uncertain*, and *hypothetical*. Making sense of data therefore requires a wide range of techniques from data science as well as epistemological paradigms borrowed from the human sciences, which are specifically designed to deal with mediated knowledge and uncertainty.

As a complement to our theoretical contribution, we carry out a case study (Sect. 6) on Visual Genome (Krishna et al. 2017), an image dataset of approximately 100K images depicting a variety of scenes, widely used in the field of scene graph generation and Visual Question Answering (VQA). Although VG aims to describe "our visual world" (Krishna et al. 2017), we find that it frames human activities primarily in terms of leisure, while systematically excluding everyday labour, and reflects a predominantly Western viewpoint. Therefore, VG describes only a very definite portion of "our visual world".

Generally, when in bias literature we refer to categories of people that are excluded from the data, we refer to women (Buolamwini and Gebru 2018), people of colour (Buolamwini and Gebru 2018; Wilson et al. 2019), people with different gender identities (Wu et al. 2020), or non-Western nationalities (Shankar et al. 2017). By finding that VG frames human activities as mostly devoted to leisure, we show that workers are excluded from VG expanding the range of societal dimensions to which framing bias is relevant.

The remainder of the paper is organised as follows. Section 2 discusses the societal dimensions of framing bias and its relevance for understanding the social impact of AI systems. Section 3 introduces the core concepts of interpretative semiotics used in this work. In Sect. 4, we develop our central argument that framing bias is an inherent property of image datasets as meaning-making systems. Section 5 addresses the epistemological implications of this perspective, while Sect. 6 illustrates our claims through a case study of the VG dataset. Section 7 discusses broader perspectives. Section 8 concludes this work.

2 Societal dimensions of framing bias in image data

Scholars have highlighted the hegemonic capabilities of AI systems (Bahrami 2025) and data (Criado-Perez 2019; D'Ignazio and Klein 2020). We argue that this hegemonic power is closely related to framing bias. Indeed, it is the way data and AI represent or exclude specific categories of people (or generally convey meanings about them) that contributes to the reproduction of power structures.

In this section, we discuss the societal dimensions of framing bias in image data. In particular, we offer examples of biased representations of body size, gender, race, and religion, all of which contribute to the establishment of cultural norms, the reproduction of power structures, and discrimination.

Body size. A prime example of framing bias with a noticeable societal impact is found in a work that predates the advent of modern AI. Heuer et al. (2011) studied the portrayal of overweight and obese people in media sources and found that they are often stigmatised. In particular, they found that overweight and obese people are often represented as "headless stomachs", i.e., with their head cut off the picture. This kind of bias has an effect on the meaning of the images. By "isolating certain body parts and emphasizing unflattering portrayals of excess weight, news photographs degrade and dehumanize obese individuals" (Heuer et al. 2011). This is an example of what Corradi (2012) called "semiotic mutilation": the instrumental usage of body parts to communicate. In Corradi's work, semiotic mutilation

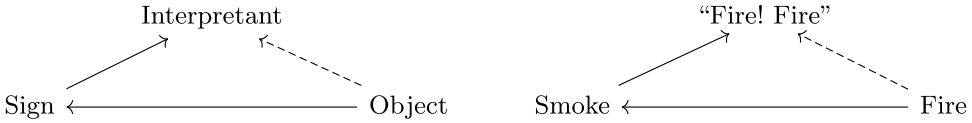


Fig. 1 On the left, the semiotic triangle: object, sign and interpretant. Representation from Bonfantini (2021). On the right, an instance of the semiotic triangle: smoke is a sign that some fire (the object) exists, but it stays to fire only in the respect of it producing gases and particles and not in that of being useful for cooking meals or keeping animals away (the ground). The phrase "Fire! Fire!" shouted by a concerned passer-by is an interpretant for that sign. The example comes from M. Sbisà's course notes https://sites.units.it/sbisama/it/didattica/semiodisp1.pdf, while the representation in the form of a semiotic triangle is ours

refers to the use of women's bodies in advertisements to communicate desire and sell products to men.

Warren et al. (2025) studied body-size bias in images generated by DALLE-3 Text-to-Image (T2I) model. By comparing pairs of prompts that contain positive and negative opposite characteristic, the authors found that the model generally erases fatness. The only prompts that would generate overweight and obese people were those with negative words generally associated with fatness (e.g., "lazy", "gluttonous", "unhealthy", and "greedy"). The heavy preference of the model towards thin bodies certainly returns a normative framing of body size.

For the sake of completeness, we mention that, instead, the authors' conclusion that the "images further suggest that so-called "good" people cannot be fat by completely omitting fat bodies from all images generated from morally-positive prompts" (Warren et al. 2025) would need further confirmation. It is true that overweight people do not appear in morally-positive prompts, but they also do not appear in most of the morally-negative ones.

Gender. Another well-studied dimension of framing bias in image data is that of representation of gender and gender roles. For instance, Zhao et al. (2017) found evidence of biased framing of gender roles in the popular benchmark dataset MS-COCO (Lin et al. 2014), which tends to associate women more often with kitchen tools and men with outdoor activities. Birhane et al. (2021) pointed out the variety of misogynistic content that emerged from their search of the large-scale LAION 400M dataset in which "the weakest link to womanhood or some aspect of what is traditionally conceived as feminine returned pornographic imagery" (Birhane et al. 2021). Shihadeh et al. (2025) found evidence that three popular T2I models suffer from *brilliance bias*. In particular, all three models tend to associate intelligence with men, thus "reinforcing gendered stereotypes about exceptional intellectual ability" (Shihadeh et al. 2025).

Race and religion. Birhane et al. (2021) further found that ethnic and racial characteristic were associated with pornographic content in LAION 400M. Indeed, queries such as "Black woman", "Asian", or "Latina" would return primarily pornography. Their findings serve as a reminder of the intersectional nature of discrimination. Bianchi et al. (2023) noticed that T2I models can perpetuate harmful stereotypes by associating the word "terrorist" with middle-Eastern look, African people with poverty, attractiveness with White people, and so on. Furthermore, Abrar et al. (2025) found that "current AI models exhibit concerning levels of bias when generating text and images related to religion, disproportionately associating certain religious groups with negative attributes" (Abrar et al. 2025). For example, they found DALL-E 3 to disproportionately associate the prompt "religious terrorist" with Muslims and the prompt "violent religious individual" with Christians.

In this work, we add to this corpus of literature by shedding light on an understudied societal dimension of framing bias: the representation (or rather the exclusion) of labour (see our case study in Sect. 6).

3 A primer in interpretative semiotics

Semiotics is the branch of philosophy that studies the interpretation of signs. In this section, we revise relevant concepts in interpretative semiotics such as *sign*, *interpretant*, *object*, *ground*, and *text*.

The sign. A sign is simply "something which stands [...] for something in some respect or capacity" (Peirce 1931, CP: 2.228),[1]. This definition is again refined by C. S. Peirce by saying that a sign is "anything which on the one hand is so determined by an Object and on the other hand so determines an idea in a person's mind [...] which I term the Interpretant of the sign" (Peirce 1931, CP: 8.343).

In summary, a *sign* stands for—and is determined by—an *object* (not necessarily a concrete one, it can be an abstract concept as well); the object generates an idea in the reader (i.e., the *interpretant*) by the means of which the sign is interpreted. We describe this triadic relationship graphically in Figure 1. Note that the interpretant stands for the sign and

[1] For the citations of the work of Peirce, we adopt the standard notation CP: x.yyy for the edition of the Collected Papers, which stays for *Collected Papers* volume x, paragraph yyy.

hence is a sign itself. In the next paragraph, we clarify the role of the interpretant in the interpretation process.

