Word Pictures: New Insights Through AI Around the Villa Laurentina by Pliny the Younger

Elena Merino-Gómez, Fernando Moral-Andrés, Blanca Querol, Pedro Reviriego, and Saverio Pisaniello

1 Introduction

The visual recreation of disappeared, unknown, or existing only in literary texts architectures has been a recurring exercise for theorists, painters, architects, and enthusiasts throughout history. The Tower of Babel (Condorelli et al., 2023; Moro, 2003), the Temple of Solomon (Lundquist, 2008; Fine, 2011), or the Labyrinth of Crete (Kotsonas, 2018; Herva & Rapakko, 2023) is part of the inventory of mythical architectures that have been known more precisely through texts than archaeological evidence. Profusely illustrated, also based on the texts that narrate them, are architectures and cities that oscillated between myth and reality until the discovery of their archaeological evidence transferred them, more or less fragmentarily, to the world of the tangible. The city of Troy (Bryant Davies, 2018), the Mausoleum of Halicarnassus (Jenkins, 2010), or the DomusAurea accounts for how the history of their graphic representation belonged, until their discovery, to the same imaginary plane inhabited by mythical or permanently disappeared

E. Merino-Gómez (✉)
Universidad de Valladolid, Valladolid, Spain
e-mail: elena.merino.gomez@uva.es

F. Moral-Andrés
Universidad Nebrija, Madrid, Spain
e-mail: fmoral@nebrija.es

B. Querol · P. Reviriego
Universidad Politécnica de Madrid, Madrid, Spain
e-mail: blanca.querol@upm.es; pedro.reviriego@upm.es

S. Pisaniello
MICROSCAPE architecture urban design ARCHITETTI ASSOCIATI. Università degli Studi di Ferrara, Ferrara, Italy
e-mail: info@microscape.it

F. Moral-Andrés et al. (eds.), *Decoding Cultural Heritage*,
https://doi.org/10.1007/978-3-031-57675-1_3

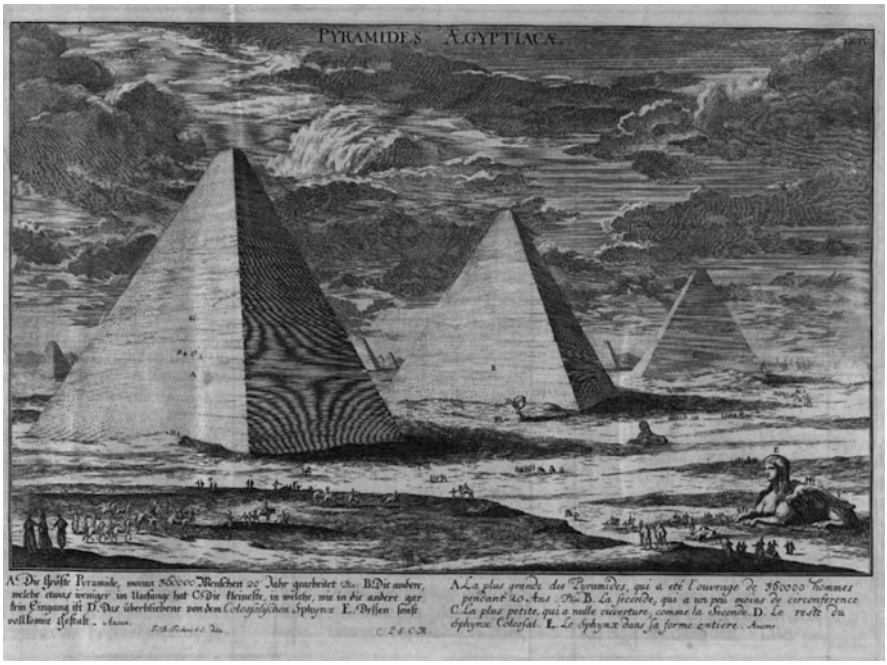


Fig. 1 Fischer Von Erlach. From *EntwurffEinerHistorischenArchitectur.* Plate IV. Sphynx and Pyramids of Giza

architectures. A similar situation of representing architectures through texts occurs when architectural theorists illustrate existing monuments they had never seen, also relying on literary sources that describe them. Fischer Von Erlach publishes illustrations, for example, of the Sphinx (Fig. 1), transcribing fragments from the texts of Pliny the Elder (1962) and Ausonius (1919) on which he bases his work. The scarcely reliable graphic result (Von Erlach, 1725, Pl. IV) discloses his lack of knowledge about the real appearance and proportions of the monument.

Among the most frequently discussed architectures based upon literary texts are the two villas that Pliny the Younger described in his letters to Gallo (Ep. 2.17. 1–29) and to other close friends (Plin. Ep. 5.6). These are the Laurentine, located near Ostia, and another one situated "in Tuscis." Since the early seventeenth century (Scamozzi, 1615, pp. 267–268; Félibien des Avaux, 1699, pp. 57–58), numerous graphic hypotheses have been continuously proposed for both buildings (Du Prey, 1994, pp. 57–78). In the early nineteenth century, a *concours d'émulation* was conducted at the École des Beaux Arts in Paris, inspired by the description of Laurentine in the letter to Gallo. This tradition underwent a revitalization in 1982 through a new competition led by notable figures, including Léon Krier, a prominent advocate of the avant-garde movement at that time.

Although it might have seemed that the traditional exercise had reached its limits with the *concours d'émulation* and the exhibitions in Paris in May 1982, followed

by those at the Musée des Beaux-Arts in Montréal from October to December 1983 (CCA, 1983), the truth is that the twenty-first century has witnessed a renewed interest in the subject. This revived interest is manifested in various initiatives, including three-dimensional digital reconstructions of Villa Laurentina (Miziołek, 2016, pp. 200–209), figurative graphic proposals (Russell Taylor, 2023), and the theoretical and visual exercises by Saverio Pisaniello (Pisaniello, 2008).

The main reason that has fueled the extensive history of conversion from the letters of Pliny to the visual representations of the Laurentine and Tuscan villas, whether in the form of plans, elevations, perspectives, models, and static or dynamic digital recreations, lies in the detailed description of their premises. In this case, we are not dealing with dwellings of widespread fame, like the Domus Aurea, nor with architectural feats or wonders of the world, such as the Pyramids. However, the descriptions of Plinian villas are possibly the clearest and most detailed in classical literature (Tanzer, 1924, p. 4). The temptation of graphic reconstruction, into which dozens of researchers have fallen in the last four centuries, is related to the particular elements of expression in the descriptions of the villas, which have allowed the identification of "virtual maps behind them" (Riggsby, 2003, p. 169).

