A Prompt-to-Image Workflow for Professional Planting Visualisation: Evaluating the Use of ChatGPT-4o in Landscape Architecture

Daniel Theidel[1,2]

[1]University of Applied Sciences, Anhalt/Germany · daniel.theidel@hs-anhalt.de
[2]University of Applied Sciences, Osnabrück/Germany · d.theidel@hs-osnabrueck.de

Abstract: Visualising planting concepts requires images that are botanically legible, spatially coherent, and seasonally plausible. While recent multimodal AI systems promise rapid, context-aware visualisations, reproducible workflows grounded in real planning data remain scarce. This paper proposes and tests a prompt-to-image workflow developed by the author for planting visualisation using ChatGPT (Version GPT-4o) (OPENAI 2025) in a realised project in Paris, based on a planting plan, species list, licence-free plant reference images, and a vegetation-free context perspective. The workflow combines structured data extraction (JSON content profile), a modular prompt architecture with documented versions (v0-v2), and image generation constrained to a red-marked target zone. Five images were generated under constant inputs to document system-immanent variability; the most plausible variant was selected as evaluation stimulus (Image A, Fig. 4). Seven experts assessed Image A using Likert ratings, an error checklist, and qualitative comments. Results indicate limited suitability for professional, species-specific planting communication (overall mean 2.34/5), with key limitations in botanical identity, morphology, phenology, and geometric scale rather than inpainting artefacts. The contribution is a clearly documented workflow and an expert-based delimitation of current use boundaries.

Keywords: Planting design, visualisation, generative AI, prompt-to-image, reproducibility, landscape architecture

1 Introduction

Visual communication of planting concepts is central to landscape architectural design, coordination, and decision making. Between hand sketch, collage, and CAD-based representation, images are an essential tool for design communication and collaboration (REKITTKE & HAYLES 2025, 636). For planting design, this task is particularly demanding: visualisations must remain botanically legible while conveying seasonal atmosphere and spatial structure.

Traditional visualisation techniques such as analogue drawings, Photoshop collages, or CAD drawings/renderings are well established but often time-consuming and limited in atmospheric expressiveness, especially for complex multi-layered plantings. Recent studio and practice reports highlight both the creative potential and the limitations of generative AI, particularly decreasing controllability and reproducibility as task complexity and detail increase, and a continued need for human intervention when precision is required (ZWANGSLEITNER et al. 2024, 992, SCHROTH & MAIER 2025, 665).

Multimodal AI models such as ChatGPT-4o (OPENAI 2025) can combine planting plans, species lists, and reference images to generate context-aware visualisations very quickly. However, reproducible workflows grounded in real planning data and systematic expert evaluations of the resulting images remain scarce. This paper specifies a prompt-to-image workflow for planting visualisation and empirically assesses a curated AI output against professional criteria.

Journal of Digital Landscape Architecture, 11-2026, pp. 16-31. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-770-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537770002.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nd/4.0/).

Research question: To what extent can a ChatGPT-4o-based workflow generate planting visualisations that are botanically and spatially accurate and atmospherically convincing?

Objectives:

- document a reproducible workflow (JSON content profile, modular prompts, constrained inpainting, controlled variant generation),
- define and structure evaluation criteria (ratings, error checklist, coded comments) across botany/geometry, atmosphere, and applicability,
- identify error patterns and boundaries for responsible professional use.

The paper contributes to the DLA theme "*Computation and Digital Technologies*" by offering a transparent, reproducible account of how multimodal AI can be embedded in planting visualisation workflows while critically discussing the limitations, risks, and professional responsibilities associated with presenting AI-generated images as "professional standard."

2 Related Work

Research and practice on AI-based visualisation in landscape architecture can be grouped into three strands that form the background for this study:

1) explorative uses of image generation in education and early design,
2) data-driven models for vegetation and planting layouts,
3) empirical studies on the adoption of AI tools in professional offices.

Explorative image generation in studios and teaching
Several recent contributions report how text-to-image systems are introduced in design studios and seminars. Rekittke & Hayles (2025) describe early experiments with Midjourney and similar tools in landscape architecture education. They show how AI images can support conceptual exploration, while also revealing limits in producing drawing-like, scaled outputs and in maintaining historical and spatial logic (REKITTKE & HAYLES 2025, 640-641). Florian Zwangsleitner et al. (2024) report on a master-level seminar combining analogue models, digital workflows and AI prompting. While students appreciated the speed and atmospheric richness of AI images, the authors emphasise decreasing reproducibility and controllability as the level of detail increases technical plans and exact planting layouts remained firmly in the domain of CAD and image editing (ZWANGSLEITNER et al. 2024, 992).