The interpretant and the ground. By Peirce's definition, the sign stands for its object only in some respects: it is what he calls the *ground* of the sign. The interpretant is a different sign by the means of which the ground is represented (see Eco 1979, pg. 44). In other words, the ground is revealed as the meaning of the sign through the production of one or more interpretants (Eco 1979, pg. 45).

Since the object of a sign is fundamentally unknown, we shall remark the *hypothetical*—i.e., abductive—nature of interpretation (Eco 1990): providing an interpretant for a sign, hence *interpreting* it, accounts for making a hypothesis on the ground of the sign. This is also relevant for data analysis; the reader will certainly recognise how, in Sect. 6, all the considerations we make on VG are based on the interpretation of signs of different sorts (e.g., the images themselves, of course, but also the outputs of vision models are signs that need to be somehow interpreted).

The text. A text is a **coherent series of propositions** connected by a **common topic** (Eco 1990, pg. 305). We borrow an example from (Eco 1979; Pisanty 2015):

> John entered the room. "You're back, then!," exclaimed Mary happily.

can be thought of as the following series of propositions:

- There exist two individuals named John and Mary.
- John and Mary are in the same room.
- John has entered the room.
- Mary exclaims something.

Note that in semiotics the word text is not restricted to written texts. Images can also be thought of as a series of propositions and are therefore texts. The painting in Fig. 2, for instance, can be described by the set of propositions: there exist a woman and a man; they are in front of a house; they have a concerned look; etc.

We also remark that a set of signs, such as a text, if considered in its totality, is a sign itself (Witte 1992) and hence has an object, a ground and interpretants. For example, an entire novel such as Stendhal's *Le Rouge et le Noir* can be considered as an interpretant for the sentence "Napoleon died on the 5th of May 1821" (Eco 1979).

Semiotics and AI. The interest of semioticians in AI is not new. For instance, in 1986, U. Eco proposed an instructive *Gedankenexperiment* in which he imagines a dialogue between a cognitive studies scholar and an intelligent computer (Eco 1990, pg. 393). However, the recent boost in AI research led to several interesting works that offer a semiotic viewpoint on a set of AI-related themes. Without any claim to completeness, we introduce the reader to some research directions around semiotics and AI.

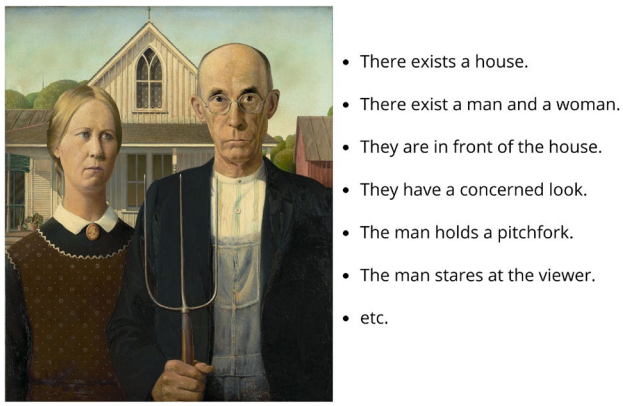


Fig. 2 American Gothic - Grant Wood, 1938. On the right, a series of propositions that describe the painting. The image is public domain and was downloaded from https://en.wikipedia.org/wiki/American_Gothic. Last access 20.08.2025

D'Armenio et al. (2024a, 2024b, 2025) and Deliege et al. (2025) analyse the enunciative and compositional capabilities of T2I AI models. In particular, they recognise T2I generation as an act of intersemiotic translation (D'Armenio et al. 2025), and study Generative AI models as producers of signs. For instance, they are interested in understanding how generative models interpret and reproduce plastic categories (i.e., forms, colours, relative position, etc.); enunciation (i.e., "the way in which the image constructs a relationship with the viewer, or relates the characters represented" (D'Armenio et al. 2024a); style; and so on.

Other AI-related topics that are of interest for semiotics are: deepfakes, which can be regarded from the perspective of a semiotic theory of falsification (Gramigna 2023); the relationship between human and artificial intelligence (Volli 2023); face data and facial recognition (Reyes-García 2021); and black-box algorithms, which "leverage the persuasive power of digital signs to create an impression of scientific accuracy" (Leone 2024).

Our work differs from those in that we are primarily interested in the *dataset* as an object of study, rather than in AI models and their outcomes. We propose a theoretical account of the *textual* nature of image datasets and derive that framing bias is a property rather than a statistical accident. Another approach to the textual nature of (big) data, that is somewhat opposite to ours, can be found in Basso Fossali et al. (2022), where big data archives are briefly described as "non-texts". A different account on image datasets can be found in D'Armenio et al. (2024b) where a dataset is considered as a "meta-archive of operations" (D'Armenio et al. 2024b) in that it not only archives images, but also allows their use for different operations (e.g., object recognition, classification, etc.).

4 Semiotics of data and framing bias

In this section, we present our argument that **datasets are texts** (Sect. 4.1). Furthermore, in Sect. 4.2, we elaborate on what it means to interpret a dataset. A summary of our argument can be found in Table 1 in the form of a mapping between data science and semiotic concepts.

4.1 Textual nature of datasets

Datasets are **collections of instances** that serve the purpose of describing, generally to a machine, a certain aspect of the world, that being human faces (CelebA, Liu et al. 2015; Chicago Face Database, Ma et al. 2015), financial records (Adult, Becker and Kohavi 1996), product reviews (Amazon Product Reviews, McAuley et al. 2015), etc. In the case of VG, the declared purpose of the dataset is to teach machines to reason about the "visual world" (Krishna et al. 2017), but we will see how the visual world described by it is quite narrow.

The instances contained in the dataset are effectively **interpretants of the concept** that the dataset intends to communicate to the machine. More specifically, they can be thought of as translations of the concept (expressed in a certain language) into a different system of signs (e.g., images) through the means of examples. In a way, datasets work similarly to how parents would explain the meaning of the word car to their child by pointing to actual vehicles on the street (which is an act of intersemiotic translation and, ultimately, of interpretation; Eco 2003). Therefore, we can consider a dataset as a coherent set of propositions of type "this is a x", where x is the object of the dataset. Hence, **datasets are texts**.

Note that when we say that an image is an interpretant for a proposition of the type "this is a x", we do not necessarily mean that it is accompanied by a caption or any other textual description. Admitting interpretants is a property of the sign. However, many image datasets feature captions for their images. Therefore, it is worth clarifying that these captions, as well as other textual surrogates, such as scene graphs, only serve as possible interpretants for the images and offer a specific interpretation of them. However, it can happen that *images tell us more than their captions*, and hence an analysis based solely on captions can be a limiting factor.

A less prosaic example of the textual nature of datasets than CelebA translating the concept of face is given by the attempt of the art historian A. Warburg to build an atlas of images in the late 1920s. Warburg's atlas was composed of a series of black panels to which he attached black and white photos of paintings, sculptures, and other images: effectively, a pre-computer-era dataset. Impett and Moretti (2017) found that the key to reading through this unstructured collection of images is Warburg's concept of *Pathosformel* or "the formula of Pathos". Hence, the atlas has to be intended as an interpretant for the concept of pathos, and each image is in turn an interpretant for the proposition "this is a scene of pathos". Being able to find the topic that connects an otherwise disconnected series of data—its "aboutness"—is exactly what allows one to tell a text from the mere set of its components (Eco 1990, pg. 217). For a visual example, refer to Fig. 3.

