Assessing the Value of Artificial Intelligence in Plant Selection

Benjamin H. George[1], Brent Chamberlain[1], Phil Fernberg[1,2], Paul Gardner[1]

[1]Utah State University/USA, [2]OJB Landscape Architecture//USA

Abstract: Planting design is a complex task that requires designers to consider a multitude of factors, many of which are ephemeral or transitory over the seasons or as a plant matures. Generative artificial intelligence (AI) may streamline the planting design process by introducing increased efficiency during the selection phase. This study examines the ability of GPT 3.5-based applications to identify suitable plants for several different planting schemas and evaluates the reliability of the prompts both within and between samples. Results suggest that both ChatGPT and direct use of GPT 3.5 via API can be a valuable planting design resource, but that there may be significant bias in the results, given the type of model selected. Understanding and mitigating this bias will be important for landscape architects who seek to use ChatGPT or other GPT models via API in planting design.

Keywords: Artificial intelligence, machine learning, planting design, horticulture, large language model

1 Introduction

Plant selection is a quintessential function of landscape architecture, where the art and science of design meet. Choosing the right plants is a pivotal design step as it can impact a range of different goals, whether design oriented (SCARFONE 2010), ecological (HUNTER 2011, RAINER & WEST 2016), or even remedial in nature (KENNEN & KIRKWOOD 2015). The process is complex, requiring a designer to balance many variables such as climate, soil type, maintenance requirements, growth, seasonality, and aesthetic preferences – and knowing which plants match the litany of variables is critically important (ROBINSON 2016). The choices made during plant selection can influence the success and sustainability of a landscape design, making it crucial for landscape architects to employ the most efficient and informed methods for this task.

The proliferation of Artificial Intelligence (AI) within landscape architecture has sparked advancements and facilitated new applications of the technology (CANTRELL et al. 2021, FERNBERG & CHAMBERLAIN 2023). Generative AI, often used in the form of natural language chatbots, could be a promising tool to streamline the plant selection process, potentially revolutionizing how landscape architects choose plants by providing data-driven recommendations, enhancing efficiency, and broadening the scope of design possibilities. The exploration of AI-driven plant selection is still nascent, leaving several questions about the usefulness, reliability, and accuracy of large language models (LLMs) for plant selection.

In this study, we examine the use of generative AI in plant selection for landscape architecture, with a focused examination of the Generative Pre-training Transformer (GPT) language model (RADFORD et al. 2018). We explore the following research questions: 1) How do different implementations of GPT, such as the ChatGPT web interface versus the application programming interface (API), influence the accuracy, variety, and distribution of plant recommendations, and 2) how do plant recommendations generated from GPT correlate among

Journal of Digital Landscape Architecture, 9-2024, pp. 277-284. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-752-6, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537752026.

different planting design themes? We anticipate this study will provide insights into how designers can better use AI tools, like ChatGPT, and the implications and limits of the technology.

2 Methods

In our study, we utilized two distinct approaches to interact with GPT 3.5 for plant selection: the *Manual* (via ChatGPT Web app) method and the *API-based* method. The two methods facilitate a direct comparison between distinct interaction modes with ChatGPT. In the manual approach, three testers independently used ChatGPT through its web interface, using an identical approach to limit variations caused by prompt history and ontological variations inferred by individual differences of how they interpret specific requests. A new chat thread was used with each prompt to maintain the integrity of each interaction and prevent influence from previous requests. Each participant used an identical set of 15 prompts for various ecological, stylistic, functional, and aesthetic theme (Tab. 1); an example being: “Provide me with exactly 20 perennials that are suitable to grow in Logan, Utah that have low water requirements. Display the result with the Common Name first, separated by a pipe, then the Scientific name. Each pair should be on a different row and rows should not be numbered." Each prompt was run ten times by each of the three testers, resulting in a list of 600 plants (three independent sets of 200 plants).

Table 1: The 15 prompt factors by theme

Ecological	Stylistic	Functional	Aesthetic
Low-water	English	Foundation Plantings	Fall foliage color
Shade	Japanese	Parking strip	Summer blooms
Clay Soil	Modern	Hedge	
High pH	Mediterranean	Small commercial	
Attract Butterflies			

Parallel to the manual method, we employed the API-based method to interact with ChatGPT to also produce 30 requests for 20 plants using the same 15 prompts. The API allows a user to interact with ChatGPT directly from code (our instance was implemented with Python). The API allows for a range of different models that are inaccessible through the web interface (OpenAI, 2024). For this study, we used the “text-davinci-003” engine which differs from the default web interface engine “text-davinci-002”. Both are part of the GPT base models used for their early phase public release. They only use data before September 2021 and can only process limited set of outcomes compared to the newest models as of the date of this publication. There are now seven GPT 3.5 models and nine GPT 4 models available (other variations also exist). The GPT 4 models require a subscription. The API can implement multiple prompts and automatically generate a report for all the prompts, resulting in significant time savings. Both the API and manual mode allow the user to select which model they would like to use. For the manual mode this needs to be inserted into the URL directly.