Data-driven and model-based approaches to vegetation
A second strand of research explores the use of machine learning and generative models for vegetation pattern generation. Xun Liu et al. (2024) experiment with GAN and pix2pix architectures to derive planting patterns from environmental data such as relief, soil and climate. Their work suggests that data-driven models can produce ecologically plausible distributions but it also underlines the need for ecological validation and specialist knowledge in training and interpreting the models (LIU et al. 2024, 204-206). In contrast to these approaches, the present paper does not aim at AI-driven or data-driven generation of planting layouts. Instead, it assumes a fixed, professionally designed planting plan and focuses on visual communication of this plan through AI-generated imagery.

Professional adoption and concerns
A third body of work examines how practitioners perceive and integrate AI in their everyday workflow. Braiden et al. (2025) present a large survey among landscape architects in North

America. Almost half of the respondents report using AI tools, mostly for text drafting, concept sketches and early visualisations. However, usage drops markedly in later phases such as planting and construction planning. Respondents express concerns regarding data security, copyright, error and the risk of miscommunication when AI-generated images are taken at face value by clients or other stakeholders (BRAIDEN et al. 2025, 628-629). Schroth and Maier (2025) describe generative AI as efficient, but constrained by limited precision and control, implying a continued need for careful human oversight in professional contexts (SCHROTH & MAIER 2025, 665, 674). Concerns about trust, liability, and stakeholder misinterpretation are also reflected in practitioner surveys (BRAIDEN et al. 2025, 629).

Yet, they also reveal a clear research gap: there is little work that links AI image generation to concrete planting plans and species lists, documents the prompt and parameter structure in a reproducible way, and evaluates the resulting images systematically with domain experts. The present study addresses this gap by proposing a JSON-based prompt-to-image workflow and an expert evaluation focusing on botanical/spatial accuracy, atmosphere and conditions for professional use.

3 Methodology

3.1 Overall Approach

This study uses a mixed-methods design that combines structured analysis of real-world planning data, development of a reproducible prompt-to-image workflow using ChatGPT (Version GPT-4o), and an expert assessment using standardised Likert ratings and qualitative responses analysed with a qualitative content analysis approach following Mayring (2015). ChatGPT (Version GPT-4o) was used for workflow specification, prompt development, and image generation only. Participant recruitment, data collection, rating aggregation, qualitative coding, and interpretation were conducted without AI (OPENAI 2025).

The study addresses the research question: ***To what extent can a ChatGPT-4o-based prompt-to-image workflow generate planting visualisations that are botanically and spatially accurate and atmospherically convincing when grounded in real planning data?***

3.2 Case Study and Planning Data

The empirical basis is a realised project in Paris, designed by a Berlin-based landscape architecture office. The office provided complete planning documents for the planting area evaluated in this study. No fictional or synthetic planning data were introduced.

The workflow was based on the following input datasets:

1) Planting plan (Fig. 1)
   Symbolic plan representation including species allocation, matrix/island structure, and density patterning.
2) Species list
   Botanical names (including cultivar where applicable), plant quality specifications (e. g., multi-stem vs. standard where provided), quantities, habitus-related notes, and additional remarks on establishment and seasonal characteristics.

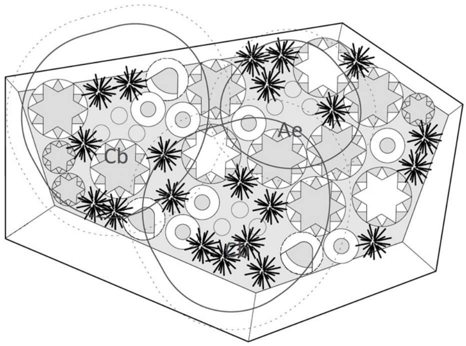

Fig. 1:
Planting plan out of the real project

3) Reference plant images
One to three representative images per species, used as morphological references for the target appearance (leaf size/shape, typical habit, texture). These images served as reference material for the generation process; they were not used as evaluation stimuli.
4) Context perspective (vegetation-free base image; Fig. 2)
A vegetation-free JPEG exported from a 3D model, used as the geometric and photographic base for inpainting. The view corresponds to an approximate eye height of 1.5 m and an approximate focal length of 35 mm. The hardscape background (walls, planter geometry, ground plane) was treated as fixed and was not to be altered during generation.
5) Seasonal target condition
Midsummer (July), representing fully developed foliage.

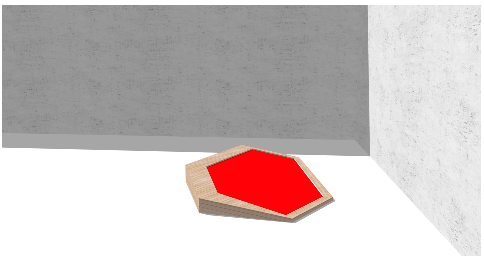

Fig. 2:
Context perspective without vegetation and red region of interest

3.3 Workflow Development (Fig. 3)

The prompt-to-image workflow was specified to translate real planting design data into a controlled visualisation within a fixed project context. The workflow was developed by the author for this study to formalise and document prompt engineering steps from real planning data to constrained image generation. The workflow comprises two consecutive layers: (A) a meta-prompt layer (v0-v2) used to elicit, validate, and formalise project inputs into a JSON content profile, and (B) an operational generation layer that converts this profile into a fixed-order inpainting prompt for image output. It follows a seven-step structure (Fig. 3) and is designed so that each stage can be documented and repeated.

