

Industrial Landscapes, Photographic Description and Generative AI: Testing GPT-5.0 ‘Thinking’ and ‘Instant’*

Elisa Angeles Fernandez

¹ Historical Laboratory of (Electro)Technology, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-166 27 Prague 6, Czech Republic. The text was created with the support of SVK26/25/F3.

Abstract

This study explores the use of GPT-5.0, both *Thinking* and *Instant* modes, for visual analysis of industrial landscapes, with a focus on hydroelectric sites. Using a preliminary comparative design on a curated image set, to assess the models’ capacity to: identify industrial elements and technical traces, employ appropriate terminology, and propose geographical, historical, and technological contexts. Overall, both versions give mostly accurate descriptions of infrastructures and changes in the terrain, and they fit well with the analytical framework that highlights technical, ecological, and historical aspects. Still, the use of context and embedded text is uneven, location guesses are uncertain, and errors often come from adding extra interpretations or missing elements. Compared to *Instant*, *Thinking* uses shorter and more technical words, but it does not actually reduce errors. We suggest that GPT-5.0 works best as a tool that helps organize and speed up the reading of photographs, rather than replacing expert analysis. Future work should expand the dataset and improve evaluation methods to better test reliability.

Keywords

Industrial Landscape, generative language, GPT-5.0, visual analysis, photographs, history of technology, hydroelectric, image-based analysis, industrial heritage.

1. Introduction

1.1. AI and landscape interpretation

In recent years, research has increasingly explored the intersection between generative artificial intelligence, image analysis, and landscape interpretation. Xing et al. (2025) survey applications of AI-generated content in landscape architecture, identifying new possibilities in site analysis, design generation, and construction management, while also warning about issues of data quality, bias, and sustainability [1]. Complementing this, Huang et al. (2025) provide a comprehensive overview of foundation models for vision and multimodality, whose capacities range from classification and captioning to change detection in remote sensing [2]. More domain-specific developments such as Tree-GPT [3], integrates large language models with expert systems for forest image interpretation, and Tan et al. (2023), who discuss multimodal models for environmental and urban planning, demonstrate the growing methodological maturity of AI tools in spatial and ecological analysis[4].

Together, these contributions indicate that multimodal AI is already being integrated into landscape and territorial studies. However, it remains largely confined to technical, planning, or environmental frameworks. The cultural and historical dimensions of the landscape, those that connect infrastructure, memory, and heritage, are still rarely addressed.

VIPERC2025: 4th International Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 3-4 December 2025

The paper is submitted to the Special Session Etech-CH

¹ ✉ angeleli@fel.cvut.cz (E. Angeles)

 0009-0000-7915-5782 (E. Angeles)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This study seeks to bridge that gap. Building upon these advances, it applies GPT-5.0 to the visual analysis of industrial and hydroelectric landscapes, exploring whether generative AI can serve not merely as a descriptive engine but as a new interpretive instrument. AI is positioned as a continuation of the long-standing practice of *reading the landscape*: a tool that translates visual evidence into textual and analytical form.

Photography, central to this experiment, functions here as both medium and method, an archive through which the technical, social, and cultural dimensions of territory can be read [5]. Rather than replacing expert observation, AI can help extend it by systematizing the recognition of infrastructural forms, environmental transformations, and historical layers across large image corpora. Contributing to a renewed dialogue between the historian and what might now be called the *datafied landscape*.

This paper presents a preliminary comparative experiment between two multimodal variants of GPT-5.0, *Thinking* and *Instant*, to evaluate their descriptive and interpretive performance. Photographs of landscape are treated here not only as technological or environmental records but also as documents, allowing us to test the extent to which generative models can mediate interpretation.

1.2. Aims and research questions

The study engages with a global corpus of hydroelectric landscapes (photographs from the United States, Mexico, Czechia, Italy, Spain, among others), situating them in relation to these historiographical traditions. The central aim is to test the descriptive and interpretive capacities of GPT-5.0 when applied to hydroelectric infrastructures.

Key questions:

- Can GPT-5.0 generate technically accurate descriptions?
- What industrial elements can it identify?
- Can it situate an image geographically or technologically?
- What kind of mediation does it perform between researcher and image?

1.3. Industrial landscapes and historiography

Industrial landscapes embody the deliberate and organized imprint that human activity leaves upon natural or agricultural terrain, transforming it for productive purposes [6]. The idea of reading such environments as historical documents was first articulated by W. G. Hoskins (1955), who revealed how successive layers of settlement and production inscribe human presence into the land [7]. Building on this, Kenneth Hudson (1963) established *industrial archaeology* as a field of inquiry that extended the interpretation of landscapes to factories, mines, and power plants, the tangible residues of industrial modernity [8]. Later, David Lowenthal (1985) reframed the landscape as a cultural construct of memory, emphasizing the ways in which societies choose to preserve, reinterpret, or erase the material traces of their past [9]. From a social-historical angle, Patrick Joyce (2003) examined industrial urban landscapes as arenas where power, labor, and everyday experience become spatially encoded [10]. Complementing this, Tim Edensor (2005) explored the affective and aesthetic dimensions of post-industrial ruins, showing how decay and obsolescence can evoke embodied encounters with modernity itself [11].