To assess the effectiveness of ChatGPT in generating plant palettes, the returned plant lists were evaluated for accuracy, variety, and distribution of plants returned. Variety and distribution are defined as statistical measures. Variety quantifies the number of different plants, while distribution is measured as the standard deviation across all plants. Accuracy establishes the confidence a designer can have in the results returned by ChatGPT, this measure was assessed qualitatively as the authors are very familiar with the plants suitable for the tests. Variety and distribution address a major criticism of AI regarding an underlying bias that each model has. In planting design, this bias could be reflected in how often certain plants are identified (variety), while the distribution of the proportion of plant recommendations could indicate a bias toward certain common species. Diversity, richness, and evenness are other measures that could be employed for this study, but these tend to measure ecological function, whereas our prompts include other factors (stylistic, function, aesthetic). For this paper we have not included measures of diversity indices.

All plant lists were reviewed to remove redundancies (ie. *Rudbeckia* x *grandiflora* and *Rudbeckia grandiflora* were combined). Two plant lists (low-water and shade) were reviewed for suitability to sample the accuracy of ChatGPT in suggesting suitable plants, representing 12% of the total number of responses.

3 Results

The manual method produces a relatively consistent plant palette within an identical prompt. Out of 600 potential plants, the number of unique species for any one set of prompts varied from 63 to 161 unique species, with 11 of the 15 prompts containing less than 120 individual plants. In comparison, the API returned a greater variety of plants, with unique species per prompt ranging from 121 to 172. Results are provided in Table 2. When using the manual method, 98% of the plants recommended by ChatGPT were suitable, based on the metrics of the prompt. The API was less accurate with its recommendations, with 85% of plants being suitable. Each list was then analysed to determine the total number of times each plant was recommended.

While the manual method produced an impressive list of plants for each prompt, ChatGPT heavily favours a select number of plants. For instance, of the 20 most common plants for the butterfly-attracting plant list (which has 90 unique species) the three most recommended plants appeared in nearly every response and the six most recommended appear in 24 responses. It is not until the 25th most recommended plant that a plant appears in less than 33% of responses. 50 plants appeared in only 3, 2, or 1 responses, meaning over half of the total species appear in 10% or fewer responses (Fig. 1).

Table 2: Results organized by factor. Number of plants indicates total unique species returned across all tests.

Factor	*Method*	*# plants*	*µ*	*SD*
Butterflies	API	132	1.51	1.32
Butterflies	Manual	90	2.22	2.58
Clay	API	130	1.58	1.38
Clay	Manual	111	1.8	2.43
Commercial	API	152	1.32	1.09
Commercial	Manual	128	1.51	2.25
English	API	144	1.39	0.96
English	Manual	95	2.11	2.48
Fall	API	172	1.16	0.59
Fall	Manual	138	1.45	1.79
Foundation	API	144	1.39	0.84
Foundation	Manual	126	1.64	2.29
Hedge	API	150	1.33	1.00
Hedge	Manual	161	1.25	1.24
High pH	API	140	1.42	1.00
High pH	Manual	103	1.94	2.61

Factor	*Method*	*# plants*	*µ*	*SD*
Japanese	API	150	1.33	0.75
Japanese	Manual	119	1.69	2.60
Low-water	API	132	1.52	1.35
Low-water	Manual	63	3.18	3.29
Mediterranean	API	150	1.31	0.96
Mediterranean	Manual	118	1.70	2.29
Modern	API	156	1.27	0.67
Modern	Manual	104	1.89	2.27
Park Strip	API	147	1.36	0.55
Park Strip	Manual	86	2.31	2.89
Shade	API	121	1.65	1.50
Shade	Manual	73	2.68	3.16
Summer Blooms	API	125	1.61	1.59
Summer Blooms	Manual	68	2.95	3.13

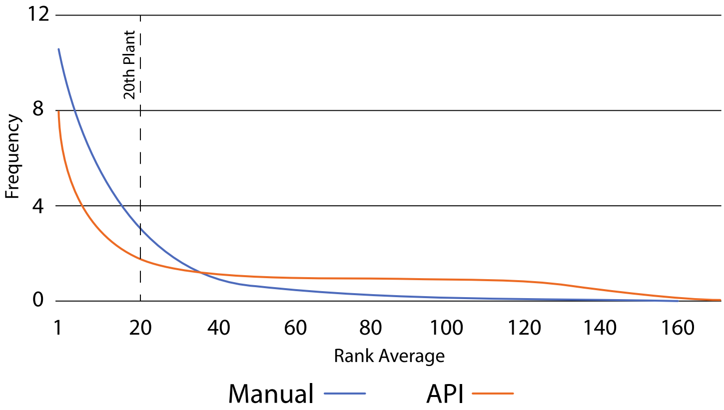

Fig. 1: The standardized average distribution of all plant species by rank order for both methods. The 20th most commonly returned plant appeared in 3.11 responses using the manual method, while only appearing in 1.93 responses using the API.