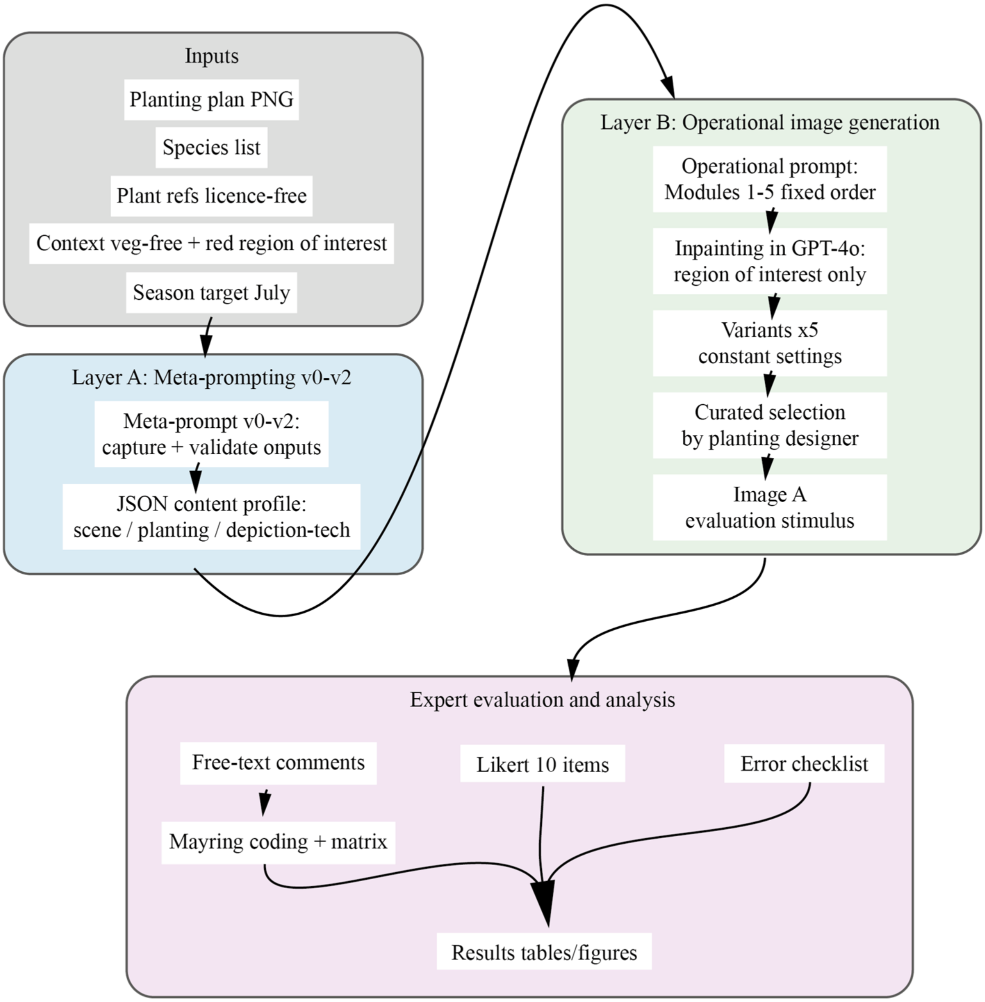

Fig. 3: Workflow development including methodology

1) Data ingestion (Layer A)
The planting plan, species list, licence-free reference plant images, and the vegetation-free context perspective were collected as the input dataset (Section 3.2). The context perspective contained a red-marked target area indicating the planting zone to be visualised.
2) Structured data extraction (Layer A)
Planting information was compiled from the planting plan and species list, including species identity, intended vertical layering (matrix/islands/shrubs/trees), and relevant habitus cues. The licence-free reference plant images were used to derive morphological descriptors (e. g., leaf character, typical habit) to support the subsequent structured description.