Another crucial conceptual tool for understanding industrial environments were appointed by Eva Semotanová (2002), for instance, situated the Czech landscape within a *longue-durée* geographical and historical framework [12]. The German tradition of *Kulturlandschaft* provided by Meyer (1998) and recent debates on *Industriekultur* [13] have further emphasized the cultural and heritage values embedded in industrial sites. Research by Tomáš Šenberger and colleagues (1995 [14]; 2021 [15]; 2023 [16]) demonstrates how former industrial and technical structures (ranging from factories to brownfields) are increasingly interpreted as integral components of cultural landscapes.

Together, these historiographical and geographical traditions have shaped a way of “reading the landscape” that bridges material evidence, cultural meaning, and historical interpretation. Building on this lineage, the present study situates AI within that same interpretive continuum: testing whether a generative model can reproduce, complement, or challenge the human gaze that has long guided the study of industrial and hydroelectric environments.

1.4. Historical background of hydroelectric landscapes

Hydroelectric landscapes emerged in the late nineteenth century, when the conversion of water power from mechanical to electrical energy marked a decisive technological shift in industrial modernity. Early known projects such as the Folsom Powerhouse (1895), Niagara Falls Power Plant (1896) inaugurated the era of dam-based electricity, integrating civil engineering, electromechanics, and territorial design. In Alpine regions steep gradients and abundant rivers fostered early investments in hydropower from the 1890s onwards, giving rise to the so-called *houille blanche* or “white coal” movement. During the interwar period, monumental constructions like Hoover Dam (1936, United States) and the Dnieper Hydroelectric Station (1932, Ukraine) came to symbolize the ideology of progress, combining large-scale state planning with modernist aesthetics of control over nature.

In the decades following WWII, hydroelectric expansion became a global phenomenon, linked to industrialization, rural electrification, and national development programs. In Scandinavia, particularly in Norway and Sweden, hydroelectric expansion became central to welfare-state policies and resource nationalism, integrating electricity networks with heavy industry and rural electrification [17]. In socialist Central and Eastern Europe, monumental projects such as the Vltava Cascade in Czechoslovakia (among them the Orlik Dam, 1961), or the Iron Gates Dam (1972, Yugoslavia–Romania) reflected state-led modernization. These infrastructures also reconfigured landscapes through the construction of workers’ housing, new settlements, and transport networks, merging technological and social planning into a unified spatial project. Large dams in Latin America and Asia, such as Koyna (India, 1981) and Itaipú (1984, Brazil–Paraguay), exemplified the postwar belief in infrastructure as an engine of modernization. This phase consolidated a techno-economic model that reshaped entire river basins, displaced communities, and produced new territorial hierarchies between energy centers and dependent regions.

At the same time, hydroelectric architecture became a laboratory for modernist aesthetics and material experimentation. From the concrete sculptural forms of the Norwegian *kraftverk*, the Art-Deco of the to the wonder of the Hoover Dam, to the monumental façades of Italian rationalism, dams and powerhouses expressed a shared twentieth-century imaginary of progress and domination over nature [18, 19]. Behind these projects were as well new social realities: seasonal labor, specialized engineering migration, and the transformation of the landscape into an industrial one, that made hydropower development a human enterprise as much as a technical one.

From the late twentieth century to the present, hydroelectric systems have entered a new stage of re-evaluation. Many early sites are now recognized as industrial heritage, while debates on sustainability and ecological impact have reframed dams as both technological achievements and landscapes of environmental transformation. Today, hydroelectric infrastructures stand at the intersection of engineering history, environmental change, and cultural identity, making them key case studies for understanding how technology inscribes itself onto territory and memory.

2. Analytical framework

The use of generative AI in the analysis of hydroelectric landscapes offers a new methodological perspective for the history of technology and industrial heritage. These landscapes, often complex and technically dense, embody traces of human intervention that extend beyond engineering design. In line with the tradition of *reading the landscape* established by Hoskins, Hudson, and their successors, AI can be conceived as a contemporary interpretive tool, one that translates visual evidence into structured textual and analytical form.

Systems such as GPT-5.0 enable researchers to process extensive image corpora, identify recurring patterns, and highlight infrastructural and environmental features (turbines, dams, sluices, reservoirs, or power lines) that articulate the dialogue between nature and technology. Beyond simple recognition, this experiment explores whether AI can capture the contextual, temporal, and cultural dimensions that shape the visual identity of industrial sites.