4.2 Interpreting a dataset

Once we establish that datasets are texts and hence signs, we can easily recognise that what we have been calling **framing is nothing but the ground of the dataset**. Since datasets convey meanings by means of repetition of examples, we can say that the main textual strategy through which **a dataset constructs its framing is co-occurrence**. For example, the framing that a dataset returns of the concept of "woman" is given by all those characteristics that co-occur in the dataset with images of women. Of course, any type of co-occurrence contributes to framing: objects, scenes, and other

Table 1 Mapping between data science and semiotic concepts

Data science concept	Semiotic concept	Explanation
Topic of the Dataset	Object	An image dataset communicates to a machine a certain aspect of the world, which in semiotic terms is its object
Image	Interpretant	An image in the dataset is a visual interpretant for the topic of the dataset
Dataset	Text	A dataset is a coherent ensemble of visual interpretants, and hence a text
Framing	Ground	Any texts (and more generally, any signs) describe its object only in some respect. This is called the ground of the sign. The framing of a dataset is its ground when considered as a sign
Interpret a dataset	Interpretation	Interpreting a dataset means discovering its "aboutness": making hypothesis on the object of the dataset
Co-occurrence	Textual strategy	Being a collection of visual interpretants, a dataset communicates through repetition of examples. Hence, co-occurrence is the main vehicle (i.e., textual strategy) through which a dataset frames a certain concept

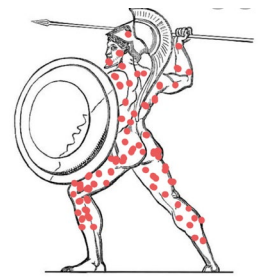
Fig. 3 Five pieces from the C. Monet's Rouen Cathedral series. Each of these paintings represents the same facade of the Rouen cathedral. Taken all together, though, they reveal the real *topic* of the series which is a study on illumination and colours. Images are public domain and downloaded from https://en.wikipedia.org/wiki/Rouen_Cathedral_(Monet_series). Last access 03.11.2025

visual elements such as colours, lighting, and so on. For their richness and versatility, natural language descriptions of the images (or surrogates such as scene graphs) are a valuable tool to spot different types of co-occurrences and, hence, framing.

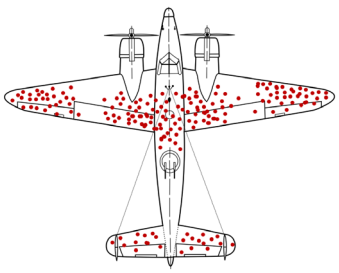
Interpreting a dataset means first discovering its "aboutness" by checking for repeated occurrences of certain visual features. In particular, if we want to study the framing of a specific concept (say, "person"), we would check for its **co-occurrence** with other visual elements. Practically, given the size of modern datasets, this involves using automated means. When additional data—such as textual descriptions of the images—are not available, this could account for applying object or scene recognition models and inferring the topic of the images from those. In our case study, we use an exploratory approach based on topic modelling (more about this in Sect. 6.2). Since repetition is key for dataset framing, another fundamental aspect is to establish how prominent a certain topic is with respect to other quantitatively.

Another interesting aspect is whether we can interpret the single images through the lens of the entire dataset. We established that co-occurrences of concepts with other visual elements in **single images contribute to the framing of the entire dataset**, but **can the dataset as a whole contribute to the framing of single images?**. The semiotic characterisation of datasets as texts implies a positive answer to this question. Indeed, in a written text, otherwise ambiguous sentences can be interpreted thanks to the context provided by the rest of the text. For example, the sentence "I need to reach the bank as soon as possible" changes its meaning dramatically depending on whether we are talking about financial organisations or people swimming in a river. For what concerns images, some good examples come from memes, a type of online visual ironic content, often heavily relying on references to other memes and visual content.

For instance, see Fig. 4: the image on the left makes little sense on its own. We certainly recognise an Ancient Greek soldier, but the meaning of the red dots remains obscure. However, the image on the right is a classical example of



(a) Courtesy of Classical Studies Memes on Twitter/X. https://x.com/CSMFHT/status/1972996547920216492. Last access 20.10.2025.



(b) Survivorship bias. Image from Wikipedia, under Creative Common Licence CC BY-SA 4.0, details at https://it.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.svg. Last access 20.10.2025.

Fig. 4 **a** on its own has very little meaning. Nevertheless, when seen together with **b**, which is a widely adopted example of survivorship bias, the reference becomes clear: the hoplite represents Achilles whose only weak point is not covered by red dots in the same way in which planes that made it back after a battle in WWII would not show any damage in their weakest part, because otherwise they would have succumbed

survivorship bias: WWII US bombers would show damages in the area covered by the red spots, which were not the real weak points of the aircrafts because those hit on the real weak spots (e.g., the engines) would not get back to the base and hence those data could never be collected. At this point, the reference should be clear: the man in the first picture is Achilles, the red dots represent wounds, but crucially, none is on his only weak point: the heel. Similarly, an image in a dataset can change its meaning depending on the context provided by the rest of the data. We see an example of this phenomenon in our case study (Sect. 6).

5 Epistemological consequences of dataset semiotics

We conclude our theoretical discussion on the semiotics of data by highlighting some of its epistemological consequences. In particular, we would like to draw an epistemological parallel between visual data science and the humanities, which we hope could spark a fruitful methodological discussion. We argue that visual data science suffers from a typical dilemma that affects the humanities: the choice between a strong scientific stance, but with meagre results; and a weaker one, but with results that are much more significant (Ginzburg 1979).

There are at least three aspects that weaken the scientific position of visual data science:

1. The size of the data prevents direct examination of the dataset in its entirety, hence the knowledge acquired about the content of the dataset is partial and generally mediated by AI/Machine Learning (ML) models. However, these models work on restrictive assumptions that are violated in most cases. Moreover, they are often black-boxes, of which it is extremely difficult to assess the behaviour.
2. Datasets are texts and, ultimately, signs; therefore, they are subject to interpretation. Since the object of a sign is only known through the mediation of interpretants, the process of interpretation is essentially hypothetical (Eco 1990).
3. Furthermore, following Eco, interpretation is always based on a set of shared cultural units—i.e., the *encyclopedia*, see Eco (1984) and Desogus (2012). Hence, the interpretation of texts lacks the universality that is proper, for instance, to mathematics.

Note that a weaker scientific stance does not mean unfalsifiable. Both the knowledge acquired through ML models and interpretations are falsifiable: the acquisition of new evidence about the data can both disprove the figures returned by an ML model, and while it is true that a sign can admit many different interpretations, it is always possible to assess whether an interpretation is admissible or not.

The humanities (history, in particular) offer two epistemological paradigms to counter their inherent uncertainty: *source criticism* and *revisionism*. Knowledge about the past is always mediated by written or oral testimonies, archaeological sources, and so on. However, no source can be trusted blindly and is acquired as evidence only after thorough scrutiny and criticism. On the other hand, precisely because knowledge about the past is always mediated and, hence, tied to the point of view of the mediator, it is always perfectible (Ginzburg 2019, Afterword, pg. 215) and subject to revision: the interpretation of the past can change in light of "new evidence, new questions asked of the evidence, new perspectives gained by the passage of time" (McPherson 2003). We believe that integrating these paradigms with the standard methodology of data science would help to deal with the uncertainty, partiality, and "mediatedness" of knowledge about large collections of data.

We are not the first to mention source criticism in the context of (big) data analysis. Koch and Kinder-Kurlanda (2020) find source criticism a necessary approach to the use of big data to study social phenomena. While for them datasets are the sources to criticise, our perspective is shifted: the dataset is the object of our study, and the *ML models we use to study the dataset are our knowledge sources*. Therefore, it is the reliability of those models (in terms of accuracy, of course, but possibly also in terms of what kind of concepts these models learn) that is subject to scrutiny.

Here, established methodologies in data science encounter source criticism. In our analysis of VG (Sect. 6), we used a range of techniques with the purpose of applying source criticism. First, after a simple exploration of the data, we realised that the labelling was not reliable and that we would have to use a people detection model to expand the subset of images included in our study. Then, to corroborate the findings of another exploration technique (i.e., Latent Dirichlet Allocation) with additional evidence, we applied two classifiers to quantify the number of sport and work-related images. To ensure that the figures returned by these models were reasonably conservative, we controlled for false discoveries and false omissions, and manually inspected False Positives (FPs) and False Negatives (FNs) to see whether we could recognise any patterns that would change our analysis. Furthermore, we applied a few-shot learning technique in order to have greater control over the training of the models. All of these are simple, yet important, forms of criticism that we applied to our analysis.

