



WIDENING ADOPTION OF GENERATIVE AI DESIGN IN CONSTRUCTION THROUGH EXPLAINABLE ARTIFICIAL INTELLIGENCE

Sanaz Zarghami, Longzhi Yang, and Pablo Martinez
Northumbria University, Newcastle, UK

Introduction

Generative AI (GenAI) has revolutionized design workflows by enabling the creation of optimized, adaptive forms using parametric models. Despite its promise, GenAI algorithms often function as black boxes, obscuring the rationale behind their outputs and impeding trust and broader adoption. Explainable AI (XAI) addresses this by improving transparency and interpretability of AI decisions. This study explores the integration of Large Language Models (LLMs), such as ChatGPT, as an XAI technique within generative AI design algorithms (GAID). Through a case study in Grasshopper using Wallacei 2.7, it is examined how LLMs can (1) interpret GAID outputs, (2) explain optimization decisions, and (3) guide users in applying sensitivity analysis as an XAI method. This study hypothesizes that integrating Large Language Models (LLMs), such as ChatGPT, into generative design workflows can enhance the explainability of optimization results and increase user trust in generative AI decisions. Findings from this initial case study indicate that LLMs such as ChatGPT may assist users in understanding the behaviour of generative design algorithms through the application of basic XAI methods.

Methodology

A case study is conducted employing a multi-objective optimization problem within Grasshopper. Using Wallacei 2.7 (see figure 1), a cylindrical form is optimized aiming to maximize volume and minimize surface area. Outputs included charts, grids, and data points related to geometry (See Figure 2 and 3). These were systematically used in iterative ChatGPT (GPT-4o mini and GPT-4.5) conversations. Prompts were crafted to evaluate:

1. LLMs' understanding of GAID principles.
2. Their capability to locate optimal solutions within visual datasets.
3. Their ability to perform and explain sensitivity analysis.

Responses were assessed for clarity, depth, and visual interpretability. Iterative prompting refined both textual and graphical outputs (See Figure 4). It is acknowledged that optimizing a cylinder presents a simplified case,

because the purpose of this initial stage is not to focus on the complexity of the form, but to evaluate the capability of LLMs in interpreting the behaviour of GAID algorithms. The use of two counteractive objectives provides a scenario that tests whether LLMs can explain an optimization process that appears inherently conflicting. While the chosen form may seem abstract, it offers a controlled environment for investigating the interpretability of LLM outputs. In future stages, this method will be applied to more complex case studies that closely resemble real-world architectural contexts.

Results

Prompts were designed in a progressive structure—beginning with general queries about GAID logic, followed by targeted instructions to test interpretation accuracy and visual generation capability (see figure 5). Dataset characteristics such as radii, height, and volume were extracted from Wallacei outputs and fed into ChatGPT in tabular or screenshot formats. All experiments were conducted with consistent parameters to ensure repeatability. Initial prompts showed GPT-4o mini could interpret Wallacei's outputs and describe optimization logic. Switching to GPT-4.5 improved visual analysis and data sensitivity. Key findings include:

Understanding GAID: ChatGPT successfully explained evolutionary logic in Wallacei outputs and interpreted optimality graphs (see figure 6).

XAI Guidance: When asked to identify optimal solutions or interpret grid-based results, ChatGPT accurately pinpointed generations and individuals based on numerical clues (see figure 7).

Sensitivity Analysis: Through stepwise prompting, ChatGPT performed sensitivity analysis by comparing variations in radius and height against volume outcomes (see figure 8). It also produced different charts showing input-output relationships (See Figure 9 and 10).

Out of 9 visual analysis tasks, GPT-4o mini failed to generate a correct chart in 4 cases, whereas GPT-4.5 succeeded in 3 out of 5 attempts after prompt simplification. Qualitatively, responses were scored on clarity, correctness, and completeness using a 3-point Likert scale by the authors, with GPT-4.5 averaging 2.6 compared to 1.8 for GPT-4o mini.

In the initial stage of this research, a taxonomy was developed to classify GAID algorithms commonly used in construction and the XAI methods applicable to them (see figure 11). This taxonomy was published in a paper presented at last year's EC3 conference (Zarghami et al., 2023).

Conclusion

This study offers an exploratory evaluation of ChatGPT's potential to assist in applying XAI methods to GAID algorithms. While the results suggest promising directions, the findings are limited to a controlled, single-case study and should not yet be considered generalisable. Broader validation across varied GAID problems and design contexts is required to substantiate these early insights. Also, its ability to autonomously generate visuals remains limited without precise human input. As LLMs evolve, their integration via plugins in platforms like Grasshopper could enable real-time feedback and enhanced decision-making transparency. However, this vision remains speculative at this stage. It assumes significant advancements in API integration, real-time processing, and prompt responsiveness, factors not explored in this paper. A key limitation of this study is the absence of comparative baselines. Future work should investigate additional XAI methods, improve LLM visualization consistency, and test other LLM variants. Another promising direction is the evaluation of open-source LLMs (e.g., LLaMA, Mistral) to reduce reliance on proprietary models like GPT. Such comparison would help identify biases, performance disparities, and practical trade-offs in applying LLMs for GAID interpretation.