To guide this analysis, the study adopts a framework structured around seven interrelated dimensions, designed to reveal the technical, ecological, and social layers that compose industrial landscapes:

1. *Contextual and technological interpretation:* Inferring geographical, historical, and technological settings from visible clues.
2. *Technical and infrastructural traces:* Identifying forms, materials, and functions of infrastructures, including their adaptation to the terrain.
3. *Ecological transformations:* Detecting environmental changes such as reservoir formation, deforestation, or river alteration associated with human intervention.
4. *Scale and territoriality:* Evaluating the extent of infrastructural networks and their relation to regional or national systems.
5. *Temporal dimension:* Recognizing signs of construction, modernization, abandonment, or heritage reuse.
6. *Social and cultural dimension:* Observing evidence of human presence, labor, or cultural integration within industrial environments.
7. *Historical dimension:* Reading the photograph as a document that reveals successive layers of human and technological transformation.

This framework resonates with broader historiographical debates on the visual construction of landscape. From 19th-century artistic depictions such as Turner's, to the emergence of industrial archaeology and Hoskins' vision of the landscape as a historical document [7], the act of visual interpretation has long been central to understanding how technology reshapes territory. Worster (1985) emphasized the role of water management in transforming socio-ecological systems [20];

Semotanová (2002) highlighted the integration of local infrastructures into wider territorial networks [12]; and Alba et al. (2021) examined the phases of ruination, modernization, and musealization that define the temporal life of industrial sites [21].

By linking these historiographical perspectives with computational methods, this framework positions generative AI not as a substitute for expert interpretation but as an instrument that systematizes and extends it, offering new means to analyze industrial and hydroelectric landscapes as layered historical documents.

2.1. Application of the framework to photographic evidence

Industrial landscapes, understood as physical, social, and cultural realities, demand integrative approaches (Alba et al., 2021). In this sense, photography has long been used to trace change, capturing material transformations and their social meanings (Collier, Inkpen & Fontana, 2015; Hoskins, 1955) [22]. Prior studies (Bisbal Grandal, 2016; López Mesa et al., 2015) show that interpretation often oscillates between naturalist, culturalist, and integrative perspectives, with photography enabling comprehensive readings [23].

Applying our seven dimensions to images requires attending not only to the most evident infrastructures (their form, function, materiality, and adaptation to terrain), but also to subtler technical imprints (terraces, cuts, rock modifications) that reveal transformation, adaptation, and control of territory. Photographic series have been used to detect ecological change (reservoir inundation, diverted watercourses, deforestation) (Collier et al., 2015) and to document how water management reshapes socio-ecological systems (Worster, 1985). At the level of scale and territoriality, historical geography helps situate local infrastructures within regional and national networks (Semotanová, 2002).

The temporal dimension is likewise crucial: photographs register phases of construction, modernization, abandonment, and heritage reuse; these can be recognized through ruination, technological updating, or musealization (Alba et al., 2021). Finally, the social and cultural dimension attends to settlements, workers' housing, and everyday spatial practices that embed human life within technical landscapes and give depth to their historical significance.

3. Materials and Methods

The corpus used in this experiment consists of 45 photographs of hydroelectric landscapes, sourced from publicly available online repositories. Images were selected primarily for their clarity, neutrality, and absence of metadata criteria intended to avoid any textual or contextual information that could bias the AI's descriptive output. Most photographs are relatively recent (likely taken within the past two decades), but exact dates could not always be verified. The priority was therefore not archival dating, but the creation of a controlled visual dataset suitable for testing the descriptive and interpretive performance of GPT-5.0, rather than a historically delimited photographic archive.

This study was conceived as a preliminary exploratory investigation, explicitly designed to compare the performance of two GPT-5.0 multimodal variants, *Thinking* and *Instant*, under identical conditions. The comparative design aimed to evaluate not only accuracy but also differences in reasoning style and interpretive strategy. To assess robustness and generalization, 25 % of the images were intentionally selected as out-of-distribution (OOD), depicting sites or features beyond the main reference domain.

Each photograph (legible, free of watermarks or annotations, and of adequate resolution) was presented individually in an isolated chat session. For every image, both models answered the same five fixed prompts, each limited to a maximum of eight lines and written in English:

- a) **Q1-Subject:** *What is the central subject of the photograph?*
- b) **Q2-Setting:** *What is the geographical and environmental setting represented?*
- c) **Q3-Infrastructures:** *Which infrastructures or industrial elements are visible?*
- d) **Q4-Transformations:** *What transformations of the terrain can be observed?*
- e) **Q5- Historical/Technological Clues:** *Which visual clues suggest the historical period or the technological phase of the photograph?*

Evaluation, aimed at identifying the presence of one or more descriptive, interpretive, or related errors, is conducted question by question according to a taxonomy of non-mutually exclusive errors (E0 applies only if none of E1–E7 is present):

- **E0 – No error:** Output accurate and consistent; visual elements described faithfully.
- **E1 – Detection errors:** Addition or omission of clearly visible elements.
- **E2 – Semantic errors:** Confusion of categories or incorrect meaning assignment (e.g., turbine vs. water wheel).
- **E3 – Descriptive errors:** Vague or generic phrasing; emphasis on marginal aspects.
- **E4 – Attribution errors:** Misattribution of objects or structures to the wrong culture, era, or context.
- **E5 – Interpretive errors:** Unsupported inferences or internal contradictions.
- **E6 – Bias-related errors:** Cultural or visual stereotypes distorting interpretation.
- **E7 – Other:** Any errors not included above.