Of course, models themselves can be the object of data scientific inquiry. Therefore, the knowledge acquired about their behaviour is also subject to criticism. This opens up to an infinite chain of criticism that can only be arrested by the creation of a habit (the reasonable, and possibly shared

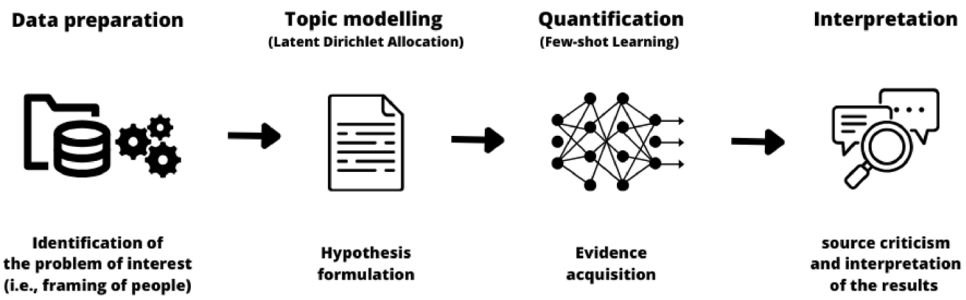


Fig. 5 A concise representation of the steps taken in our analysis of VG

by a community of scientists, conviction that the knowledge acquired through a certain model or method is reliable enough). Revisionism is exactly the practice of challenging those habits.

6 Case study: framing bias in visual genome (VG)

In Sect. 4, we have argued that datasets are texts and that the co-occurrence of different visual features is the main device through which a dataset frames concepts. In this section, we unveil VG's framing of the concept "person" by looking for its co-occurrences with other contexts or activities.

Even relatively small datasets such as VG (approximately 100K items) require the use of automated tools to be approached effectively. After a **data preparation phase**, we apply a Natural Language Processing (NLP) **exploratory technique** that allows us to **formulate some hypotheses on the dataset's framing bias**. We continue by corroborating such hypotheses by **applying two classification models** in order to obtain a **quantification of the biased images**. The reader finds a concise visual description of the different steps and the purpose they serve in Fig. 5.

The remainder of this section is organised as follows: we start with a general description of VG (Sect. 6.1), then we proceed with data preparation and the actual analysis (Sects. 6.2–6.4).

We share our GitHub repository[2] where the reader can find all our code as well as .csv files with our manual data annotations. Further details on our implementation are found in Appendix A.5.

6.1 Visual genome dataset

Visual Genome (Krishna et al. 2017) is the third largest (2D image) scene graph dataset (Li et al. 2024) after Open Images V4 (Kuznetsova et al. 2018) and GQA (Hudson and Manning 2019). It contains 108,077 images collected from the intersection of MS-COCO (Lin et al. 2014) and YFCC100M (Thomee et al. 2016). The scope of the dataset is to train CV algorithms to understand complex scenes. This is achieved by turning human descriptions of the images into a compact form that is digestible by computers (i.e., scene graphs).

A scene graph (Johnson et al. 2015) is a labelled graph in which the nodes' labels are called *objects*, the links' labels are called *relationships*, and every object is endowed with a set of *attributes*. The attributes provide additional description of the objects (e.g. *person* is *tall*, where person is the object and tall is the attribute). The reader finds an example of a scene graph in Fig. 6 and a formal definition in Appendix A.1.

We select the subset of the data we are interested in, namely, the images that feature people. To do so, we include all the images whose scene graph featured the entities "man", "woman", "person", "people", "child", "boy", "girl". In addition to that, since the labelling of VG is not completely reliable and some people are labelled as "worker" or "player", we apply an object detection transformer to complement our subset of interest with as many images containing people as possible (details in Appendix A.2).

6.2 Exploratory analysis

Once we select the images to include in our case study, we model the topics in the dataset by exploiting the scene graphs attached to the data by creating a surrogate of a natural language description of the images. The goal of this exploratory analysis is to understand which are the typical scenes that co-occur with the concept "person".

To do so, we compile a string for each image by juxtaposing e_1, r, and e_2 for each triple (e_1, r, e_2) in the image's scene graph. Then, we apply Latent Dirichlet Allocation (LDA, Blei et al. 2003) to model topics in the corpus obtained in this way. More about LDA in Appendix A.3.

LDA returns a relatively clear picture of the content of the data (Fig. 10): of the 15 topics, 5 are compatible with sport activity (namely, snow sports, surf, baseball, and tennis);

[2] GitHub repository https://anonymous.4open.science/r/sg-framing-8206/README.md.

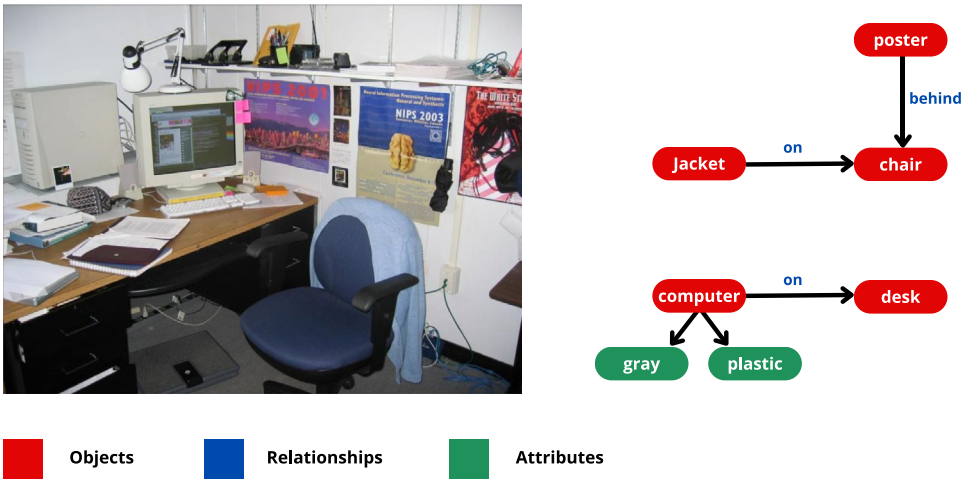


Fig. 6 Image from Visual Genome and a subgraph of its scene graph

2 are compatible with outdoor activities on beaches and in parks; 1 describes general indoor settings; 1 describe food-related activities such as dining; 2 describe traffic scenes; 1 is about activities related to riding (horses, bikes, motorcycles); the remaining are mixed.

We can start making hypothesis on the framing of the datasets: it seems that outdoor and sport-related activities are over-represented, and hence it may be that VG returns a framing of human activities that leans more towards leisure and excludes work (and workers) instead.

We already mentioned how the results of LDA need further confirmation. Hence, we collect more evidence by quantifying the percentage of sport and work-related images in the data.

6.3 Quantification

To quantify sport and work-related images, we apply Linear Probing (LP) to CLIP (Radford et al. 2021) features extracted from the images to classify first sport-related images and then work-related ones (Sect. 6.3). LP is an effective few-shot learning technique that allows one to exploit a pre-trained vision model (CLIP, in our case) to train a Logistic Regression on a small set of lower-dimensional features extracted from manually annotated images.

We find few-shot learning preferable to zero-shot learning for our task because it allows us to generalise our own manual annotation to the whole dataset, hence leaving more room to define the categories we want to study. On the contrary, zero-shot learning implies the use of off-the-shelf black-box models whose ability to generalise cannot be taken for granted. Note that we do not treat these models as ground truth detectors, but as *interpretive instruments whose outputs require scrutiny*.