However, despite its superiority in terms of clarity and detail compared to any other description from antiquity, Plinian narrative leaves numerous gaps for interpretation. Helen H. Tanzer, after analyzing all the solutions published until the 1920s, found them "strangely unlike" (Tanzer, 1924, p. ix). According to Montfaucon (1729, p. 129), referring to the Laurentine villa, "Il y a dans cette description bien des endroits difficiles à entendre".[1] While the French scholar refers to translation and interpretation issues with the vocabulary to justify the differences between his interpretation and that of Félibien des Avaux (1699), he ultimately concludes that "quand on traduit des descriptions aussi détaillées que celle-ci, et aussi pleines de mots extraordinaires, qu'on ne peut entendre qu'à demi, il faut souvent deviner malgré qu'on en ait; et quand on est reduit là, chacun devine à sa manière"[2] (Montfaucon, 1729, p. 129). It is precisely the need to "guess" that missing part, carried out "à samanière" by each researcher, that has resulted in such disparate outcomes in attempts to recreate Pliny's villas.

The emergence of generative artificial intelligence represents a new opportunity to update the classic exercise of translating into images the architectures described in texts. The exercise that Adolfo Natalini dismissed as "bizarre and anti-modern" (Pisaniello, 2008, p. 28) is, however, open to renewal through generative experimentation of images based on texts and other images.

[1] "In this description [of the Laurentine], there are many places that are difficult to understand." (Montfaucon, 1729, p. 129) Unless otherwise indicated, the rest of translations are our own. The orthographic conventions are respected in the transcription in the different languages specific to each period.

[2] "When translating detailed descriptions, especially those filled with extraordinary words that can only be half-understood, one often has to guess despite having them; and when one is reduced to that point, each one guesses in their own way." (Montfaucon, 1729, p. 129)

For the observation of the AI's performance in its ability to recreate images based on descriptions, we have focused on the Villa Laurentina. The text lists up to 38 distinct rooms (Riggsby, 2003, p. 170), in addition to other transitional spaces, courtyards, and outdoor areas. While the description of the Villa Tuscana is more extensive, the one dedicated to the Villa Laurentina is more explicit from an architectural standpoint (Du Prey, 2018, p. 471); therefore, it should be more easily translatable into images.

2 From Text to Image Through Artificial Intelligence

Artificial Intelligence has been part of computer science for decades and used in many applications mostly related to classification and pattern recognition (Winston, 1984). One classical example is the recognition of hand-written digits, in which a model is trained with a large dataset of digits to learn to recognize new digits presented to the model at a later time (Deng, 2012). There are many models that have been developed to learn, being Neural Networks possibly the most popular one. Neural networks resemble the behavior of the human brain by using basic elements that try to emulate neurons that are connected in layers (Jain, 1996). Initially, the number of neurons and layers was low, the data available to train the networks was limited, and neural networks were only capable of solving simple problems. However, as larger datasets and more powerful computing platforms (Sze, 2017) start to become available, larger and deeper networks can be trained to solve more complex problems. For example, Graphical Processing Units (GPUs) originally intended to accelerate computer graphics have played a key role to accelerate the execution of large neural networks (Peddie, 2023), and huge datasets with billions of images and captions are now publicly available to train large models (Schuhmann, 2022).

Initially, those large networks were still mostly used to recognize objects or classify images. However, as the networks can potentially be trained to produce any desired output, schemes that can generate images using neural networks started to appear in the form of Generative Adversarial Networks (GANs). GANs combine two networks: generator and discriminator. The generator tries to produce images that the discriminator cannot differentiate from real images. As both networks learn, the images produced by the generator get closer to the real images (Creswell, 2018). A GAN can be trained to generate, for example, faces that are quite realistic. More recently, other algorithms known as diffusion have been developed to generate images. They are also based on complex neural networks and are capable of producing highly realistic images (Rombach, 2022).

Another important development was the creation of neural network-based systems that combine image and text, being able to identify, for example, texts that correspond to images (Radford, 2021). These systems can also be used to guide an image generator, for example, a GAN to generate images that correspond to a given text. This led to the explosion of AI-based text to image generators with

the introduction of DALL-E (Ramesh, 2021), MidJourney, Stable Diffusion, or Leonardo Ai. These tools are capable of generating images that represent a given text and are now used by millions of persons in different areas that include graphical design, art (Anantrasirichai & Bull, 2022), architecture (Ploennigs, 2023), or video games. They have also been used to complete unfinished architectures (Merino Gómez, 2023).

These text-to-image tools are pre-trained on huge datasets of text and image pairs and are capable of generating a wide range of images. However, in some applications, it is beneficial to generate a specific type of images. This can be done by performing a fine-tuning of the pre-trained models (Ruiz, 2023) using a small dataset of tens or hundreds of images. This feature is now supported by tools such as Leonardo AI and opens new possibilities to reproduce specific architectural styles.

One potential use of these AI text to image generation tools that to the best of our knowledge remains unexplored is the recreation of architectures for which only a textual description is available. In the rest of this work, we pursue this study focusing on the Villa Laurentina described by Pliny in his letter to Gallo.

3 The Elusive Villa

Despite the impossibility of recreating the exact materiality of Pliny's villas, considering that they might have been more than just an exercise in the epistolary genre (Barbiero, 2023; Du Prey, 2018, p. 467), some researchers have attempted to construct synthetic solutions taking into account all possible parameters. Robert Castell (1728), for example, investigated all surviving Latin textual sources mentioning any facet of the Roman villa. He ultimately provided graphic documentation coherent with the compiled information (Fig. 2). Almost two centuries later, Helen H. Tanzer (1924) critically compiled the graphic tradition of all previous authors who had addressed the issue to present her synthetic version of the result.

Both synthesis processes involve operations similar to what artificial intelligence can handle simultaneously and automatically today. On the one hand, text-to-image generation processes can generate images from prompts, which are textual sequences. On the other hand, it is possible to train and customize a pre-trained AI through fine-tuning with a reduced set of images to guide and substantiate the results based on existing graphic solutions. Once the fine-tuning is completed, both with textual and image data, it is possible to integrate both processes into a single one by introducing prompts to operate through the model customized with the image dataset. In this way, one could expect a synthetic solution that combines existing textual information with graphic information generated by researchers over the centuries.