3) Creation of a JSON-based content profile (Layer A)
   The extracted planning information was encoded into a structured JSON content profile capturing scene parameters (camera and seasonal target), planting composition and layering, and style and technical requirements (Section 3.6). The JSON content profile functioned as the stable, machine-readable input basis for prompt construction.
4) Prompt construction using modular prompt blocks (Layer B)
   A modular prompt architecture was used to separate planning logic (species composition and spatial/layering intent derived from the JSON profile) from depiction logic (style, atmosphere, and technical instructions). The prompt modules and their assembly are specified in Section 3.5.
5) Image generation via inpainting guided by the red-marked target area (Layer B)
   Image generation was performed in ChatGPT (Version GPT-4o) using the vegetation-free context image as the fixed base. The red-marked target area in the context image served as the sole generation zone. The prompt included an explicit rule to generate vegetation only within the red-marked area and keep all non-marked areas (hardscape, walls, planter geometry, ground plane) unchanged.
6) Iterative prompt precision adjustments (Layer B)
   A small number of targeted prompt adjustments were made during workflow finalisation to reduce unintended changes outside the target area and to stabilise key planting hierarchy cues (see Section 3.8 for the documented adjustment log). Adjustments were limited to specific modules while the remaining prompt structure was kept stable.
7) Production under constant conditions: Five images were generated under constant inputs and settings to document system-immanent variation (Section 3.9).

3.4 Prompt Architecture, Versioning, and Assembly (v0-v2)

Prompt versions v0-v2 refer to the meta-prompt used for project capture and parameterisation. They are distinct from the operational generation prompt that produces the images. To ensure transparent control of image generation, a modular prompt system was used and documented in three stages (v0-v2). v0 defines the prompt baseline (required input data and the domain-specific objective). v1 adds an interactive procedure for structured data elicitation (follow-up questions and interim summaries). v2 denotes an initial/meta-prompt that systematises data elicitation, validation, and the subsequent assembly of the generation prompt. v2 yields a JSON content profile as the structured data basis and an operational generation prompt derived from it. Thus, v2 governs data acquisition and prompt assembly, whereas the operational generation prompt controls the image output only.

v0 specified required inputs and the domain objective (species-specific planting visualisation within a fixed context image). v1 added an interactive elicitation routine (mandatory follow-up questions on camera, season, planting layers, and negative criteria) and interim validation summaries. v2 formalised a structured output schema (JSON content profile with fixed keys for scene/planting/style) and a fixed assembly rule that maps JSON fields to the five operational prompt modules (Modules 1-5), including the region of interest constraint as a mandatory technical rule.

ChatGPT (Version GPT-4o) was used for prompt development (v0-v2) and multimodal image generation (OpenAI 2025). Plant references consisted exclusively of licence-free images and were used only as morphological cues. The operational generation prompt was organised into five modules and always assembled in a fixed order: scene/perspective, planting description, depiction style, season/light, and technical instructions including negative criteria. The generation area was not defined via a black-and-white mask; instead, it was specified by a red-marked target zone in the context image (generate only within the marking; keep all other image areas unchanged).

Key excerpts from the final meta-prompt (v2), including the fixed module structure and control rules, are provided in Box 1.

Box 1: Excerpts from the Final Meta-Prompt (v2, Modular Structure)

Scene & perspective: "Use the provided vegetation-free courtyard perspective as fixed base (eye height ~1.5 m, ~35 mm). Do not change walls, planter geometry, or ground plane."

- Planting structure: "Depict the planting strictly according to the species list and layered logic (matrix / islands / shrubs-trees), using the provided reference images as morphological cues."
- Style: "Specify the intended depiction style (e. g., photorealistic) and require realistic shadows and clear species separation."
- Season & light: "Specify target season and lighting (here: midsummer/July; soft evening light) as fixed depiction parameters."
- Spatial rule (target zone): "Generate vegetation only inside the red-marked planting area in the context image; keep all non-marked areas unchanged."
- Negative criteria: "No people, furniture, vehicles, additional objects; avoid painterly looks, oversaturation, repetition artefacts, distortions, unrealistic plant forms."

3.5 Derivation of the Generation Prompt from v2

The operational inpainting prompt was assembled from the validated JSON content profile using a fixed module order (scene/perspective; planting; depiction style; season/light; technical constraints and negatives). The depiction parameters (photorealistic; July; soft evening light) and the region of interest rule were held constant for the generation of the variants. Any refinements were limited to individual modules and logged (see Section 3.8).

3.6 JSON Content Profile (Result of Data Extraction and Analysis)

The JSON content profile provides the structured data basis for image generation. It was created after reviewing and cross-checking the planting plan, species list, licence-free plant reference images, and the vegetation-free context perspective, and it encodes all planning information relevant for visualisation in a machine-readable format. The profile serves three purposes: consistent transfer of key design parameters into the prompt modules, separation of planning logic (species, layering, distribution, scale) from depiction logic (style, lighting, season), and transparent documentation of inputs to support reproducibility.

The profile comprises three levels:

1) Scene (camera/perspective parameters, target season, lighting assumptions),
2) Planting (matrix/island logic, species lists by vertical layer, optional height/habitus attributes),
3) Depiction/Technical settings (depiction style, desired depth effect, aspect ratio).

An excerpt of the resulting JSON content profile, which serves as the stable input for prompt assembly, is shown below.