This taxonomy (E0–E7) was developed specifically for this study, drawing on general frameworks used in multimodal and text-generation error analysis. It adapts common distinctions such as detection, semantic, and interpretive errors to the visual-historical context of industrial landscape interpretation.

Because this is a preliminary exploratory phase, outputs were evaluated by a doctoral researcher in the history of science and technology, with training in landscape architecture and experience in the analysis of industrial heritage, who also administered the stimuli. Although this means the procedure was non-blind, the adopted taxonomy emphasizes manifest, directly observable indicators, such as presence or absence of dams, penstocks, or transmission towers, so discrepancies are straightforward when comparing images and textual outputs. In this context, masking the model identity was not expected to influence results.

The study was conducted using the ChatGPT interface rather than the API. This choice was guided by the intent to replicate a realistic, non-programmatic use case closer to how researchers in the

humanities may interact with generative AI tools. However, future work should test API-based configurations to enhance reproducibility and parameter control.

4. Results

4.1. Preliminary comparative observations

In the first half of the experimental sample, some differences between the two GPT-5.0 versions became evident. *Thinking* generally employed a richer and more technical vocabulary in its descriptions, although in certain cases it could have used more specialized terminology. It was notably strict in its self-evaluations, frequently flagging minor issues, and in two cases it refrained from exploiting textual cues present within the photograph to infer location. In contrast, *Instant* produced more generic but still largely correct answers. Its self-evaluations were more lenient and it did not display explicit metacognitive actions such as “checking the image,” which were occasionally observed in the *Thinking* outputs.

These tendencies suggest that *Thinking* engages in more elaborate reasoning chains while *Instant* favors speed and generality. Nevertheless, in both cases the overall accuracy of industrial element detection remained comparable and broadly consistent.

4.2. Final Comparative Results: Instant vs. Thinking

-Stylistic and Content-Based Differences Between Instant and Thinking Responses

Looking at the comparison, there are two very different ways of answering (Table 1). One style (*Thinking*) preferred short, precise, and technical phrasing, almost like a checklist. The other (*Instant*) leaned toward longer, more narrative explanations, adding descriptions and context.

Table 1 Stylistic and Content-Based Differences Between Instant and Thinking Responses. Generated by the author using experimental results.

Aspect	<i>Thinking</i>	<i>Instant</i>
Style	Technical, compressed, stronger vocabulary, slightly more specialized tone.	Narrative, expansive, flowing sentences, added descriptions.
Length	Shorter in: Q1 (100%), Q2 (100%), Q3 (97.8%), Q4 (88.9%), Q5 (88.9%), Control (28.9%).	Generally longer; looking to fulfill the 8 lines statement. Only in Control was it shorter in most cases.
Q1 (Subject)	Concise, list-like precise terms.	Full sentences, more scene-setting.
Q2 (Environment/Setting)	Brief, direct, technical.	Narrative, with extra context and atmosphere.

Aspect	<i>Thinking</i>	<i>Instant</i>
Q3 (Infrastructure)	Compact lists, domain-specific language (e.g., erosion forms, engineered features).	Expanded explanations, background mechanisms.
Q4 (Transformations)	Technical terms and tight bullet points.	More elaboration, descriptive of processes.
Q5 (Historical/Technological Clues)	Tight summaries, evidence-tagged lines.	Extended into timeframes and plausibility notes.
Recognition of iconic content	Correct when strong visual cues present (e.g., Hoover Dam).	Equally effective.
Observed limitations	Sometimes did not exploit embedded text; self-evaluation strict, overly conservative.	Less restrictive, often added helpful but redundant context.

In the first group of questions (Q1 and Q2), the contrast was especially clear: concise notes versus flowing sentences with extra scene-setting. In the middle questions (Q3 and Q4), the compact style relied on lists and technical terms, while the longer style expanded with background details. For historical or temporal clues (Q5), the brief approach stuck to evidence-based summaries, while the other stretched into timeframes and plausibility notes. In the control questions, the gap narrowed, since the narrative mode often justified its reasoning at length.

Both ways were able to spot well-known sites like Hoover Dam, but neither managed with less famous places such as Orlik or Necaxa. At times, the shorter approach could have made better use of text embedded in the images, and it also judged itself more strictly than necessary.

Overall, one style focused on density and precision, while the other emphasized readability and context. The result is a trade-off: concise answers that get to the point quickly versus longer ones that provide a fuller narrative.