Even in the design of the experimentation methodology, significant limitations become evident, both concerning the creation of prompts and the selection of images. The first difficulty lies in the language used to formulate prompts. The understanding that image generators have for complex Latin expressions is still

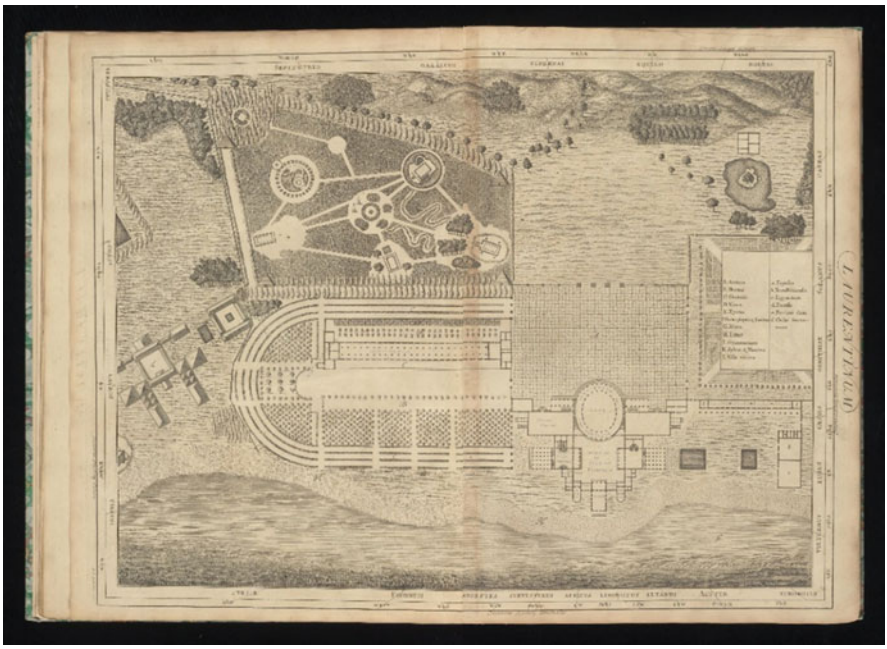


Fig. 2 Robert Castell (1728), The Laurentine from *The villas of the ancients illustrated*

very limited, as shown in Figs. 3 and 4. This compels us, in this initial approach, to discard the original language and consequently opt for a translation. However, the nuances of Latin, the challenges of an uncommon lexicon, and the different semantic contents that this lexicon has had throughout the history of architecture give rise to dozens of interpretative variations in all translations undertaken so far. Montfaucon (1729, p. 129) asserted that he was "sure that if ten skilled people did it [the translations of the Epistles by Pliny the Younger] separately, there would not be one that agreed entirely with the other".[3]

By choosing a version in a language other than Latin, the interpretative bias of the author who translated texts, not precisely simple given the vast amount of "mots extraordinaires" used by the original author, is being introduced. Fully aware of this limitation, we have chosen the English language and have referred to the reference translation by William Melmoth, present in the esteemed Loeb Classical Library collection (Pliny the Younger, *Letter to Gallus*, 1915b), to derive functional prompts for our purposes.

Another significant limitation lies in the length of the textual sequence recognized by the software. Leonardo Ai only admits prompts of 1000 characters, and

[3] "[. . .]; je suis sur que si dix hábiles gens en faisaient à part la traduction, il n'y en auroit pas un que convint en tout avec l'autre." Montfaucon (1729, p. 129)

Fig. 3 Result from Leonardo-Ai, from Stable Diffusion-Dream Shaper v7 after entering the prompt: "Cuius in prima parte atrium frugi, nectamensordidum; deindeporticus in D litteraesimilitudinemcircumactae, quibusparvolasedfestiva area includitur. Egregiumhacadversustempestatesreceptaculum; namspecularibus ac multomagisimminentibusrectismuniuntur." ("The entrance hall is plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. This affords a capital retreat in bad weather, as it is sheltered by glazed windows, and much more by overhanging eaves." Pliny the Younger, Ep. 2.17. 4 translated by William Melmoth in Pliny the Younger (1915a)) Pliny the Younger, *Ep*. 2.17. 4

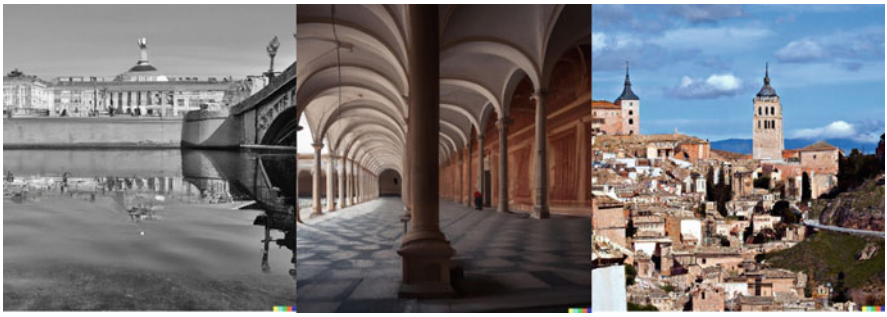


Fig. 4 The result from DALL-E2, from OpenAI, after entering the same prompt as in Fig. 3

DALL·E2, of 400, which, in the English language, average around 60 and 150 words, respectively. The translated version by Melmoth of the letter to Gallus describing the Laurentine contains 2030 words and 11,190 characters, values that are well beyond the specified range.

The syntactic and semantic complexity of the text will also hinder the obtaining of precise outputs. Prompt engineering, among other strategies, advises splitting complex tasks into simpler subtasks, summarizing long documents piecewise and constructing a full summary recursively (OpenAI, 2024), which will undoubtedly introduce variations conditioned by the ways of fragmenting and simplifying the prompts. In this chapter, we will be working with the customization of the Stable

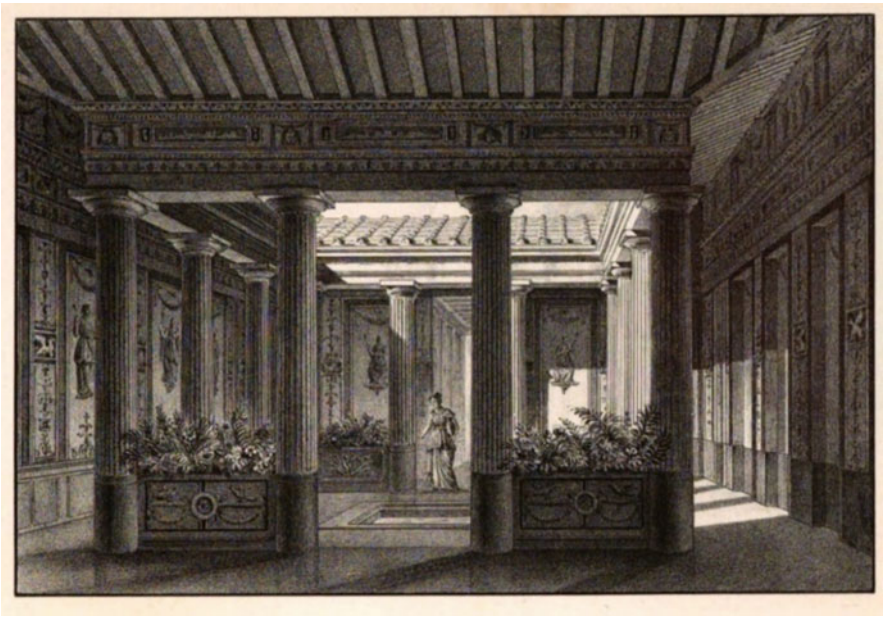


Fig. 5 Partial view of the Villa Laurentina, depicting the peristyle according to the reconstruction proposed by Haudebourt (1838, p. 102)

Diffusion v1.5 tool, given its ability to handle the insertion of more extensive prompts.