JSON Content Profile (Excerpt, Full Version Documented in the Project Archive):


```
{
  "scene": {
    "camera": {"focal_length_mm": 35, "eye_height_m": 1.5},
    "season": "July",
    "lighting": "soft evening light"
  },
  "planting_area": {
    "matrix": {"species": "Hakonechloa macra", "coverage_pct": "60-80"},
    "islands": [
      {"species": "Rodgersia podophylla", "height_cm": "60-100"},
      {"species": "Podophyllum 'Spotty Dotty'"}
    ],
    "shrubs_trees": [
      {"species": "Aralia elata", "form": "multi-stemmed"},
      {"species": "Catalpa bignonioides", "leaf": "large cordate"}
    ]
  },
  "style": {
    "mode": "photorealistic",
    "depth_of_field": "soft",
    "color_profile": "natural midsummer greens"
  }
}
```

(Version ChatGPT-4o) (OPENAI 2025)

An excerpt of the operational inpainting prompt used for the five constant-condition generation runs is provided in Box 2.

Box 2: Operational Generation Prompt (Excerpt):

“A polygonal raised wooden planter … neutral architectural background … focus on the planting structure inside the raised bed … multi-layered planting (upper layer / climbers / mid & ground layers) … ultra-detailed photorealistic planting rendering … midsummer … soft evening light …”
“Generate vegetation only inside the red-marked target area; do not alter concrete walls, wooden planter, or ground plane.”
“Avoid repetition artefacts, distortions, unrealistic plant forms. Negative prompt: exclude people, furniture, cars, artificial objects, over-saturation, painterly effects.”

(*Note: The full operational prompt is documented in the project archive; only a short, verifiable excerpt is shown here* (OPENAI 2025).)

3.7 Image Generation within the Red-Marked Target Zone

Image generation was performed with ChatGPT (GPT-4o) using the vegetation-free context perspective as fixed base image. The planting area was constrained by a red-marked region of interest. The operational prompt enforced the rule to generate vegetation only inside this region and keep all non-marked areas unchanged. No manual post-processing (e. g., Photoshop retouching) was applied to the generated variants. The selection of Image A was based on plausibility review only (see Section 3.10).

3.8 Documented Prompt Refinements

During the development of the final prompt, a few targeted refinements were made and systematically logged to differentiate stable elements from newly introduced stabilising constraints:

1) Vertical hierarchy of a woody species: Following feedback on the intended dominance in the upper vegetation layer ("Catalpa slightly taller"), the planting description was extended with an explicit height/crown specification to reinforce the intended hierarchy.
2) Stabilisation of the non-vegetated context: After an unsatisfactory output the technical module was tightened to prevent changes outside the planting area (keep existing geometry unchanged; generate vegetation only within the target zone).
3) Target-zone rule as a standard constraint: The coupling of the generation area to the red marking in the context image was formalised as an unambiguous rule and adopted as a standard element of the final prompt.

These adjustments should be understood as minimal interventions to improve controllability, without modifying the underlying data basis (planting plan, species list, reference images).

3.9 Image Generation under Constant Conditions (Five Outputs)

To capture system-immanent variability, five outputs were generated under constant settings. Across all five outputs, the following elements were kept identical: the vegetation-free context image (JPEG), the red-marked target zone as spatial constraint, the JSON content profile (final version), the prompt-module order (1-5), and depiction parameters for style, season, and lighting (photorealistic visualisation; midsummer condition; soft evening light). Negative criteria were also held constant across outputs (e. g., no people, furniture, or additional objects; no painterly effects; no oversaturation; no obvious repetition patterns) (Figs. 4-8).

Stochastic effects of generation (e. g., non-fixable random initialisation/seed in the utilised system) could not be controlled deterministically. Differences between the five outputs are therefore treated as system-immanent variance and made transparent through repeated generation.

Figs. 4-8:
Five image variants based on the same prompt (from top left to bottom left) (Version GPT-4o, OPENAI 2025)

3.10 Curated Test Image

For the expert evaluation, a curated test image (Image A, Fig. 4) was defined as the standardised stimulus. From the five images generated under constant conditions, the person responsible for the realised planting design selected the variant that most plausibly represented the intended species composition, vertical layering, and midsummer effect. This selection serves as a practice-oriented safeguard to ensure that the evaluation stimulus is not dominated by a singular outlier (e. g., strong artefacts), but represents the most plausible image under identical conditions.

3.11 Expert Evaluation

The curated AI image (Image A, Fig. 4) served as the standardised stimulus for the expert evaluation (n = 7). The evaluation comprised ten itemised Likert ratings (1-5) covering Botany/Geometry (4 items), Atmosphere (4 items), and Application (2 items); short free-text comments; and an error checklist (none/low/mid/high) plus an overall frequency judgement (rare/sometimes/often).