On average, for a typical list of 20 plants, 30% are in nearly every response, 60% come from a subset of 30 plant species, and the remaining 10% show an inconsistent pattern. In contrast to the manual method, the API had a distinctly different pattern of plant recommendations. For the API, no plant appears on more than 30% of responses, and nearly 90% were suggested only once or twice. To quantify the variety of plants, the standard deviation was calculated for the total count of unique species in each list, which shows that the API (SD=1.04) was substantially more diverse than the manual method (average SD=2.48). In a typical dataset, a lower standard deviation would usually indicate less variety, but in the case of these data, the frequency of recommendations are highly skewed toward fewer plants with the manual method, while other plants are recommended at a very low frequency.

Finally, correlations were calculated between different prompts. The total count of plants for each prompt were used to identify if Chat GPT was recommending plants based on the input requirement or where really just reusing plants with little attention to the requirements. A selected sample of paired lists was chosen where it was theorized that a clear correlation should or should not exist based. For instance, it would not be expected that plants from the low-water and English garden lists would correlate because these represent two distinctly different climactic growing conditions. Seven correlations were tested, and all returned a result as would be expected (Tab. 3). Statistical significance and the *R-value* was used to determine the test. We acknowledge some Pearson's r *values* are somewhat low (e. g. 0.5), but some of the variability would be expected.

Table 3: Example scenarios comparing two factors (e. g. "Low-water", "English") showing expected correlations compared with statistical results. T/F indicates if expectation matches the statistical outcome.

Comparison	**Expected**	**Result**	**T/F**
Low-water x English	No correlation	r(24) = -.021, p = .325	T
Low-water x Clay soil	Correlation	r(40) = .526, p = <.001	T
English x Modern	No correlation	r(35) = -.034, p = .846	T
English x Mediterranean	No correlation	r(26) = .089, p = .666	T
Low-water x Parking Strip	Correlation	r(48) = .738, p =<.001	T
Butterflies x Summer Blooms	Correlation	r(38) = .489, p = <.002	T
Shade x Parking Strip	No correlation	r(13) = -.05, p = .87	T

4 Discussion

Initial results suggest that ChatGPT exhibited high accuracy, consistency, and task suitability in its plant recommendations. It also showed a significant response bias towards certain plants or plant groups. This is especially apparent with the manual method where, for instance, *Perovskia atriplicifolia* (Russian Sage) was in the top 20 plants for all but two factors (Japanese and Shade) and was the most recommended plant for seven of the factors. At the same time, many commonly used plants were rarely recommended on lists with a relevant factor (e. g. *Agastache*, *Epilobium*, or *Yucca*, in the water-wise list). The API was significantly less biased, but still favoured a select number of plants at 3-4 times the rate of other plants. However, the API was also less reliable in recommending suitable plants. This would suggest that the manual interface has additional guardrails or reinforcement. Expressions of AI "hallucinations" were almost non-existent, with only two responses – *Hylotelephium* 'Autumn Blaze' and *Dolichis cormoides* (Fig. 2).

The correlation results showed similarity or dissimilarity as expected between the lists and factors tested (Tab. 3). This suggests that ChatGPT meets expectations within a factor, which is encouraging for the use of ChatGPT in plant selection tasks defined by discrete factors. Yet while ChatGPT appears to be able to produce viable recommendations for plant palettes, it is important to note that these results may be biased, producing a limited selection that may lead to excessive conformity in the plant palette, especially if using the base manual model.

Fig. 2: AI representation of planting beds using the plant palette generated by ChatGPT for the low-water prompt

We suspect this is due to statistical models used within the different GPT models, especially since we tested the GPT Base model. On a single site this may not be an issue; however, if ChatGPT were to be used widely across the industry, the conformity bias would become noticeable and could have impacts on both aesthetic and environmental qualities in the built environment. Designers can mitigate this bias through quality control of their prompts for each project and developing protocols to supplement ChatGPT's suggestions when appropriate to improve variety in the planting palette. For instance, intentionally repeating the same prompt to maximize the number of individual plant species returned could be an effective strategy to identify lesser-used plants. However, this requires a systematic approach that not all designers may utilize.