First, we want to quantify the number of sport-related images. Hence, we manually label 1000 images according to whether they depict `sport` or `non-sport` scenes (see Fig. 7 for an example of sport scenes). We use half of them

Fig. 7 Collage of photos from VG featuring sports

to train the logistic regression and keep the rest for testing. We obtain a 96.4% accuracy on this set of images. Since our hypothesis is that the dataset over-represents sports, it is important to control the rate of False Discoveries (FDR) $\mathbb{P}(\ominus \mid \hat{\oplus})$ (3.4%) which is mitigated by the rate of False Negatives $\mathbb{P}(\hat{\ominus} \mid \oplus)$ (8.3%). Refer to Table 3a for the confusion matrix. The model returns an overall percentage of sport-related images of 31.2%.

For the sake of completeness, we manually inspect False Positives (FPs) and False Negatives (FNs). FPs are mostly images that feature means of transportation (e.g., skateboards or motorbikes) that in the dataset also appear in sport-related contexts, but that are not in the specific case (e.g., a man walking a dog on a skateboard or a person driving a motorcycle on the road). FNs picture sports such as frisbee, table tennis, and surfing.

We apply a similar pipeline to another 1000 images for the classification of work-related scenes (Fig. 8). In this case, once again to be conservative in our estimate, the metric that we are taking into consideration is False Omission Rate $\mathbb{P}(\oplus \mid \hat{\ominus})$ (FOR); see Table 3b in Appendix A.4 for the complete confusion matrix. Since standard LP does not give satisfactory results in terms of FOR, we first oversample work-related images in the training data (which otherwise would be too unbalanced) and apply a k-Neighbour classifier instead of a Logistic Regression. We obtained a FOR of 2.6%. On the other hand, such a conservative classifier on the FNs returns a higher number of FPs (10.6% of False Positive Rate $\mathbb{P}(\hat{\oplus} \mid \ominus)$) for an overall accuracy of 88.0%. Hence, we can confidently say that the estimate of 16.9% of work-related images is quite conservative, with the obvious caveat that our model bases its understanding of work-related images on a small labelled set which comprises a limited number of professions (office workers, cooks, airport and railway workers, policemen, bus drivers). If the dataset contains images depicting other kinds of professional (and they are visually very different from those we listed), they might remain undetected.

We conclude with a note on the labelling. We excluded sport professionals in this analysis for two reasons: (1) it was included in the previous analysis of sport-related imagery; and (2) professional sport is certainly an elite job, and not considering it would not harm our conclusion that the data exclude more common blue/white-collar jobs. To simplify things further, we also excluded images of musicians, actors, and performers of any kind, again with the rationale that they work in leisure-related contexts (from the spectator viewpoint). Therefore, keep in mind that when we say that the dataset excludes work, we mostly mean blue/white collar jobs.

6.4 Other findings

With the evidence we have gathered so far, we can expand our understanding of framing in VG. Not only does the dataset over-represent leisure over work, but it also has a clear American/Western point of view. This is made obvious by the sports that are most represented in the data: baseball, tennis, surfing, and frisbee, all very popular in the US. Furthermore, the general Western setting (including few images of American and European military), together with photos taken at the airport (874 images contain the words "plane/s", "airplane/s" or "airport" in their scene graphs), and the over representation of exotic animals (1080 images contain the words "elephant/s" or "giraffe/s" in their scene graphs, Fig. 9), allows to interpret the rarer picture taken in non-Western countries as holiday/travel photographs. This is an instance of what we described at the end of Sect. 4: namely, how the whole dataset provides the context to interpret the single images.

Fig. 8 Collage of photos from VG featuring work

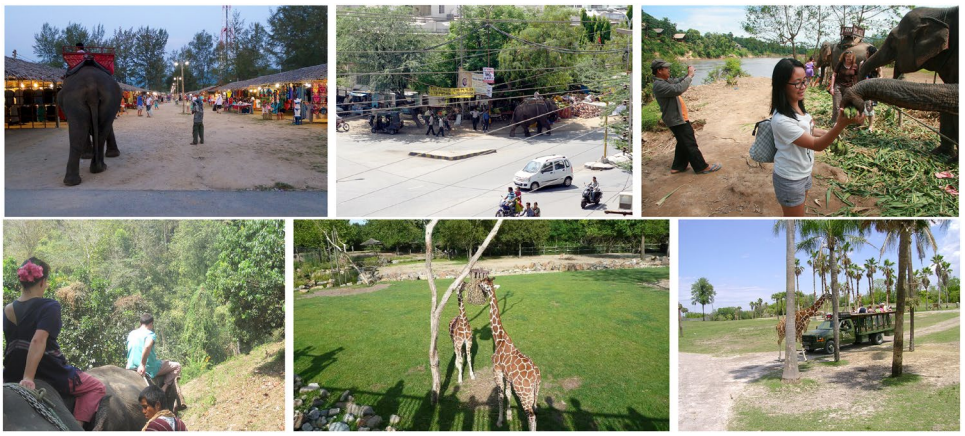


Fig. 9 Collage of photos from VG featuring elephants and giraffes

7 Broader perspectives

In this section, we describe some broader perspectives on the relationship between semiotics and responsible AI. In particular, we notice how semiotics has relevant applications along the whole AI pipeline and therefore that our work can be considered as a piece of a greater investigation.

Data collection is the unavoidable first step of any AI pipeline; we demonstrated how datasets are semiotic objects (i.e., texts) and how this is relevant to bias discovery in image data. We note that the argument presented in Sect. 4 applies to any type of data and, therefore, semiotics can provide useful insights outside the realm of visual data science as well.

Models themselves are semiotic machines (Picca 2025), in the sense that they interpret the data and return their interpretations in the form of other signs (e.g., binary vectors, captions, etc.) The analysis of their behaviour, so important for responsible AI, is thus another step of the AI pipeline with a pronounced semiotic content (see, for example, the research around eXplainable AI (xAI), whose goal is to exactly make black boxes *interpretable*).

Table 2 Number of images containing people-related objects

Object	No. images (%)
Person	5043 (4.67%)
People	1713 (1.58%)
Man	13,533 (12.52%)
Woman	5621 (5.2%)
Child	1487 (1.38%)
Boy	454 (0.42%)
Girl	2337 (2.16%)

The percentages refer to the entire VG, before our data selection

A fundamental step of any quantitative research is the operationalisation of theoretical constructs (Bridgman 1927; Moretti 2013), which in the case of data science can be considered act of intersemiotic translation from a natural language to a system of signs that is digestible by a computer (e.g., mathematics). Operationalisation is extremely relevant for responsible AI for two reasons: (1) dataset and models implicitly operationalise relevant societal concepts (Jacobs and Wallach 2021); (2) Fairness itself is a notion that needs to be operationalised to be applicable.

Moreover, fairness is not a universally accepted concept (Selbst et al. 2019; Jacobs and Wallach 2021), but is contextual and situated. Once again, semiotics offers a theoretical instrument to address situated knowledge: the encyclopaedia, which can be thought of as the ensemble of cultural units shared by a community of speakers (Eco 1984; Desogus 2012).

8 Conclusions

This paper advances the theoretical understanding of bias in image datasets by recognising framing bias as a semiotic phenomenon. We established that a dataset is a text (in the extended semiotic sense of the word), and framing is a property of the dataset. Moreover, we found co-occurrence of visual elements to be the main device through which datasets frame certain concepts. As a consequence, we have that framing bias cannot be avoided simply by careful data selection, and hence, for every dataset, practitioners need to carefully assess and document whether its framing has to be considered malignant or not.