It is also important to consider that the necessary fragmentation of the text for the insertion of prompts should imply that the resulting outputs will be fragmentary as well. If, for example, the text "[. . .] entrance-hall, plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. Sheltered by glazed windows and much more by overhanging eaves" (Plin. *Ep*. 2.17.4) is introduced, we are providing a partial description that would, therefore, only allow obtaining partial results of the villa, in line with what some authors like Haudebourt (Fig. 5) or Bouchet (Fig. 6) contributed for different parts of the residence.

When addressing the selection of images for training, various difficulties also concur. The first limitation arises when customizing the tool. To do this, the graphic proposals for the reconstruction of the Laurentina have been listed, from the first version by Scamozzi (1615, pp. 267–268), through all those appearing in the renowned work by Du Prey (1994), to the current update with proposals identified after the year 1994. In total, more than 50 reconstructions have been considered, many of which are represented in various graphic documents, such as plans on different levels, plans with and without garden surroundings, elevations, sections, perspectives, as well as some interior views and models. All this results in a large number of images—more than 200—which also exhibit great heterogeneity.

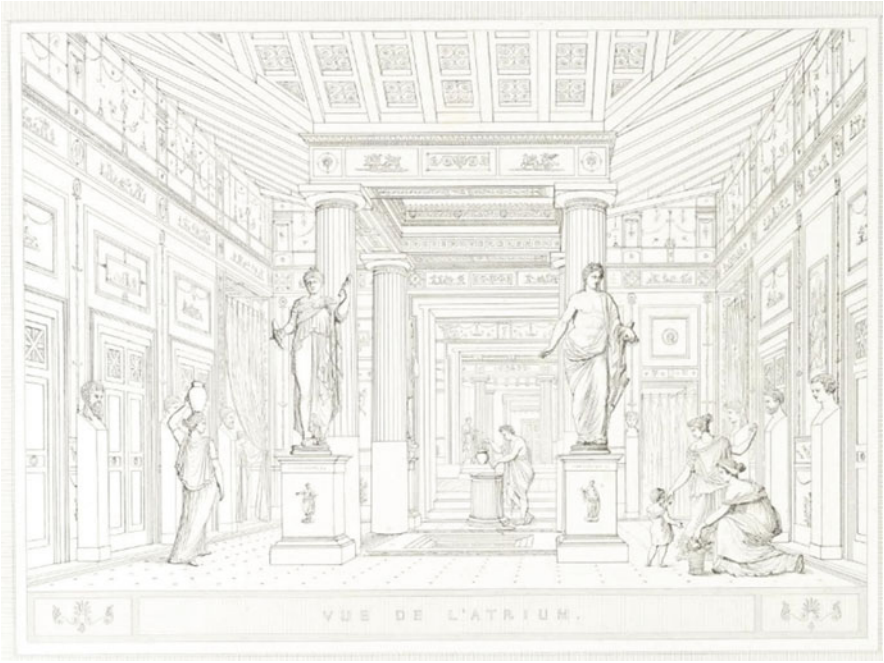


Fig. 6 Partial view of the Villa Laurentina, depicting the atrium according to the reconstruction proposed by Bouchet (1852, p. 41)

For the customization of models in Stable Diffusion v1.5, it is only possible to work with a dataset of a maximum of 40 images. The limitation of 40 images, from the outset, prevents training with all versions of the Laurentina reconstruction throughout history, unlike what Helen Tanzer did in the aforementioned graphic analysis. In addition, there are significant differences in the characteristics of the images. Their heterogeneity introduces noise, making it challenging for the tool to operate accurately once customized. For these reasons, two subsets of images with different criteria of homogeneity are chosen for customization:

- *First customization (floor plans):* A first subset of 30 images is chosen, featuring the main floor plan of the Villa Laurentina as interpreted by the following authors:

  Scamozzi (1615)
  Félibien des Avaux (1620)
  Robert Castell (1728)
  Potoki (1779)
  Márquez (1796)
  Lepage (1818)
  Macquet (1818)
  Normand (1818)

Hirt (1827)
Stier (1832)
Haudebourt (1838)
Canina (1840)
Schinkel (1841)
Bouchet (1852)
J.G.S. Jr.: Unidentified Columbia student (1864)
Loring (1864)
Cowan (1889)
Winnefeld (1891)
Magoun (1895)
Tanzer (1921)
Van Buren (1943)
Adam (1981)
Montes (1981)
Santelli (1981)
Sunderman (1981)
Treuttel (1981)
Weber and Laroche (1981)
Ricotti (1984)
Civitelli (2017)
Russell and Taylor (2020)

- *Second customization (general views):* A second subset of 30 images is selected, comprising general perspectives of the villa, derived either from drawings or models, by the following authors:

Krubsacius (1760)
Hirt (1827)
Haudebourt (1838)
Canina (1840)
Bouchet (1852)
Unidentified Author from Accademia di San Luca Rome (1855)
Pember (1947) (One black and white picture of the model and one colored picture of the same model)
Farina (1981) (One colored picture of a model)
Haut (1981)
Ustarroz and ´iñiguez (1981)
Krier (1981) (Five black and white perspectives, one colored perspective, and four pictures of the model)
Fernando Montes (1981) (One black and white axonometry and three colored pictures from the model)
Sunderman (1981)
Jean-Pierre Adam (1981)
Serge Santelli (1981)
Treuttel (1981)
Hardy (2014)

For the first training subset, the 30 images of the main floor plan, each from different authors, have been edited to a square format and converted to black and white so that the different colors do not distract or create confusion during the training, leaving us with a more consistent group of images. The example prompt used for training was "A floor plan." For the second subset, another 30 images are selected, this time from 17 authors, offering different views for some of them. They have also been edited in a square format, and watermarks have been removed from those documents that had them. The example prompt was "Roman villa." The images from both subsets have a resolution of 768 × 768 pixels.

4 Results: Before and After Training

4.1 Recreation Using Generic Models

In order to conduct an analysis of the effectiveness of the trained model, it will be necessary to compare the results with the performance of the generic model, i.e., without customization. Two situations are analyzed: the first one with literal prompts that partially describe the house and the second one with a general prompt summarizing the villa, as the length of the complete description by Pliny is not accepted by Stable Diffusion v1.5

The first partial literal prompt inserted into the generic model is: "An ancient Roman entrance-hall, plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. Sheltered by glazed windows and much more by overhanging eaves" (Plin. *Ep*. 2.17.4). The phrase "An ancient Roman" is added to the literal sentence from the epistle to provide context. With the literal prompt, without the remark "An ancient Roman . . . " it has been verified that the AI generates contemporary "entrance halls."