It is important to note that lack of novelty in planting palettes is not only an LLM problem but an ongoing problem in the profession (RAXWORTHY 2013). Similar observations about bias could be made about human designers who are constrained by their knowledge of plants. When, where, and how one was educated in planting design, as well as who they practice with, tend to favour certain palettes, follow prominent design trends, and produce a degree of conformity in many designs. Whether using GPT models, learning from planting design courses in university, or through limited exposure to a variety of planting palettes, designers always run the risk of excessive conformity.

The challenge with AI is that it may become a tool that helps to inform planting designs and it may become less flexible with planting variety unless specifically told to make additional recommendations. With the low cost of these tools, designers could become reliant on these models, reinforcing the reuse of similar plants, especially if the project is in a novel environment to the designer. As these technologies progress, and if they become more common industry-wide and through educational exercises, it would be wise to develop strategies to employ GPT or other LLMs and continuously evaluate their effectiveness. Since most designers, at least in the near term, will not be using the API method (which offered a higher variety of plants), it will be important to quality control GPT for plant recommendations. It is clear the number of models continues to expand, so we recommend users attempt to find a combination of models that offers both statistically consistent plant palettes and others that offer a wider range of variety as a way of learning about new plants.

Beyond accuracy, variety, and distribution, the use of AI in planting design may have significant impacts on processes and workflows. First, based on its high reliability, AI may be able to facilitate landscape architects to more easily work in unfamiliar climates and ecosystems. Often a firm might hire a local landscape architect or horticultural specialist to develop the planting plan when working outside the firm's home region, or it would be necessary to invest a large amount of research time to select the plant palette. This outsourcing or investment could potentially be automated and make it possible for more firms, and smaller firms, to compete for a greater number of projects across the globe, and thus disrupt some historically consistent business norms.

5 Conclusion and Outlook

Based on our specific task and factors, ChatGPT appears to be an efficient and effective tool to aid a designer in plant selection. Conformity bias in the responses, especially with the manual method, may be concerning, especially if designers become unquestioningly reliant on ChatGPT for plant selection. A seasoned designer with a good knowledge of plants could use ChatGPT to help identify a wider plant selection, while bringing a gravitas to the evaluation of the suitability of the proposed plants. However, this study only utilized one style, one location, and one site condition at a time, nor did it compare quality or efficiency in GPT outputs with those of an unaided human designer. More research needs to be done in multi-factor prompting and current practice vs human-AI comparisons so as to more rigorously vet how useful this would be toward a range of planting design challenges. Additional research should also examine how specific a prompt can be before the model starts to return erroneous responses, as preliminary testing suggests that adding too much information to a prompt returns inaccurate results that responds inconsistently to the prompt. Furthermore, there needs to be an examination of the aesthetic quality of plant palettes generated by ChatGPT, as suitability or variety in plant selection does not guarantee a pleasing aesthetic outcome. As we refine our understanding of how best to both interact and use ChatGPT or other LLMs for planting design, we expect this application of AI will provide valuable affordances to landscape architects.

References

CANTRELL, B., ZHANG, Z. & LIU, X. (2021), Artificial intelligence and machine learning in landscape architecture. In: The Routledge Companion to Artificial Intelligence in Architecture. Routledge.

FERNBERG, P. & CHAMBERLAIN, B. (2023), Artificial Intelligence in Landscape Architecture: A Literature Review. Landscape Journal, 42 (1), 13-35. https://doi.org/10.3368/lj.42.1.13.

HUNTER, M. (2011), Using Ecological Theory to Guide Urban Planting Design: An adaptation strategy for climate change. *L*andscape Journal, 30 (2), 173-193. https://doi.org/10.3368/lj.30.2.173

KENNEN, K. & KIRKWOOD, N. (2015), Phyto: Principles and Resources for Site Remediation and Landscape Design. Routledge.

OpenAI (2024), Models. OpenAI. https://platform.openai.com.

RADFORD, A., NARASIMHAN, K., SALIMANS, T. & SUTSKEVER, I. (2018), Improving language understanding by generative pre-training.

RAINER, T. & WEST, C. (2016), Planting in a Post-Wild World: Designing Plant Communities for Resilient Landscapes. Timber Press.

RAXWORTHY, J. (2013), Novelty in the Entropic Landscape: Landscape architecture, gardening and change. Doctoral Dissertation.

ROBINSON, N. (2016), The Planting Design Handbook. Routledge.

SCARFONE, S. C. (2010), Professional Planting Design: An Architectural and Horticultural Approach for Creating Mixed Bed Plantings. John Wiley & Sons.