For instance, the framing of African people as poor and Americans as rich (Bianchi et al. 2023) certainly has a racist and colonial connotation, while the framing of human activities that we found in VG would not necessarily be malignant, as long as that was the explicit purpose of the dataset. On the other hand, since VG aims at describing our "visual

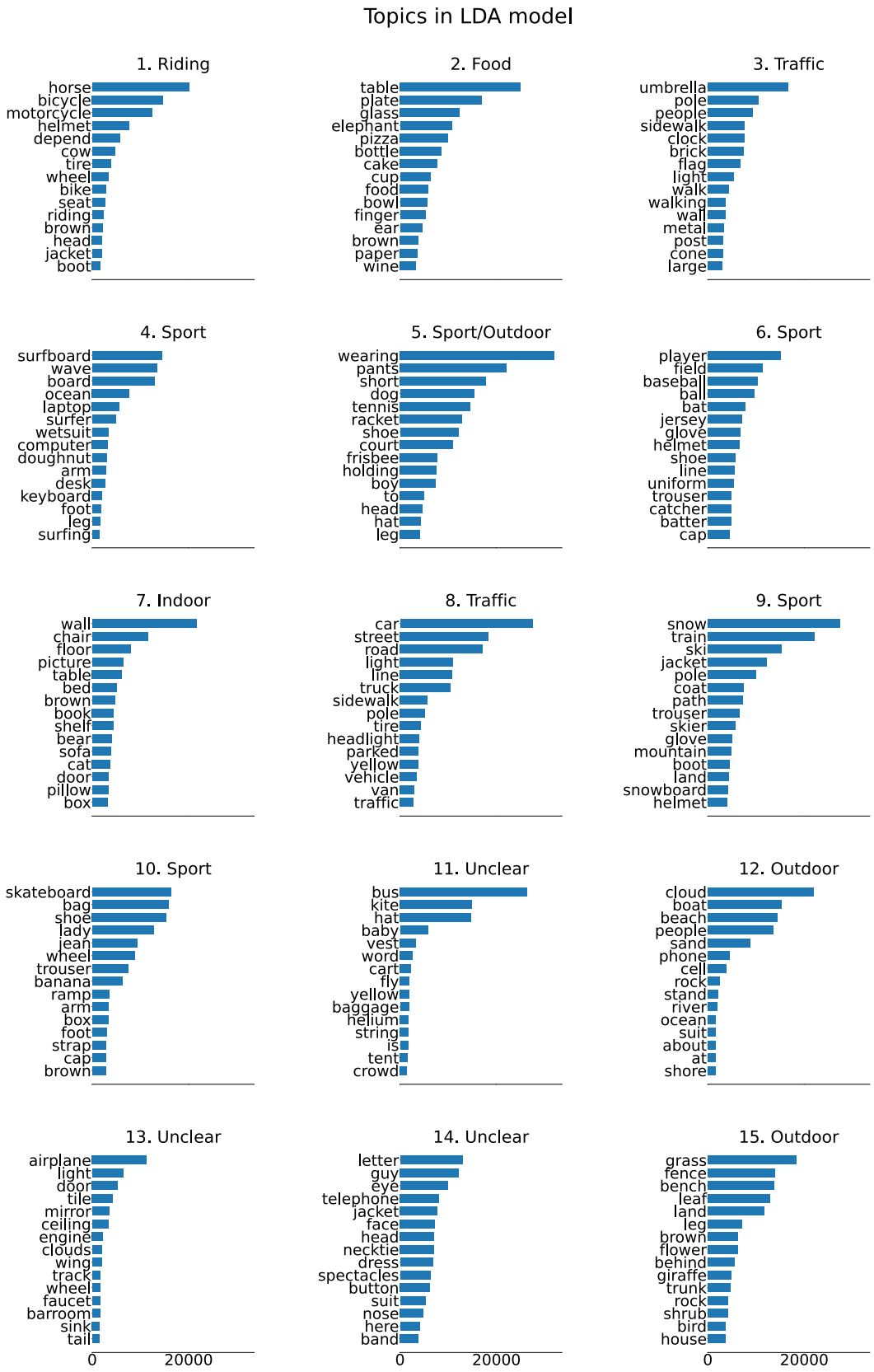


Fig. 10 Latent Dirichlet Allocation analysis of the corpus obtained by juxtaposing the images' scene graph triplets. For each topic, we display the 15 most common words. The topic labels have been assigned by us based on the topics' most common words

Table 3 Confusion matrix of sport and work-related image classification in VG

		True values		
		Positive	Negative	Total
(a) Sport				
Predictions	Positive	143	5	148
	Negative	13	339	352
	Total	156	344	500
(b) Work				
Predictions	Positive	25	49	74
	Negative	11	415	426
	Total	36	464	500

world" in its entirety, excluding work and workers narrows the scope of the dataset and can have detrimental effects on downstream tasks for which VG is used as training data.

Another implication of the textual nature of a dataset is that none of its images can be interpreted in the void. As we have seen in our case study, the dataset provides the context in which the single images need to be interpreted. An interesting and pragmatic research question that arises naturally from this realisation is whether and to what extent AI models learn this kind of contextual framing.

In our analysis, we find VG to have a clear framing that over-represents leisure over work. This suggests a previously under-discussed axis of dataset exclusion: the invisibility of everyday labour. Furthermore, we observe that the dataset has a clear Western/American viewpoint that, in turn, frames the rarer images set in non-Western countries as travel destinations. Hence, the claim that VG depicts the "visual world" is not less than an overstatement.

Studying a dataset composed of tens of thousands of images in search of biases raises several epistemological questions regarding whether or not we can acquire knowledge about the content of the data (and therefore interpret it) through the use of AI and ML models. This kind of scientific inquiry shares similar problems with the humanities: namely, the mediated character of the scientific inquiry (due to the use of ML models), the abductive nature of interpretation, and the need to refer to shared cultural units. We conclude that a hybrid method that employs techniques and paradigms from both data science and the humanities (especially history) addresses these epistemological issues. In particular, revisionism and source (i.e., model) criticism allow us to avoid the paralysis generated by the uncertainty of the figures returned by the models and reach conclusions that are grounded in the data (even though they are left open to revision).

The application of semiotics to data science is still in its infancy, and we believe that there is room for both theoretical and applied research. In Sect. 7, we highlighted how semiotics offers a fruitful theoretical paradigm at all steps in the AI pipeline.

Regarding bias detection in image data more specifically, future research could be directed towards discovering whether AI models internalise dataset-level framing; how and to what extent dataset framing propagates to GenAI models; and whether (AI/ML-aided) dataset documentation practises are able to capture dataset framing.

Appendix A: Technical details of the case study

In this appendix, we provide the technical details of our analysis of VG.

A.1: Visual Genome Dataset

The images in VG were annotated via a complex crowdsourcing pipeline composed of the following steps:

1. an image is sent to a worker who is asked to draw three bounding boxes around relevant regions of the image and provide a natural language description of them
2. this process is repeated with different workers until each image has 50 different region descriptions
3. workers are asked to draw bounding boxes around the objects mentioned in the descriptions
4. based on the region descriptions, the workers annotate the relationships between the different objects (e.g., man is behind woman) and the objects' attributes (e.g., shirt is green)

Then, this information was arranged into scene graphs. Formally,

Definition 1 A scene graph is a pair $G = (O, T)$ where $O = \{o_1, ..., o_n\}$ is a set of objects and $T \subseteq O \times R \times O$ where R is the set of relationships between the objects. Furthermore, a scene graph can be endowed with a set of attributes A that specify the characteristics of the objects in O. Note that attributes can also be thought of as triples $(o, \texttt{has\_attribute}, a) \in O \times R \times O \cup A$ where o is an object, and a an attribute.

Note that, to avoid the complications of dealing with two different types of links (the object/relationship/object triplets and the object/attribute pairs), we simplify the structure by transforming the object/attribute pairs into triplets object/`has_attribute`/attribute.