Figure 1 shows the intake of an earth dam in Nexaca, Mexico, an example of the hydroelectric landscapes included in the corpus, alongside a comparative table summarizing GPT-5 *Thinking* and *Instant* outputs in Table 2. While both recognize the same structures, the former provides denser technical detail, and the latter situates the scene more narratively within its environment. In this case, GPT-5’s *Thinking* and *Instant* produced notably different descriptive tendencies. The *Thinking* mode emphasized the technical and structural aspects of the dam (identifying the intake tower, machinery, and engineering details), while *Instant* generated a more general visual and contextual interpretation focused on the surrounding environment. Exemplifying with these differences how “*reasoning depth*” influences the model’s treatment of industrial and ecological features within the same visual input.

Figure 1. Intake of an earth dam in Nexaca, Mexico. The hydroelectric landscape photograph used as visual input, with comparative GPT-5 Thinking and Instant outputs. *Source: Elisa Angeles Fernández, 2025.*



Table 2 The comparative outputs are summarized in text rather than reproduced as images. *Generated by the author using experimental results.*

	Thinking	Instant
Q1	Concrete tower structure in the middle of a green reservoir labeled “FENIX,” with yellow crane and industrial platforms, likely part of a hydroelectric facility.	Concrete hydraulic structure labeled “FENIX,” rising from green water, with directional signs (“SUR,” “NORTE”) and yellow machinery, surrounded by forested hills.
Q2	Mountainous, forested highlands with steep slopes and a man-made reservoir; dense and abundant vegetation contrasts with the concrete tower.	Tropical or subtropical valley landscape, steep slopes and environment.
Q3	Intake/control tower, gantry crane, railings, ladders, signage (“FENIX,” “SUR,” “NORTE”), cables and lighting poles.	Large concrete tower with yellow crane, railings, signage, and electrical cables; part of a dam or hydroelectric complex.
Q4	Flooded valley forming an artificial lake; clearing, erosion scars, and altered shoreline visible.	Submerged valley floor and modified slopes; deforestation for construction access and maintenance tracks.
Q5	Modern corporate logos, standardized railings, and safety colors typical of late-20th/21st-century industrial maintenance.	Contemporary industrial design, modern logos, and reinforced concrete typical of current hydroelectric facilities.

-Error evaluation and frequency across the experiment

The evaluation revealed notable differences in error distribution across response modes. Error-free outputs (E0) were the most frequent overall, accounting for 50.0% of *Instant* answers and 36.4% of *Thinking* answers. However, interpretive errors (E5) emerged as a major source of inaccuracy, representing nearly one-third of responses in both conditions (31.9% instant; 33.3% thinking). Detection errors (E1) also appeared consistently, with a higher prevalence in reasoning-based outputs (24.4%) compared to instant ones (12.8%). Other error types, such as semantic (E2), descriptive (E3), and attribution errors (E4), occurred less frequently ($\leq 3.5\%$). Bias-related (E6) and unclassified errors (E7) were absent. The results indicate that while correct outputs are common, errors linked to overinterpretation, and element detection remain the most salient challenges, please, check Table 3 and Figures 2,3,4,5,6,7.

Table 3 Distribution of error types across instant and reasoning responses

Code	Error name	Instant (n)	Instant (%)	Thinking (n)	Thinking (%)
E0	No error	113	50.0%	82	36.4%
E1	Detection errors	29	12.8%	55	24.4%
E2	Semantic errors	2	0.9%	4	1.8%
E3	Descriptive errors	2	0.9%	7	3.1%
E4	Attribution errors	8	3.5%	2	0.9%
E5	Interpretive errors	72	31.9%	75	33.3%
E6	Bias-related errors	0	0.0%	0	0.0%
E7	Other	0	0.0%	0	0.0%

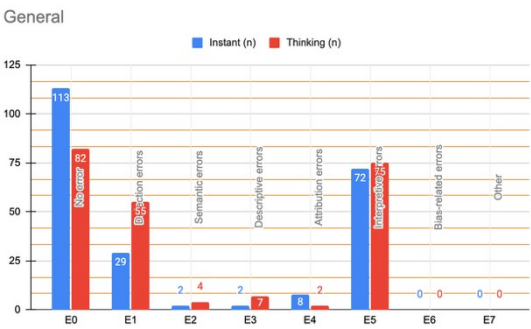


Figure 2: General error distribution.
(Generated by the author using experimental data.)

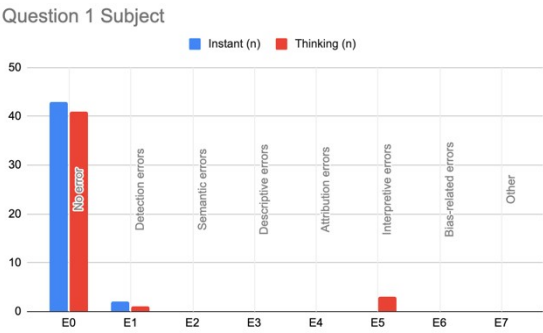


Figure 3: Error distribution in Q1: Subject
(Generated by the author using experimental data.)