The results from Fig. 7 demonstrate that the generated images do not go much beyond the creation of an entrance hall. The portico from the textual description does not appear in any of them, and the understanding of "glazed windows and much more by overhanging eaves" is also very limited. Substituting the word "entrance hall" in the prompt, whose contemporary meaning may be distorting, with the Latin word "atrium," also present in the specialized English lexicon of historical architecture,[4] does not yield much better results (Fig. 8).

The non-customized model manages to understand that the word "atrium" corresponds to a space, usually uncovered. However, it omits in two out of the three outputs the characteristic of glazing and, in all three cases, as in the previous

[4] The word has been present in the Oxford English Dictionary since 1577 (OED, 2023): **atrium**, n. a court. The central hall or court of a Roman house; [. . .] a court. In a public building, a usually skylit central court rising through several storeys and surrounded by galleries at each level with rooms . . .

Fig. 7 Results with the prompt corresponding to Plin (*Ep*. 2.17.4) with the non-customized model

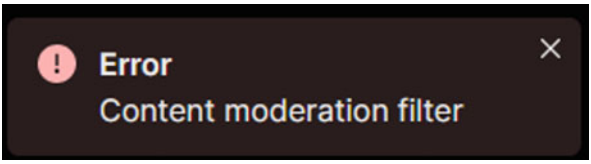
Fig. 8 Results with the prompt corresponding to Plin. *Ep*. 2.17.4 with the non-customized model, replacing "entrance hall" with "atrium"

example, a substantial part of the description referring to the "portico in the form of the Letter D" and the "overhanging eaves" that agreeably shelter the portico.

An experiment is also conducted with a general prompt of less than 1000 characters that summarizes the main parts of the villa: "An ancient Roman villa from the first century by Pliny the Younger. Simple atrium to D-shaped portico with a cozy area. Sheltered, pleasant hall opens to a sea-facing dining room with panoramic views of seas, court, and distant woods. Sunlit rooms with sea views. Hall angle boosts warmth. Circular chamber with bookcase cupboard. Cozy bedchamber. Side for slaves, yet guest-friendly. Sunlit parlour with sea views, connected rooms

Fig. 9 Results with the summary derived from the comprehensive description of the villa, with the model still untrained

Fig. 10 Policy warning



for year-round comfort. Spacious bath area, ball court. Turrets above over sea views. Sheltered dining room overlooks the garden, insulated from sea roar. Lush garden with trees. Banqueting room boasts views. Adjoining apartments overlook villa entrance and kitchen garden. Spacious gallery with double windows facing the sea. Sun-warmed terrace defends against winds, offers shade in summer. Open windows ensure air circulation. Charming detached building with sea-view winter room, versatile chamber, and bed-chamber insulated from noise."

Once again, the non-customized model provides unconvincing results (Fig. 9), being unable to respond to the multiple parts of the prompt. As in the previous case, it recognizes only the first words of the description and only in a very limited way. It also introduces architectural elements unrelated to the villa architecture of the first century, such as wrought-iron railing and domestic furniture, among many other details. From the recognition of the initial part of the text, even when inserting a summary of the complete description, a partial output of the building is derived.

It is also noteworthy the limitation due to the application's policy when the summary text is introduced. The inclusion of the word "slaves" prevents the generation of images, issuing an error message: Fig. 10. It is necessary, therefore, to replace the word "slaves" with "servants" to obtain results.

Fig. 11 Results with the model trained on the first dataset: "floor plans"

4.2 Customized Model

The results for the same prompts with custom-trained models are examined below. First, we input the partial prompt referring to the vestibule through the customized model with floor plans: "An ancient Roman atrium, plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. Sheltered by glazed windows and much more by overhanging eaves" (Plin. *Ep*. 2.17.4). It would be expected that a model customized with floor plans would be able to handle examples of vestibule layouts. In Fig. 11, it can be observed that the AI has merely imitated the chromatic and graphic features of the inserted drawings for training, without correctly addressing the floor plan content they represent. The result, however, consists of elevations, with and without perspective, in which the concept of the entrance to a dwelling predominates.

If we force the prompt by adding "**An architectural plan** of an ancient Roman atrium . . . ," the result is that of general floor plans of the whole, without attribution of any criteria for defining spaces, based on those used in the model training. In no case does it adhere to the recreation of an atrium in the context of a Roman dwelling: Fig. 12.

Subsequently, in the same way, the second model, trained with "general views," is requested to produce images related to the vestibule (Fig. 13) and the entire villa (Fig. 14), using the same prompts as in the previous case. In Fig. 13, the trained model ignores the partial description introduced and provides a general view. The solutions entail a simple generic emulation of the graphics and perspective views, formally related to what was used for training.

In Fig. 14, we observe results that, at first glance, may seem more coherent; however, once again, the model generates a complete villa that does not follow the description in the text. Instead, it introduces formal variations using the most frequent graphic styles in the training. For instance, a conspicuous feature like the portico in the form of the letter D does not appear in any of the generated images despite the fact that both datasets contain various solutions representing the semi-

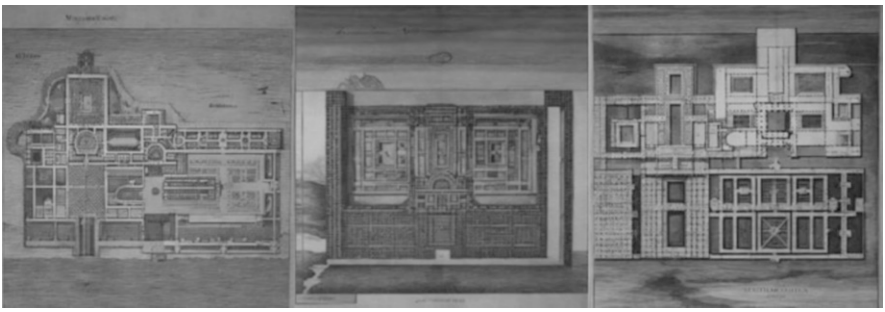


Fig. 12 Results with the model trained using the first dataset, "floor plans," requesting a partial floor plan of the vestibule

Fig. 13 Results with the model trained using the "general views" dataset, requesting a partial image of the vestibule according to the description: "An ancient Roman atrium, plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. Sheltered by glazed windows and much more by overhanging eaves" (Plin. *Ep*. 2.17.4)

circular portico resembling the letter D. Although one-third of the input images of the training "general views" correspond to photographs of models, in the more than 20 tests conducted, the model never returned images of a photographic nature or representing models.