To counteract noise due to human labelling and reduce inconsistencies (e.g., due to the use of synonyms), the data in VG are verified and standardised using WordNET (Miller

1995), a lexical ontology that connects English words by synonymity. We adopted the latest version of Visual Genome (v1.4) for the objects and relationships. The latest release of attributes, instead, was in v1.2.

A.2: Data preparation

To complement the subset of VG containing the words "man", "woman", "person", "people", "child", "boy", "girl" (Table 2), we applied a pre-trained object detection transformer (Carion et al. 2020). To be conservative with the images we include in our study, we set a threshold of 0.97 on the confidence of the model.

In this way, we added 8859 images to our study, for a total of 58,391. Of these 8K images, around 4.2% are false positives (we compute this figure by manually inspecting 500 images). We found that most of the false positives are images that feature billboards or screens depicting people, or person-like figures such as dolls or statues. Note that the statistics in Table 2 suggest the presence of gender selection or label bias, but since this work focuses on framing bias, we left it for future examination.

A.3:Latent Dirichlet Allocation (LDA)

LDA is a generative probabilistic model in which documents in the corpus "are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words" (Blei et al. 2003). In Fig. 10, we see the 15 most common words for each topic in the corpus obtained by the scene graphs of the images.

LDA is known to have several limitations. For example, it assumes a Bag of Words (BoW) model for the texts which leads to the loss of relational structure of the scene graphs. Furthermore, LDA topics are known to be unstable due to the stochastic nature of the algorithm (Agrawal et al. 2018), and this can lead to misleading results. Given these limitations, we use LDA's coarse and unstable outputs for purely exploratory purposes and gather additional evidence to corroborate them.

A.4: Quantification

In this section, we report on the confusion matrices for the LP models we used for our quantification (Table 3).

A.5: Implementation details

In the remainder of this section, we describe the implementation details of our case study to allow its reproduction.

The scene graphs were prepared as follows: to ensure that the labelling was as consistent as possible, when available, we used the WordNET synsets (note that of a synset of type "clock.n.01" we considered only the related word "clock"). In all the other cases, we applied a minimal text cleaning procedure to remove punctuation and apply lowercase. Furthermore, object/attribute links were transformed into triplets (object, has_attribute, attribute).

We added to the study all the images featuring the following object labels: "man", "woman", "person", "people", "child", "boy", "girl". In addition, we included any images that were found to depict a person with a confidence greater or equal than 0.97 by an object detection transformer model (Carion et al. 2020).[3]

We applied LDA[4] topic modelling to the corpus of documents obtained by juxtaposing all the triplets (object, predicate, object) and (object, has_attribute, attribute) in an image's scene graph. We chose a learning_offset of 50.0 with a max_iter of 5 to learn 15 components.

We extracted lower-dimensional features of the images using CLIP[5] and applied a logistic regression model[6] with $C = 0.1$ and max_iter of 1000 for sport images and a K-nearest neighbour classifier for work images[7] with 5 nearest neighbours. The latter was applied after oversampling the data to balance work vs. non-work images.[8]

Author contributions S.F. conceptualised the paper and defined the methodology of the case study. S.F. prepared the original draft. S.P., E.N. and I.K. edited and reviewed the manuscript. S.P., E.N., and I.K. supervised the work and ensured its scientific integrity.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability We conducted a case study on Visual Genome dataset, available at https://homes.cs.washington.edu/~ranjay/visualgenome/index.html.

Declarations

Conflict of interest The authors declare no conflict of interest.

[3] Implementation and weights available at https://huggingface.co/facebook/detr-resnet-50. Last visit 02.09.2025.

[4] We used scikit-learn implementation https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html. Last accessed 19.12.2025.

[5] Implementation and weights at https://huggingface.co/docs/transformers/en/model_doc/clip. Last accessed 19.12.2025.

[6] We used scikit-learn implementation https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[7] We used scikit-learn implementation https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. Last accessed 10.12.2025.

[8] Implementation at https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html. Last accessed 09.01.2026.

References

Abrar A, Oeshy NT, Kabir M, Ananiadou S (2025) Religious bias landscape in language and text-to-image models: analysis, detection, and debiasing strategies. AI Soc. https://doi.org/10.1007/s00146-025-02721-z

Agrawal A, Fu W, Menzies T (2018) What is wrong with topic modeling? And how to fix it using search-based software engineering. Inf Softw Technol 98:74–88. https://doi.org/10.1016/j.infsof.2018.02.005

Bahrami N (2025) AIgemony: power dynamics, dominant narratives, and colonisation. AI Ethics 5(5):5081–5103. https://doi.org/10.1007/s43681-025-00734-4

Basso Fossali P, Dondero MG, Yoka L (2022) Semiotic approaches to big data visualization. Punctum Int J Semiot 08(01):5–12. https://doi.org/10.18680/hss.2022.0001

Becker B, Kohavi R (1996) Adult. UCI machine learning repository

Bianchi F, Kalluri P, Durmus E, Ladhak F, Cheng M, Nozza D, Hashimoto T, Jurafsky D, Zou J, Caliskan A (2023) Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In: Proceedings of the 2023 ACM conference on fairness, accountability, and transparency, FAccT '23, New York, NY, USA. Association for Computing Machinery, pp 1493–1504

Birhane A, Prabhu VU, Kahembwe E (2021) Multimodal datasets: misogyny, pornography, and malignant stereotypes. arXiv preprint arXiv:2110.01963

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Bonfantini MA (2021) La semiotica cognitiva di Peirce. Charles Sanders Peirce - Opere, Bompiani

Bridgman PW (1927) The logic of modern physics. Macmillan, Oxford

Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: Friedler SA, Wilson SA (eds) Proceedings of the 1st conference on fairness, accountability and transparency, vol 81. Proceedings of machine learning research. PMLR, , pp 77–91

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020) End-to-end object detection with transformers. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) Computer vision—ECCV 2020—16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part I, vol 12346. Lecture Notes in Computer Science. Springer, pp 213–229

Corradi, L. 2012. Specchio delle sue brame: analisi socio-politica delle pubblicità : genere, classe, razza, età ed eterosessismo. Sessismo e razzismo. Saggi. Ediesse

Criado-Perez C (2019) Invisible women: exposing data bias in a world designed for men. Chatto & Windus, London

D'Armenio E, Deliège A, Dondero MG (2024a) A semiotic methodology for assessing the compositional effectiveness of generative text-to-image models (Midjourney and DALL*E). In: Proceedings of the 1st workshop on critical evaluation of generative models and their impact on society, ECCV 2024. F.R.S.-FNRS - Fonds de la Recherche Scientifique. Springer

D'Armenio E, Deliège A, Dondero MG (2024b) Semiotics of machinic co-enunciation: about generative models (Midjourney and DALL*E). Signata. Annales des Sémiotiques 15(15)

D'Armenio E, Dondero MG, Deliège A, Alessandro S (2025) For a semiotic approach to generative image AI: on compositional criteria. Semiot Rev. https://doi.org/10.71743/ee5nrx33

Deliege A, Dondero MG, D'Armenio E (2025) Revisiting Wölfflin in the age of AI: a study of classical and baroque composition in generative models. J Imaging. https://doi.org/10.3390/jimaging11050128

Desogus P (2012) The encyclopedia in Umberto Eco's semiotics. Semiotica. https://doi.org/10.1515/sem-2012-0068

D'Ignazio C, Klein LF (2020) Data feminism. The MIT Press, Cambridge

Eco U (1997) [1984] Semiotica e Filosofia Del Linguaggio. Einaudi

Eco U (2020) [1979] Lector in fabula: La cooperazione interpretativa nei testi narrativi. La nave di Teseo