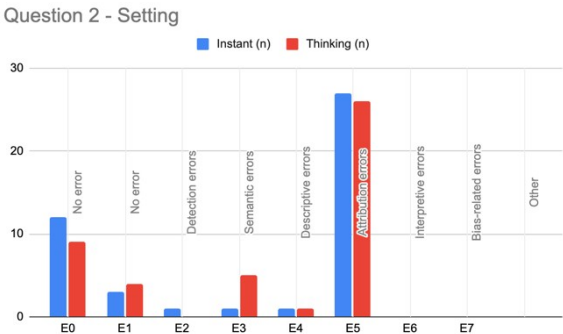


Figure 4. Error distribution in Q2: Setting.
(Generated by the author using experimental data.)

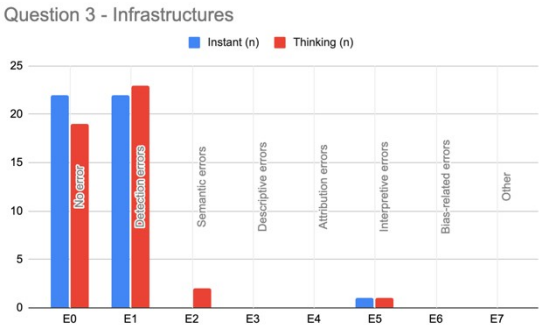


Figure 5. Error distribution in Q3: Infrastructures.
(Generated by the author using experimental data.)

Question 4 - Transformations

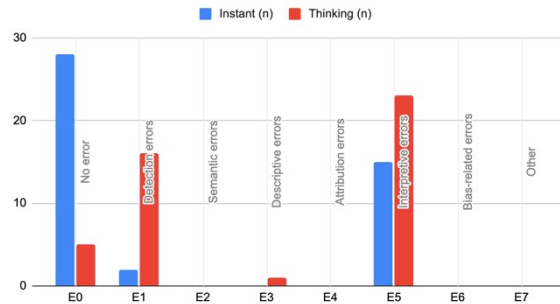


Figure 6. Error distribution in Q4: Transformations
(Generated by the author using experimental data.)

Question 5 Historical/Technological Clues:

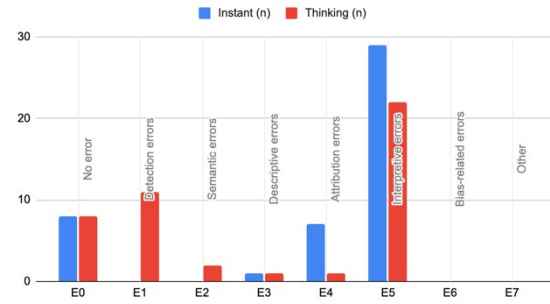


Figure 7. Error distribution in Q5: Historical
(Generated by the author using experimental data.)

What stands out in these results is that, although many answers were correct, two types of mistakes kept appearing very often: interpretive errors and detection errors. Interestingly, these problems were not reduced when the system gave more elaborate, “thinking” answers, in fact, detection errors became more frequent. This shows that while the model can describe images accurately, it often struggles when it tries to explain too much or add details that are not really there.

5. Discussion

As this is a preliminary investigation, the results should be taken as indicative rather than conclusive. Nevertheless, the comparative design between GPT-5.0 *Thinking* and *Instant* provides us evidence of how these distinct model versions interpret industrial landscapes and opens possibilities for future studies.

With respect to the guiding questions posed in the introduction, the analysis suggests that GPT-5.0 can indeed produce technically accurate descriptions of landscapes and infrastructures, particularly the *Thinking* mode. It used precise terminology and structured answers, even noticing features such as erosion processes, dam components, and sluice structures, among others.

Both models were consistently reliable with iconic sites, while its responses were more cautious or limited when confronted with less known or distinct environments.

In terms of industrial elements, the AI successfully identified hydroelectric stanchions, substations, spillways, sluices, train lines, concrete constructions, etc., often linking them to a broader functional or historical context, quite accurately. This shows us a capacity to distinguish infrastructures from natural formations, although the level of detail remains descriptive rather than technical in the expert sense. We can also note a partial capacity for geographical and technological contextualization. When landmarks were well recognizable or inscriptions were visible, GPT-5.0 could situate the images within a plausible environmental and historical setting. However, it remained imprecise for a finer geolocation.

Although human cross-checking was conducted during evaluation, no major divergence was observed between the researcher’s reading of the images and the AI-generated descriptions. Differences remained mostly in the degree of precision or vocabulary rather than in interpretation: both human and model consistently identified the same infrastructures and landscape transformations. This outcome suggests that, at least for clearly legible photographs, GPT-5.0 tends to align with expert visual observation, though without the contextual depth or historiographical reasoning that characterize human analysis.