Free-text comments (n=6) were segmented into meaning units (n=53; one unit = one discrete statement, sentence or clause). A deductive coding scheme was derived from the evaluation framework (Botany/Geometry, Atmosphere, Application) and the predefined error checklist, resulting in a codebook of main categories and sub-codes (MAYRING 2015, 97). Multiple coding of a meaning unit was allowed where a statement addressed more than one aspect. Coding was performed by the author; a second pass was used to check internal consistency. The full coding matrix is documented in the project archive. For transparency, the coding matrix documents all coded units and category assignments (53 meaning units from 6 participants; 11 main categories; multiple coding allowed, resulting in 86 category assignments) (documented in project archive).

4 Results

4.1 Sample Characteristics

Seven participants completed the evaluation (n=7). Reported roles were: landscape architecture practice (n=2), research/teaching (n=2), and dual role "both" (n=3). Countries reported were Germany (n=3), Switzerland (n=1), Austria (n=1), Germany/Switzerland (n=1), and one missing country entry (n=1). Professional experience ranged from 2 to 25 years (median 12). Six participants provided qualitative comments in at least one free-text field (n = 6), while one participant submitted only the Likert ratings (n = 1).

Qualitative comments were segmented into meaning units (n = 53) and coded deductively (MAYRING 2015). Table 1 summarises the main categories by number of meaning units assigned, number of contributing participants, and code assignments (multiple coding allowed).

Table 1: Main category summary of qualitative coding following Mayring: meaning units (n), participants (n), and code assignments

Main category	Meaning units (n)	Participants (n)	Code assignments (n)
Botanical accuracy	14	6	21
Professional applicability	9	6	12
Visual realism	8	4	10
Spatial structure	7	4	9
Professional risk / no-go	8	5	8
Lighting/depth	5	5	6
Future work	5	5	5
Seasonality	5	5	5
Matrix legibility	4	4	4
Uncertainty	4	2	4
Professional risk / miscommunication	2	2	2

Note. Participants (n) refers to commenting participants (n = 6). A total of 53 meaning units were coded; multiple coding is possible, so category totals can exceed 53. Of the 12 predefined main categories in the codebook, 11 occurred in the material.

These three indicators help summarise the qualitative material: meaning units reflect topic density, contributing participants show how widely a theme was shared, and code assignments capture recurrence under multiple coding. In this form, Table 1 provides a compact evidence base for which qualitative themes dominated the expert feedback and how consistently they occurred.

4.2 Rating Outcomes for Image A (Likert 1-5)

Only Image A was assessed with itemized Likert ratings (10 items, scale 1-5; 1 = very poor, 5 = excellent). Across all items, the mean score was 2.34/5 (median 2.20). The three predefined dimensions yielded the following mean scores:

- Botany/Geometry (4 items): 2.21/5
- Atmosphere (4 items): 2.64/5
- Application (2 items): 2.00/5

4.3 Botanical and Spatial Accuracy

Image A was rated on four botany items (Likert 1-5). Mean scores were 2.29 for species identity, 1.86 for morphology, 2.00 for habit & layering, and 2.71 for matrix readability (Tab. 2). Identity, morphology, and habit/layering showed consistently low distributions (≤ 2 by 5/7 participants; 71%), while matrix readability was rated 3 by 5/7 (71%). Comments emphasised coarse recognisability ("*Genus is recognizable, but species and cultivars are not*", P1, translated) and insufficiently differentiated plant structure ("*same stem form across different species*", P6, translated). Several remarks also referenced mismatches in vertical hierarchy (canopy–understorey relations), suggesting that spatial planting logic was not consistently maintained.

Table 2: Likert ratings for Image A by dimension and item (n = 7; scale 1-5 scale, 1 = very poor, 5 = excellent; reported as mean and share of low (≤2) and high (≥4) ratings)

Dimension	Item	Mean	% ≤ 2	% ≥ 4
Botany	Species identity	2.29	71.4	28.6
Botany	Morphology	1.86	71.4	0.0
Botany	Habit & layering	2.00	71.4	0.0
Botany	Matrix readability	2.71	28.6	0.0
Atmosphere	Seasonality	2.86	42.9	28.6
Atmosphere	Lighting	3.00	28.6	28.6
Atmosphere	Depth	2.71	42.9	28.6
Atmosphere	Overall atmosphere	2.00	71.4	14.3
Application	Communication value	2.43	71.4	28.6
Application	Professional trustworthiness	1.57	71.4	0.0

4.4 Atmospheric Quality

Atmospheric quality was assessed with four Likert items (1-5). Mean scores were 2.86 for seasonality, 3.00 for lighting, 2.71 for depth, and 2.00 for overall atmosphere. Ratings showed a split: lighting clustered at 3–4 (5/7; 71%), whereas overall atmosphere concentrated at 1-2 (5/7; 71%). Comments mirrored this pattern, with some participants finding depth/lighting convincing ("*depth and lighting are good and convincing*", P4, translated) and others noting limited spatial impression or inconsistencies ("*hardly any depth*", P3; "*shadows seem inconsistent*", P6, translated). Overall, local rendering cues performed better than the holistic atmospheric impression.