Many other visual features are detected in the generated images that are unrelated to the textual description. For example, in the third and fifth images of Fig. 14, the villa appears inscribed in a circle and a hexagon, respectively, a feature completely extraneous to the content of the text. All generated images follow a diagonal composition, even though more than a third of the training images provided had frontal views without a dominant diagonal line. The results, therefore, remain purely formal, with a predominant tendency, as expected, to reproduce the characteristics of the images with greater rate of recurrence in the dataset and with a residual impact derived from the textual instructions.

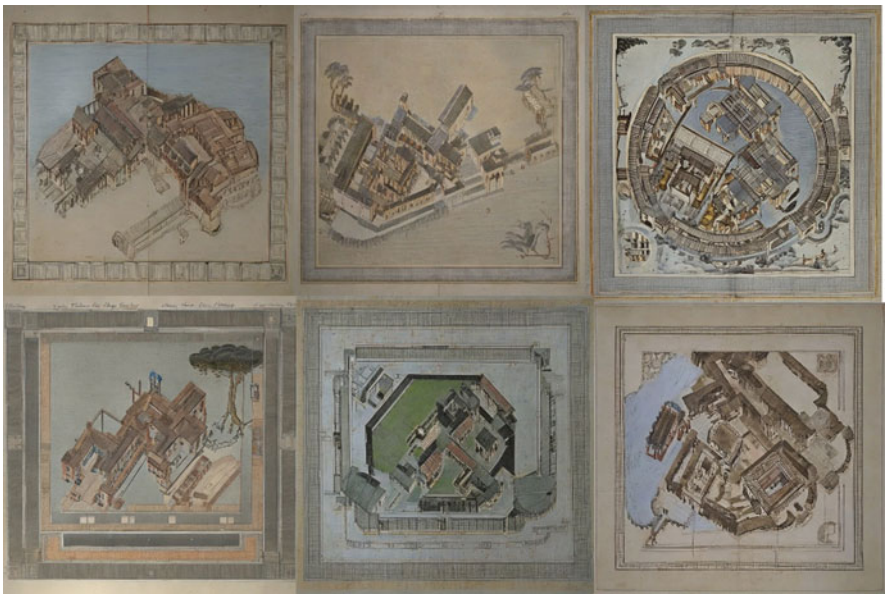


Fig. 14 Results with the model trained using the "general views" dataset, requesting a global image of the villa according to the elaborated 1000-word synthesis prompt

5 Discussion

The results generated by the AI after customizing the model show significant limitations in simultaneously addressing the textual nature of the prompts and the graphics provided during training. Once the model is customized, visual data generally prevails over textual data. When the influence of the prompt is detected in the results, as is the case with the general inputs in Fig. 11, it is very restricted, limited to the first words of the prompt: "An ancient Roman atrium . . . " Additionally, the images respond defectively to the lexicon, representing "entrance doors," which are exterior elements, instead of "atria," which are interior spaces within the dwelling, although near the entrance. It is observed that, due to the absence of partial *atrium* images in the training data, the model simplifies the term, assimilating it to an entrance door, a much more common architectural element.

Unlike what happened in the case of Fig. 8, the model does not generate architecture with "ancient Roman" features; instead, it produces elements reminiscent of an undefined neoclassicism. All the results obtained are in black and white, consistent with the color range of all the images introduced in the "floor plans" training. They exhibit a drawing aesthetics typical of the representation codes of elevations from the eighteenth and nineteenth centuries, which is also broadly based on the outline of the floor plans introduced for training.

When the prompt for the "ancient Roman atrium" was forced by completing the context with "**An architectural plan** of an ancient Roman atrium...," the model did not grasp the specificity of the space and simply imitated the general floor plans (Fig. 12). The only difference compared to the images generated using the "general views" training is that, in this case, the floor plans are generated with an orthogonal arrangement relative to the frame of the plan. This is because in the "floor plan" training dataset, 85% of the plans had the same arrangement. In the case of attempting to generate atriums through the "general views" customization, general floor plans are also generated, with no specificity dedicated to the atriums. The difference this time is the emulation of the predominant visual features of the views: polychromy and diagonal composition (Fig. 13).

The capabilities of generative AI, even when trained with highly specific visual subjects, are limited to producing variations on the introduced models, devoid of any criteria beyond the generation of diverse images. Every author who has addressed the topic, since the seventeenth century, has overlaid their own interpretative experience of the text onto their background of images seen or projected throughout their training or profession. As a result, different proposals have always been generated. However, AI does not filter the results based on different interpretations. The reading of the text is always identical to itself, and the goal is varied production. This pursuit of variation for variation's sake is guided by its fundamental aims of never generating the same image twice, mirroring the way Large Language Models (LLMs), when presented with the same prompt, generate different texts.

Leon Battista Alberti consistently expressed concerns throughout his work regarding the challenges in the transmission, reproduction, and generation of images, always subordinate to the author's arbitrariness, whether copying someone else's drawing or depicting reality. His attempts at a univocal identification of the image become evident in the numerical encoding of the plan of Rome in his *Descriptio Urbis Romae* (Alberti, 2000). Alberti's goal of consistently generating the same image through a pre-digital age digitization (Carpo, 2008) was exactly the opposite of generative AI, which systematically proposes different results. The text that has reached us with the description of the Laurentine Villa narrates an extremely complex program, difficult to parameterize, with significant stylistic and formal gaps, making its complete and unambiguous reconstruction challenging. However, one might expect from an automated system, such as generative AI, that at least those parts described in greater detail in the text would be minimally addressed. The combination of textual prompts and previous training images should yield results that address both aspects, if not simultaneously, at least in a balanced manner. However, the bias towards the image is prevalent, and the generated images do not correspond to the information provided by Pliny's epistolary ekphrasis.

The proposed exercise aimed to have the AI interpret prompts in a manner similar when approaching the design of an architectural project based on a written account: a program of needs is described, some formal aspects are hinted, and, through visualizing previous representation forms, the ways of drawing the architecture to be designed could be constrained. Judging by the results, it can be stated that, at this point, AI is not capable of satisfactorily responding to the development of a specific

architectural program through its textual description. The design methodology, which addresses both textual and visual aspects, both total and partial, is mimicked, albeit rudimentarily, through the fragmentation of the prompts, with not much better outcomes.

6 A Last Attempt

The history of recreations of the Villa Laurentina evolves with the trends in architecture, as demonstrated by the proposals of diverse authors such as Justo Solsona, Fernando Montes, or Serge Santelli (Pinon & Culot, 1982, pp. 206–214). In the twenty-first century, the work of Saverio Pisaniello revisits the theme with a metaphysical approach. The architect pays homage to the words of Franco Rella, "we never have a direct route that leads us to the ancient" (Pisaniello, 2008, p. 27), through a proposal that is unique and different from all others, as if it were artificial intelligence. However, it is the result of a thorough and intimate reading and understanding of the text—an aspect that, as demonstrated in this chapter, AI is still far from achieving.