The experiment makes it easier to understand the kind of mediation performed between the researcher and the image. GPT-5.0 acts as an interpreter translating visual details into words, groups and labels features, and often comments on its own answers. This caution sometimes made the answers less direct but also shows how much tools can support researchers' work. The idea that "AI acts as an interpreter" must, however, be understood in a limited sense. While the model translates visual evidence into textual form, its process of interpretation differs fundamentally from that of the historian or archaeologist. Human interpretation draws on contextual, archival, and theoretical knowledge, weaving multiple sources and temporal layers into an argument about meaning and causality. AI, in contrast, operates through probabilistic associations within training data, producing correlations rather than explanations. Its "reading" of the image remains descriptive and pattern-based, lacking the reflexive and critical dimension that characterizes historical inquiry. Recognizing this difference helps clarify the complementary role of AI, not as a substitute for expert judgment, but as a structured assistant that systematizes and accelerates visual analysis while relying on human interpretation for depth and validation. Again, the notion that "AI acts as an interpreter" should also be approached critically. Unlike the engineer or historian, whose interpretation is guided by contextual and archival knowledge, AI operates through probabilistic associations based on training data. This difference reveals both its potential (to uncover visual regularities at scale) and its limitation, as it may reproduce aesthetic biases or omit socio-historical nuance.

Evidence of such biases appeared across the dataset. Both GPT-5.0 modes often aestheticized industrial infrastructures through expressions such as "*majestic*," "*impressive*," or "*spectacular*," portraying dams and power plants as monumental works of modernity. Iconic Western sites like Hoover Dam were precisely recognized, whereas non-Western or less monumental landscapes (e.g., Orlik, Necaxa) received generic or uncertain descriptions. This pattern suggests a visual bias toward Western industrial imagery and a technocentric reading of the landscape that privileges engineering over social or environmental dimensions. These tendencies also raise ethical and historiographical questions regarding what can be called *datafied heritage*: when the interpretation of cultural landscapes depends on training data and algorithmic vision, the archive itself becomes mediated by unequal visibility and cultural bias.

Finally, when compared to the analytical framework, the descriptions match the main dimensions, the model can point out technical traces, ecological changes, and historical layers, even if its account stays descriptive rather than expert.

Beyond interpretive bias, the experiment also points to methodological implications for the history of technology. Multimodal AI could assist in the cataloguing and cross-referencing of archival photographs, identifying recurring technical elements or construction typologies across vast image collections. The automated "reading" of images transforms historical documentation into *datafield heritage* (a searchable, analyzable layer of cultural memory) while requiring historians to critically mediate how such data are interpreted and contextualized.

6. Conclusion and Future Perspectives

This preliminary study shows that Chat GPT-5.0 can describe industrial landscapes in a very accurate form, recognize key infrastructures and to some degree, place an image into a quite precise geographical and historical context. And giving more reliable results at iconic sites or with clear visual clues, and it was inconsistent when it came to reading text and pinpoint locations.

The comparison between versions shows that *Thinking* offers more technical and self-critical descriptions, while *Instant* answers more quickly and being overall general. Despite these

differences, both achieved good and correct readings of the elements, confirming that multimodal models can support photographic interpretation of historical landscapes, although without replacing the expert eye.

In relation to the analytical framework, the AI managed to identify technical traces, ecological transformations and temporal dimensions, but with limitations. To move forward, it will be necessary to vary the set of images and improve evaluation methods.

Beyond this exploratory test, the approach demonstrates how multimodal AI can assist heritage scholars in systematically documenting and comparing visual archives. It may be particularly useful for large collections where manual interpretation is time-consuming, offering a complementary tool for the descriptive phase of analysis rather than replacing human judgment.

At the same time, epistemological questions arise when the act of “seeing” is partially delegated to an artificial intelligence. Who becomes the author, the model?, the human prompter?, or their interaction?. From the perspective of photographic theory and visual culture, AI does not merely interpret but also reconfigures the gaze itself, translating visual information into textual form. Future work should therefore address these epistemic shifts through an expanded methodology combining AI-assisted reading with multi-coder human validation, ensuring that interpretation remains a critically controlled process.

Limitations:

This is a preliminary, single-coder, non-blind study on 45 images (25% OOD). The images should not contain metadata, and the order of insertion into the AI was randomized. Corpus based on recent, online photographs, not a historical archive. Single-coder, non-blind evaluation.

Data & Code Availability

Table of results will be made available in a public repository upon acceptance.

The visual corpus used in the experiment consisted of hydroelectric landscape photographs collected from Google Images and used exclusively for research purposes. All files were stripped of metadata, metatags, and alt text to prevent copyright infringement. No copyrighted visual material is included in the published dataset.