Eco U (2020) [1990] I Limiti dell'Interpretazione. La nave di Teseo

Eco U (2024) [2003] Dire quasi la stessa cosa: Esperienze di traduzione. La nave di Teseo

Fabbrizzi S, Papadopoulos S, Ntoutsi E, Kompatsiaris I (2022) A survey on bias in visual datasets. Comput Vis Image Underst 223:103552. https://doi.org/10.1016/J.CVIU.2022.103552

Ginzburg C (2019) [1976] Il Formaggio e i Vermi: Il Cosmo Di Un Mugnaio Del '500. L'oceano Delle Storie, Adelphi

Ginzburg C (2023) [1979] Spie. Radici di un paradigma indiziario. Adelphi Edizioni, Milano, pp 157–202

Gramigna R (2023) Some remarks on fakes and deepfakes. A new epistemic challenge, Semiotica e intelligenza artificiale, vol 48. Aracne. pp 45–64

Guo P, Sun H, Xing S, Li J (2025) A study on the visual rhetorical differences in national image representation of China and the United States by generative artificial intelligence: an empirical analysis based on large multimodal models. J Inf Technol Politics. https://doi.org/10.1080/19331681.2025.2566181

Heuer CA, McClure KJ, Puhl RM (2011) Obesity stigma in online news: a visual content analysis. J Health Commun 16(9):976–987

Hudson DA, Manning CD (2019) GQA: a new dataset for real-world visual reasoning and compositional question answering. In: IEEE conference on computer vision and pattern recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019. Computer Vision Foundation/IEEE, pp 6700–6709

Impett L, Moretti F (2017) Totentanz: operationalizing Aby Warburg's 'pathosformeln'. Technical report, Stanford Literary Lab

Jacobs AZ, Wallach H (2021) Measurement and fairness. In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, FAccT '21, New York, NY, USA. Association for Computing Machinery, pp 375–385

Johnson J, Krishn R, Stark M, Li LJ, Shamma DA, Bernstein MS, Fei-Fei L (2015) Image retrieval using scene graphs. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 3668–3678

Khosla A, Zhou T, Malisiewicz T, Efros AA, Torralba A (2012) Undoing the damage of dataset bias. In: Proceedings of the 12th European conference on computer vision—volume Part I, ECCV'12, Berlin, Heidelberg. Springer, pp 158–171

Koch G, Kinder-Kurlanda K (2020) Source criticism of data platform logics on the internet. Hist Soci Res Historische Sozialforschung 45(3):270–287

Krishna R, Zhu Y, Groth O, Johnson J, Hata K, Kravitz J, Chen S, Kalantidis Y, Li L, Shamma DA, Bernstein MS, Fei-Fei L (2017) Visual genome: connecting language and vision using

crowdsourced dense image annotations. Int J Comput Vis 123(1):32–73. https://doi.org/10.1007/S11263-016-0981-7
Kuznetsova A, Rom H, Alldrin NG, Uijlings JRR, Krasin I, Pont-Tuset J, Kamali S, Popov S, Malloci M, Kolesnikov A, Duerig T, Ferrari V (2018) The Open Images dataset V4. Int J Comput Vis 128:1956–1981
Leone M (2024) Semiotics of the black box: on the rhetorics of algorithmic images. Vis Commun 23(3):426–451. https://doi.org/10.1177/14703572241247120
Li H, Zhu G, Zhang L, Jiang Y, Dang Y, Hou H, Shen P, Zhao X, Shah SAA, Bennamoun M (2024) Scene graph generation: a comprehensive survey. Neurocomputing 566:127052. https://doi.org/10.1016/J.NEUCOM.2023.127052
Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnik CL (2014) Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer Vision - ECCV 2014, Cham. Springer International Publishing, pp 740–755
Liu Z, He K (2025) A decade's battle on dataset bias: are we there yet? In: The thirteenth international conference on learning representations, ICLR 2025, Singapore, April 24–28, 2025. OpenReview.net
Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of international conference on computer vision (ICCV)
Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: a free stimulus set of faces and norming data. Behav Res Methods 47:1122–1135
McAuley J, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, SIGIR '15, New York, NY, USA. Association for Computing Machinery, pp 43–52
McPherson J (2003) Revisionist historians. Perspective on History
Miller GA (1995) WordNet: a lexical database for English. Commun ACM 38(11):39–41. https://doi.org/10.1145/219717.219748
Moretti F (2013) 'operationalizing'. New Left Rev II(84):103–119. https://doi.org/10.64590/daw
Peirce CS (1931–58) The collected papers of Charles Sanders Peirce [CP]. In: Hartshorne C, Weiss P, Burks AW (eds). Harvard University Press
Picca D (2025) Not minds, but signs: reframing LLMS through semiotics. arXiv preprint arXiv:2505.17080
Pisanty V (2015) From the model reader to the limits of interpretation. Semiotica 2015(206):37–61. https://doi.org/10.1515/sem-2015-0014
Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. In: Meila M, Zhang T (eds) Proceedings of the 38th international conference on machine learning, ICML 2021, 18–24 July 2021, Virtual Event, vol 139. Proceedings of machine learning research. PMLR, pp 8748–8763
Reyes-García E (2021) Face value: analyzing and visualizing facial data. Lexia-Rivista di semiotica
Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, Vertesi J (2019) Fairness and abstraction in sociotechnical systems. In: Proceedings of the conference on fairness, accountability, and transparency, FAT* '19, New York, NY, USA. Association for Computing Machinery, pp 59–68
Shankar S, Halpern Y, Breck E, Atwood J, Wilson J, Sculley D (2017) No classification without representation: assessing geodiversity issues in open data sets for the developing world. arXiv preprint arXiv:1711.08536
Shihadeh J, Ackerman M, Loker D (2025) What does genius look like? Investigating brilliance bias in AI-generated images. AI Soc. https://doi.org/10.1007/s00146-025-02752-6
Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, Borth D, Li L (2016) YFCC100M: the new data in multimedia research. Commun ACM 59(2):64–73. https://doi.org/10.1145/2812802
Torralba A, Efros AA (2011) Unbiased look at dataset bias. In: The 24th IEEE conference on computer vision and pattern recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011. IEEE Computer Society, pp 1521–1528
Volli U (2023) Il mito delle due intelligenze e la semiotica, Semiotica e intelligenza artificiale. 48:91–103. Aracne
Warren J, Weiss GM, Martinez F, Guo A, Zhao Y (2025) Decoding fatphobia: Examining anti-fat and pro-thin bias in AI-generated images. In: Chiruzzo L, Ritter A, Wang L (eds) Findings of the association for computational linguistics: NAACL 2025, Albuquerque, New Mexico. Association for Computational Linguistics, pp 4724–4736
Wilson B, Hoffman J, Morgenstern J (2019) Predictive inequity in object detection. arXiv preprint arXiv:1902.11097
Witte SP (1992) Context, text, intertext: toward a constructivist semiotic of writing. Writ Commun 9(2):237–308. https://doi.org/10.1177/0741088392009002003
Wu W, Protopapas P, Yang X, Michalatos P (2020) Gender classification and bias mitigation in facial images. In: 12th ACM conference on web science, WebSci '20, New York, NY, USA. Association for Computing Machinery, pp 106–114
Zeng B, Yin Y, Liu Z (2024) Understanding bias in large-scale visual datasets. In: Globerson A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak J, Zhang C (eds) Advances in neural information processing systems, vol 37. Curran Associates Inc, Red Hook, pp 61839–61871
Zhao J, Wang T, Yatskar M, Ordonez V, Chang K (2017) Men also like shopping: reducing gender bias amplification using corpus-level constraints. In: Palmer M, Hwa R, Riedel S (eds) Proceedings of the 2017 conference on empirical methods in natural language processing, EMNLP 2017, Copenhagen, Denmark, September 9–11, 2017. Association for Computational Linguistics, pp 2979–2989