In a final twist, a last training session of Stable Diffusion v1.5 is conducted with a dataset of 40 images from Saverio Pisaniello's work *Esistenza Minima* (2008). His visual production includes readings of fragmented spaces, as his subtitle announces: "stanze, spazîdellamente, reliquario," as well as global solutions. The fragmentation, necessary for architectural projection and the comprehension of any text, is ultimately encompassed in its overall result. All of this is presented in a wide variety of visual formats, such as floor plans, sections, models, partial views, and interpretations that are incorporated into the training: Fig. 15.

The results only serve to confirm everything explained in the previous sections. In Figs. 16 and 17, the outcomes corresponding to the partial prompt "An ancient Roman atrium . . . " and the one summarizing the entire villa with a 1000-word text are presented, respectively. As seen in Fig. 11, the model mistakenly associates the

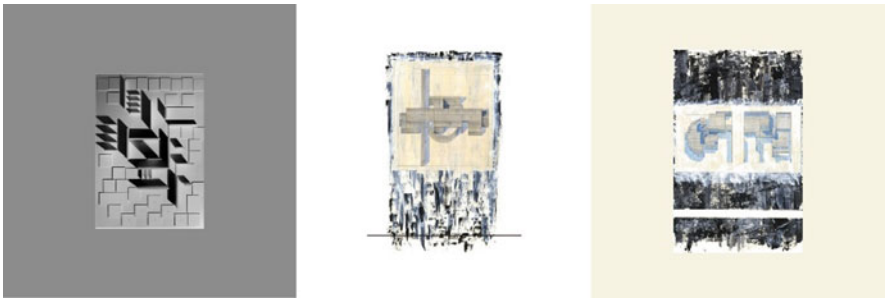


Fig. 15 Three of the 40 original images from Saverio Pisaniello's text *Esistenza Minima*. They have been standardized to the square format required by training at sizes of 768 × 768 pixels

Fig. 16 Images generated after training with the dataset of Saverio Pisaniello and the prompt: "An ancient Roman atrium, plain, but not mean, through which you enter into a portico in the form of the Letter D, which includes a small, but agreeable area. Sheltered by glazed windows and much more by overhanging eaves" (Plin. *Ep*. 2.17.4)

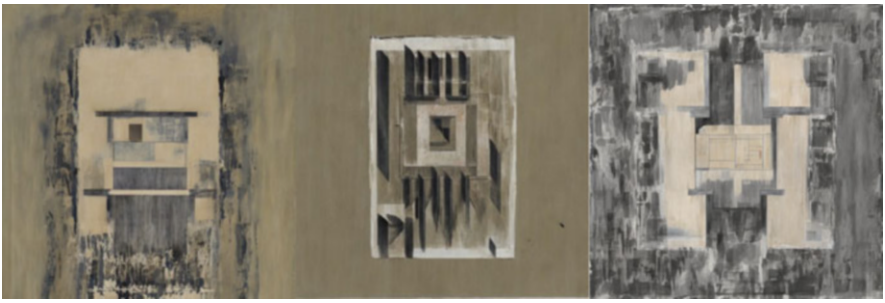


Fig. 17 Images generated after training with the dataset of Saverio Pisaniello and the 1000-word prompt summarizing the Laurentine

concept of "atrium" with that of an entrance. It merely mimics an entrance elevation, failing to respond to significant parts of the description such as the "portico in the form of the Letter D" or the "overhanging eaves." The results in Fig. 16 differ from those in Fig. 17 in that they, at least, present a partial view, while those in Fig. 17 offer a global vision of the villa, albeit consistently lacking its main features. It is noteworthy that there is a complete absence of any visual reference to an "Ancient Roman villa from the first century . . . ." The dominant force of visual representation results in outcomes that completely disregard even the initial words of the prompt, which are usually the ones that the text-to-image translation responds to.

7 Conclusions

"Cette maison insaisissable," the Roman villa referred to by Pinon (1982, p. 11), as well as described by Pliny the Younger, remains as elusive in the present as throughout its entire history of attempts at graphic formalization. The opportunity

presented by AI to translate texts into images and to influence the visual aspects of those images seems to herald new ways of embodying, conceiving, projecting, completing, or reconstructing architectures. However, the generative capacity in this field is still limited.

The obtained results are no less speculative than the long tradition of emulating the Laurentine, and they also present notable interferences and comprehension errors on the part of Stable Diffusion. The translation of prompts into images is confined—and not always—to the initial terms of the textual chain, and there is a general tendency towards semantic simplification when highly specific terminology is introduced, distorting the final outcomes.

This rudimentary and flawed understanding of AI currently hinders progress beyond the realm of experimentation or the generation of very general sketches. Therefore, it is incapable of defining an architectural program with minimal complexity or addressing a program of needs for the generation, for instance, of an architectural project.

It becomes evident, likewise, the need to advance in prompt engineering to achieve better results through clear textual structures, allowing us, perhaps 1 day, to approach the skill of Mercury, capable of being clear and perfectly understandable, "without any sign of hands: only with words."[5]

Acknowledgments This work has been possible in part thanks to the FUN4DATE project (PID2022-136684OB-C22) funded by the Spanish AgenciaEstatal de Investigación (https://doi.org/10.13039/501100011033).

References

Alberti, L. B. (2000). *Descriptio Urbis Romae.* (M. Furno, M. Carpo, Edits., M. Furno, & M. Carpo, Trads.). Droz.

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review, 55*, 589–656.

Ausonius. (1919). *Ausonius* (Vol. 1) (H. G. Evelyn White, Trad.). William Heinemann.

Barbiero, E. A. (2023). Let's get real: Ekphrasis, reality and fiction in Pliny's epistles. In C. R. Jackson & J. Solso (Eds.), *res vera, res ficta: Fictionality in Ancient Epistolography*. De Gruyter. https://doi.org/10.1515/9783111308128-010

Bouchet, J.-F. (1852). *Le Laurentin: maison de campagne de Pline-le-consul restitué d'après sa lettre à Gallus*. Chez l'auteur, rue de Seine, 18, et chez les principaux marchands d'estampes.

Bryant Davies, R. (2018). *Troy, Carthage and the Victorians: The Drama of classical ruins in the nineteenth-century imagination*. Cambridge University Press.

Carpo, M. (2008). Alberti's media lab. In C. Mario & F. Lemerle (Eds.), *Perspective, projections & design. Technologies of architectural representation* (pp. 47–64). Routledge.

Castell, R. (1728). *The Villas of the ancients illustrated*. Author's Own Expense.