4.5 Practical Applicability and Professional Standard

Mean scores were 2.43 for communication value and 1.57 for professional trustworthiness (Likert 1-5). Trustworthiness was rated 1 by 5/7 participants (71%), whereas communication value clustered at 2 in 4/7 cases (57%). Comments emphasised limited client-facing usability ("*I would not go to a client with this*", P2, translated) and described use primarily for conveying mood to non-experts (P3). One participant noted process constraints, stating that complete planting plans and species lists are often available only late in execution planning (P4). Overall, perceived applicability was constrained by low professional trust and workflow timing.

4.6 A/B Preference (Image A vs. Baseline B)

All participants preferred the baseline visualisation (Image B; Fig. 9) over the AI-based visualisation (Image A; Fig. 4) (**B chosen by 7/7; 100%**).

Preference reasons most frequently referred to higher perceived realism and depth in B, and clearer recognizability of plant habit and hierarchy. Representative statements include: "***more realistic, has spatial depth and high recognizability of plants***" (P3, translated), "***sterile and smooth… B has more depth and looks more realistic***" (P4, translated), and "***better depiction of habitus and hierarchy; better spatial impact***" (P1, translated).

Fig. 9:
Visualisation from the real project served as a baseline and was created conventionally using Photoshop collage

4.7 Error Checklist (Intensity and Frequency)

Participants assessed predefined error types using four intensity levels (none/low/mid/high) and reported overall frequency (rare/sometimes/often). Across participants (n=7), the highest mean severities (coded 0-3 for aggregation) were observed for phenology error (mean 2.00), species confusion (mean 1.86), and geometric scale error (mean 1.57). Intermediate severities were recorded for foreign objects (mean 1.29) and oversaturation/colour shift (mean 1.29), while density/distribution was lower (mean 1.14). Inpainting artefacts showed the lowest severity (mean 0.43) (Tab. 3).

Table 3: Error checklist for Image A (n = 7): mean severity scores (0-3) and share of ratings by intensity level (% high, % mid+high)

Error type	Mean severity (0-3)	% high	% mid+high
Species confusion	1.86	42.9	71.4
Geometric scale error	1.57	14.3	71.4
Phenology error	2.00	42.9	71.4
Density/distribution	1.14	14.3	28.6
Inpainting artefacts	0.43	0.0	0.0
Foreign objects	1.29	14.3	57.1
Oversaturation / colour shift	1.29	14.3	42.9

Intensity distributions illustrate the contrast between error types: inpainting artefacts were rated none by 4/7 participants and low by 3/7 (0/7 mid or high). By comparison, species confusion was rated high by 3/7 and phenology error was rated high by 3/7.

Overall problem frequency was reported as often by 3/7, sometimes by 2/7, and rare by 2/7 participants.

5 Discussion

The evaluation indicates that the AI-generated planting visualisation (Image A) does not meet a professional standard for species-level planting representation in its current form. This is evidenced by low botany/geometry ratings (mean 2.21/5) and consistently low scores for species identity, morphology, and habit/layering (each rated ≤2 by 71% of participants). In parallel, the error checklist shows the highest severities for phenology error (mean severity 2.00/3), species confusion (1.86/3), and geometric scale error (1.57/3). Overall, the quantitative findings match the qualitative statements that recognition is often limited to the genus level, with species and cultivars remaining indistinguishable ("*genus recognizable, but species and cultivars are not*", P1 translated) and that plant structure is insufficiently distinctive ("*same stem form across different species*", P6 translated). Notably, the dominant botanical and geometric error types identified in the ratings and checklist emerged despite the workflow's control measures (JSON-based input structuring, fixed module order, red-marked target-zone-only inpainting, and constant negative criteria across five generation runs; Sections 3.3-3.6), indicating that, within this configuration, semantic plant fidelity and scale relations, rather than typical inpainting artefacts, were the primary limitations.

A key implication is that "plausible planting" in a visual sense should not be conflated with botanical correctness or spatial correctness. While lighting and some local depth cues received comparatively higher ratings (lighting mean 3.00/5; depth mean 2.71/5), overall atmosphere remained low (mean 2.00/5), and participants overwhelmingly preferred the non-AI baseline (Image B; 7/7). Within the photorealistic configuration tested here, the images can appear visually convincing while still containing systematic botanical and geometric errors. This limits their reliability when accurate, detailed species identification, growth form, seasonal state, and scale relationships are critical.