Acknowledgements

The author thanks the Historical Laboratory of (Electro)Technology (FEE, CTU in Prague) for supporting the project. The text was created with the support of SVK26/25/F3.

Declaration on Generative AI

During the preparation of this work, the author used Chat-GPT-5.0 in order to: Grammar and spelling check. Further, the author used Chat-GPT-5.0 for the experiment in order to: analyze images. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Xing, Y., Gan, W., Chen, Q., & Yu, P. S. (2025). AI-generated content in landscape architecture: A survey. *arXiv* (arXiv:2503.16435). <https://arxiv.org/abs/2503.16435>
- [2] Huang, Z., Yan, H., Zhan, Q., Yang, S., Zhang, M., Zhang, C., Lei, Y., Liu, Z., Liu, Q., & Wang, Y. (2025). A survey on remote sensing foundation models: From vision to multimodality. *arXiv* (arXiv:2503.22081). <https://arxiv.org/abs/2503.22081>
- [3] Du, S. Q., Tang, S. J., Wang, W. X., Li, X. M., & Guo, R. Z. (2023). Tree-GPT: Modular large language model expert system for forest remote sensing image understanding and interactive analysis. *arXiv*. <https://doi.org/10.48550/arXiv.2310.04698>
- [4] Tan, C., Cao, Q., Li, Y., Zhang, J., Yang, X., Zhao, H., ... Mai, G. (2023). On the promises and challenges of multimodal foundation models for geographical, environmental, agricultural, and urban planning applications. *arXiv*. <https://doi.org/10.48550/arXiv.2312.17016>
- [5] Bisbal Grandal, I. (2016). *El estudio del paisaje por medio de la fotografía: Desarrollo de una metodología interpretativa* [Tesis doctoral, E.T.S. Arquitectura (UPM)]. <https://doi.org/10.20868/UPM.thesis.40527>
- [6] Borsi, F. (1975). *Le paysage de l'industrie*. Archives d'Architecture Moderne.
- [7] Hoskins, W. G. (1955). *The making of the English landscape*. Hodder & Stoughton.
- [8] Hudson, K. (1963). *Industrial archaeology: An introduction*. John Baker.
- [9] Lowenthal, D. (1985). *The past is a foreign country*. Cambridge University Press.
- [10] Joyce, P. (2003). *The rule of freedom: Liberalism and the modern city*. Verso.
- [11] Edensor, T. (2005). *Industrial ruins: Spaces, aesthetics and materiality*. Berg Publishers.
- [12] Semotanová, E. (2002). *Historical geography of the Czech lands* (2nd ed.). Historical Institute of the Academy of Sciences of the Czech Republic.
- [13] Meyer, H. (2025). *Industriekultur – Quo vadis?* Halle-Saale: Mitteldeutscher Verlag.
- [14] Šenberger, T. (1995). *Rekonstrukce výrobně-technických staveb k novým účelům* (2nd ed.). Prague: Národní technické muzeum.
- [15] Popelová, L., & Šenberger, T. (Eds.). (2021). *Průmyslové dědictví – evidence, výzkum, ochrana, nové využití = Industrial heritage – recording, research, protection, re-use*. České vysoké učení technické v Praze.
- [16] Popelová, L., & Šenberger, T. (Eds.). (2023). *Poválečný industriál: Historie a budoucí využití*. Česká technika – ČVUT.
- [17] Kaijser, A., & Hedin, J. (1995). *Nordic energy systems: Historical perspectives and current issues*. Science History Publications.
- [18] Banham, R. (1986). *Concrete Atlantis: U.S. industrial building and European modern architecture, 1900–1925*. MIT Press. <https://archive.org/details/concreteatlantis0000banh>

- [19] Forty, A. (2000). *Words and buildings: A vocabulary of modern architecture*. Thames & Hudson. https://arch3281fall14.wordpress.com/wp-content/uploads/2014/09/adrian-forty_words-and-buildings_introduction.pdf
- [20] Worster, D. (1985). *Rivers of empire: Water, aridity, and the growth of the American West*. Pantheon Books. <https://archive.org/details/riversofempirewa00wors>
- [21] Alba Dorado, M. I., et al. (2021). El paisaje industrial: Aproximación al diseño de unas bases metodológicas para su estudio, puesta en valor e intervención. *ACE: Architecture, City and Environment*, 16(48), 10501. <https://doi.org/10.5821/ace.16.48.10501>
- [22] Collier, P., Inkpen, R., & Fontana, D. (2015). The use of historical photography in environmental studies. *Cybergeog: European Journal of Geography*. <https://doi.org/10.4000/cybergeog.4019>
- [23] López Meza, I., Bisbal Grandal, I., & Aroztegui, C. (2015). Refotografiando paisajes mineros para analizar el cambio: El caso del poblado minero de Lota en el sur de Chile. *Arquiteturarevista*, 11(1), 11-21. <https://www.redalyc.org/pdf/1936/193642162003.pdf>