[5] "Mercuriumferuntvelmaximeobhanc remdivinumhabitumquodnullosignomanus: sedsolisverbis quae diceretitadiceretut plane intelligeretur". (Alberti, 1512, p. Lib. VI. VII)

CCA. (1983). *The Villas of Pliny and classical architecture in Montréal*. Canadian Centre for Architecture. Retrieved from https://www.cca.qc.ca/en/events/2798/the-villas-of-pliny-and-classical-architecture-in-montreal

Condorelli, F., Tramelli, B., & Luigini, A. (2023). Digital Turris Babel. Augmented release of Athanasius Kircher's Archontologia. In A. Giordano, M. Russo, & R. Spallone (Eds.), *Beyond digital representation* (pp. 83–95). Springer. https://doi.org/10.1007/978-3-031-36155-5_6

Creswell, A. W. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine, 35*(1), 53–65.

de Montfaucon, B. (1729). Le laurentin, maison de campagne de Pleine le jeine, décrit par lui-même. Chapitre XIV. In *L'Antiquité expliquée, et représentée en figures. Tome troisième. Les usages de la vie. Première partie* (pp. 125–129). Chez Florentin Delaulne et autres. Retrieved from https://arachne.uni-koeln.de/Tei-Viewer/cgi-bin/teiviewer.php?manifest=BOOK-804114

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine, 29*(6), 141–142.

de Du Prey, P. (1994). *The Villas of Pliny from antiquity to posterity*. University of Chicago Press.

de Du Prey, P. (2018). Conviviality versus Seclusion in Pliny's Tuscan and Laurentine Villas. In A. Marzano & G. P. Métraux (Eds.), *The Roman Villa in the Mediterranean Basin* (pp. 467–475). Cambridge University Press.

Félibien des Avauxn, J.-F. (1699). *Les plans et les descriptions de deux des plus belles maisons de campagne de Pline le consul*. Chez Florentin & Pierre Delaulne ruë S. Jacques, à L'empereur & au Lions d'or.

Fine, S. (2011). *The Temple of Jerusalem: From Moses to the Messiah*. Brill.

Haudebourt, L.-P. (1838). *Le Laurentin, maison de campagne de Pline de jeune; resstituée d'après la description de Pline*. Chez Carilian Goeury.

Herva, V.-P., & Rapakko, J. (2023). Insides, outsides and the labyrinth: Knossos, palatial space and environmental perception in Minoan Crete. *Journal of Social Archaeology, 23*(3), 264–285. https://doi.org/10.1177/146960532311867

Jain, A. K. (1996). Artificial neural networks: A tutorial. *IEEE Computer, 29*(3), 31–44.

Jenkins, I. (2010). The Mausolea of Halicarnassus. Bulletin of the Institute of Classical Studies. Supplement. *Exploring Ancient Sculpture: Essays in Honour of Geoffrey Waywell, 104*, 121–1305. Retrieved from https://www.jstor.org/stable/44214902

Kotsonas, A. (2018). A cultural history of the Cretan Labyrinth: Monument and memory from prehistory to the present. *American Journal of Archaeology, 122*(3), 367–396. https://doi.org/10.3764/aja.122.3.0367

Lundquist, J. M. (2008). *The temple of Jerusalem: Past, present, and future*. Praeger.

Merino Gómez, E. M. (2023). Unfinished architectures: A perspective from artificial intelligence. *EGA Expresión Gráfica Arquitectónica, 28*(48), 254–267.

Miziołek, J. (2016). *The Villa Laurentina of Pliny the younger in an eighteenth-century vision*. "L'Erma" di Bretschneider.

Moro, S. (30 de April de 2003). *Re-constructing Babel. The history and creative possibilities of a myth, explored through text and installation*. University of Central Lancashire.

OED. (2023). *Oxford English Dictionary*. Retrieved from https://www.oed.com/

OpenAI. (7 de January de 2024). *Prompt engineering*. Retrieved from https://platform.openai.com/docs/guides/prompt-engineering

Peddie, J. (2023). *The history of the GPU-eras and environment*. Springer Nature.

Pinon, P. (1982). *La Laurentine et l'invention de la maison romaine* (pp. 11–51). (P. Pinon, & M. Culot, Eds.). Paris: Éditions du Moniteur.

Pinon, P., & Culot, M. (1982). *La Laurentine et l'invention de la villa romaine* (P. Pinon, & M. Culot, Eds.). Éditions du Moniteur.

Pisaniello, S. (2008). *Esistenza minima. Stanze, spazî della mente, reliquiario*. Firenze University Press.

Pliny, t. Y. (1915a). Letter to Domitius Apollinaris. In *Letters* (Vol. 1) (W. Melmoth, Trad., pp. 377–396). William Heinemann.

Pliny, t. Y. (1915b). Letter to Gallus. In *Letters* (Vol. I) (W. Melmoth, Trans., pp. 150–165). William Heinemann.
Pliny. (1962). *Natural history. Volume X. Libri XXXVI-XXXVII.* (M. Eichholz, Trad.). William Heinemann.
Ploennigs, J. (2023). AI art in architecture. *AI in Civil Engineering, 2*(1), 8.
Radford, A. K. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763). PMLR.
Ramesh, A. P. (2021). Zero-shot text-to-image generation. In *International conference on machine learning* (pp. 8821–8831). PMLR.
Riggsby, A. M. (2003). Pliny in space (and time). *Arethusa, 36*(2), 167–186. https://doi.org/10.1353/are.2003.0018
Rombach, R. B. (2022). High-resolution image synthesis with latent diffusion models. In *IEEE/CVF conference on Computer Vision and Pattern Recognition* (pp. 10684–10695).
Ruiz, N. L. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *IEEE/CVF conference on Computer Vision and Pattern Recognition* (pp. 22500–22510).
Russell Taylor, A. (April de 26 de 2023). *Pliny's Villa at Laurentum*. Retrieved from https://rtarchitects.co.uk/2023/04/26/plinys-villa-at-laurentuma/
Scamozzi, V. (1615). *L' Idea Della Architettvra Vniversale: Diuisa in X. Libri* (Vol. 1). Expensis auctoris.
Schuhmann, C. B. (2022). Laion-5B: An open large-scale dataset for training next generation image-text models. In *Advances in Neural Information Processing Systems* (pp. 35, 25278–25294).
Sze, V. C. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12), 2295–2329.
Tanzer, H. H. (1924). *The Villas of Pliny the younger*. Columbia University Press.
Von Erlach, F. (1725). *Entwurff einer historischen Architectur*. s.n.
Winston, P. H. (1984). *Artificial intelligence*. Addison-Wesley Longman Publishing.

Index

F. Moral-Andrés et al. (eds.), *Decoding Cultural Heritage*,
https://doi.org/10.1007/978-3-031-57675-1