From a professional and ethical perspective, the observed error profile implies a high risk of miscommunication if AI images are used without safeguards in contexts where planting decisions have contractual, ecological, or cost consequences. The low trustworthiness ratings (mean 1.57/5; 71% rated 1) and explicit participant constraints ("*I would not go to a client with this*", P2, translated) indicate that, under the conditions evaluated in this study (GPT-4o model version and the documented prompting and workflow configuration), AI images should not be used as stand-alone evidence for execution-related deliverables (e. g., planting specifications, procurement, maintenance instructions, or construction coordination). In competition or participatory settings, several statements point to disclosure expectations ("*transparency that this is AI-based*", P6 translated), which aligns with the need to prevent stakeholders from interpreting AI images as faithful depictions of specified species, seasonal effect, or spatial proportions.

A limited use case remains plausible under strict framing as an early-stage, low-stakes communication artifact to explore spatial mood or support discussions with non-expert audiences. This requires clear AI labelling, an explicit statement of the intended level of botanical fidelity, and a domain review cross-check against the planting plan and species list. In this study, the low severity of inpainting artefacts (mean 0.43/3) suggests that typical "image editing" artefacts were not the dominant limitation; rather, the limiting factors were primarily semantic plant fidelity (identity, morphology, phenology) and spatial consistency (scale and layering). Accordingly, future technical improvements should prioritize constraint-based control of plant identity and seasonal state, as well as explicit anchoring of geometric scale and vertical hierarchy.

Despite the limited image-level performance in this GPT-4o-based setup, the JSON content profile and modular prompt architecture provide a reusable structure for integrating planning data and documenting constraints in AI-assisted visualisation workflows. However, the present findings reflect a specific model–workflow configuration; replication should therefore compare both updated models and revised prompting protocols under transparently versioned specifications.

6 Conclusion and Outlook

This study evaluated a prompt-to-image workflow for AI-based planting visualisation in a single realised project context with an expert sample (n = 7). The AI-generated image (Image A, Fig. 4) achieved an overall mean rating of 2.34/5 across 10 Likert items (1-5), with comparatively low performance in Botany/Geometry (2.21/5) and Application (2.00/5). Professional trustworthiness was the weakest application criterion (mean 1.57/5; 5/7 participants rated 1), and all participants preferred the baseline visualisation (Image B, Fig. 9) over Image

A (7/7). Taken together, the findings delimit the workflow's current suitability for professional planting visualisation: in this GPT-4o-based configuration, the dominant limitations concern botanical and spatial validity (species identity, morphology, habit/layering, phenology, and geometric scale), rather than image-editing artefacts.

Image A (Fig. 4) and Figures 5-8 were generated with ChatGPT (Version GPT-4o) (OPENAI 2025). AI use was limited to workflow specification, prompt development, and image generation (Section 3.1).

Exploratory re-runs with newer model versions are planned, but require the same expert evaluation protocol before any performance claims.

Accordingly, the primary outlook is empirical and methodological:

1. re-run the same evaluation protocol across emerging multimodal image models to quantify whether botanical and spatial validity improves under identical constraints;
2. broaden testing beyond a single realised case by including multiple planting typologies and contexts to reduce the risk of over-generalisation
3. extend the workflow toward explicit seasonality control and time-based simulation once model/tool affordances allow such constraints to be expressed and checked reliably. Where possible, prompts, structured content profiles, and evaluation materials should be provided as supplementary resources to support reproducibility.
4. Future work should test the workflow in other design-visualisation contexts and benchmark the structured, JSON-guided prompting approach against ad-hoc prompting under controlled conditions to assess gains in transparency and repeatability.

References

BRAIDEN, H., CHAMBERLAIN, B., GEORGE, B. H. & FERNBERG, P. (2025), AI in Practice: Professional Survey Findings from Landscape Architects in North America. JoDLA – Journal of Digital Landscape Architecture, 10-2025, 626-633. doi: 10.14627/537754059.

LIU, X., YANG, N. & TIAN, R. (2024), Reinventing Planting Design in Landscape Architecture: A Generative AI Approach. JoDLA – Journal of Digital Landscape Architecture, 9-2024, 202-209. doi: 10.14627/537752020.

MAYRING, P. (2015), Qualitative Inhaltsanalyse: Grundlagen und Techniken. 12. Aufl. Beltz, Weinheim.

OPENAI (2025), ChatGPT (Version GPT-4o) [Large language model]. https://chat.openai.com/ (28.12.2025).

REKITTKE, J. & HAYLES, A. (2025), Early Days of AI Image Generation in Landscape Architecture. JoDLA – Journal of Digital Landscape Architecture, 10-2025, 634-644. doi: 10.14627/537754060.

SCHROTH, O. & MAIER, A. (2025), Integrating Generative Artificial Intelligence into the Landscape Architecture Design Process. JoDLA – Journal of Digital Landscape Architecture, 10-2025, 665-675. doi: 10.14627/537754063.

ZWANGSLEITNER, F., HABJANIČ, G. & KNEGENDORF, A. (2024), AI as a Tool in the Landscape Architecture Design Process. JoDLA – Journal of Digital Landscape Architecture, 9-2024, 987-994. doi: 10.14627/537752093.