References

- Anand, K., Wang, Z., Loog, M., & van Gemert, J. (2020). Black Magic in Deep Learning: How Human Skill Impacts Network Training. <http://arxiv.org/abs/2008.05981>
- Avital, M. (2007). Innovation through generative systems design. *Information Systems Journal*, 17(3), 231–254.
- Bang, Y. et al. (2023). A Multitask, Multilingual, Multimodal Evaluation of ChatGPT. <http://arxiv.org/abs/2302.04023>
- De Cremer, D., Bianzino, N. M., & Falk, B. (2023). How Generative AI Could Disrupt Creative Work.
- Esposito, M. et al. (2025). Generative AI for Software Architecture. <http://arxiv.org/abs/2503.13310>
- Gerlings, J., & Shollo, A. (n.d.). Reviewing the Need for Explainable Artificial Intelligence (xAI).
- Hu, X., Liu, A., & Dai, Y. (2024). Combining ChatGPT and knowledge graph for explainable machine learning. *Journal of Engineering Design*.
- Khan, M. A. et al. (2025). Smart buildings: XAI for energy management. *Energy Reports*, 13, 2066–2081.
- Kookalani, S. et al. (n.d.). Building design automation from generative design.
- Love, P. E. D. et al. (2023). Explainable AI in Construction. *Advanced Engineering Informatics*, 57.
- Robertson, J. et al. (2024). A generative AI prompt protocol. *Business Horizons*, 67(5).
- Sokol, K., & Flach, P. (2021). Foundations of Explainable AI. <http://arxiv.org/abs/2112.14466>
- Wang, W. et al. (2021). Data-driven low-energy urban design. *Renewable Energy*, 179, 2016–2035.
- White, J. et al. (2023). A Prompt Pattern Catalog for ChatGPT. <http://arxiv.org/abs/2302.11382>
- Zarghami, S., Kouchaki, H., Yang, L., & Martinez Rodriguez, P. (n.d.). Explainable AI in Generative Design for Construction.

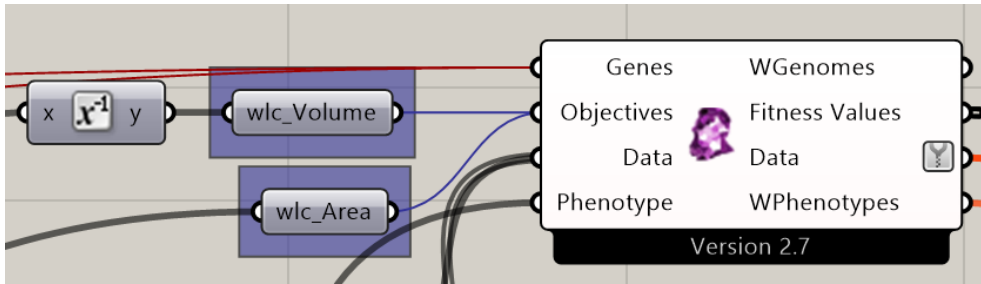


Figure 1:Wallacei plugin in grasshopper environment



Figure 2: Results from wallacei

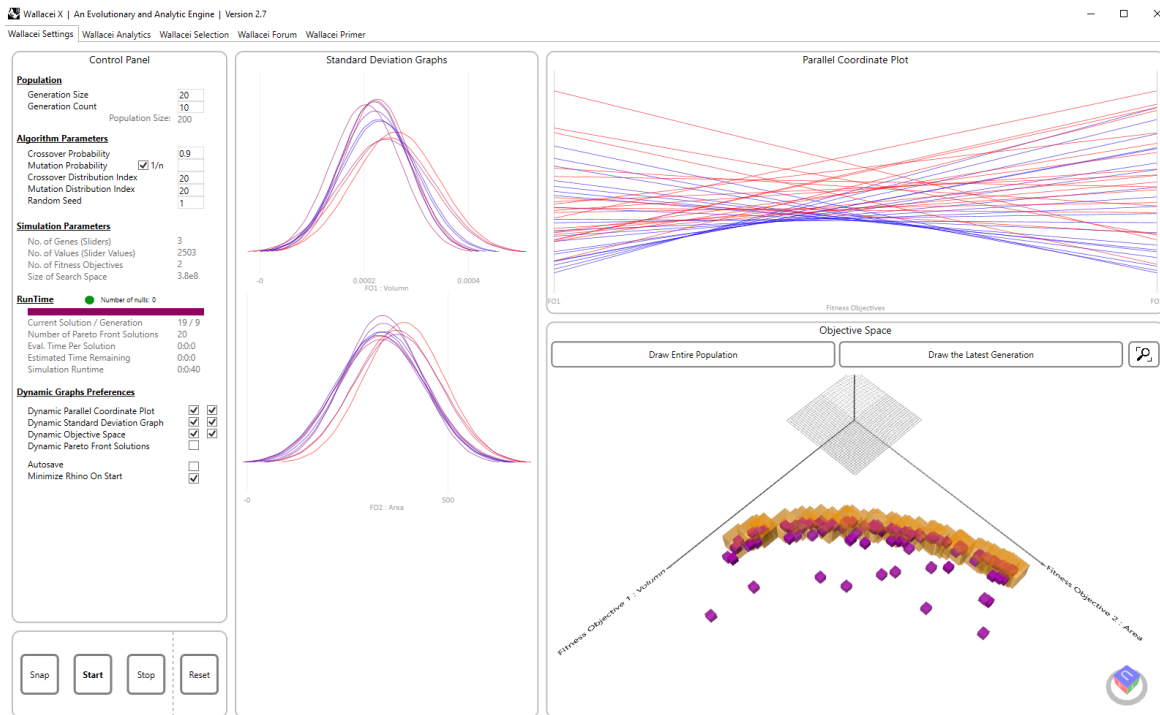


Figure 3:Results for the defined optimization problem

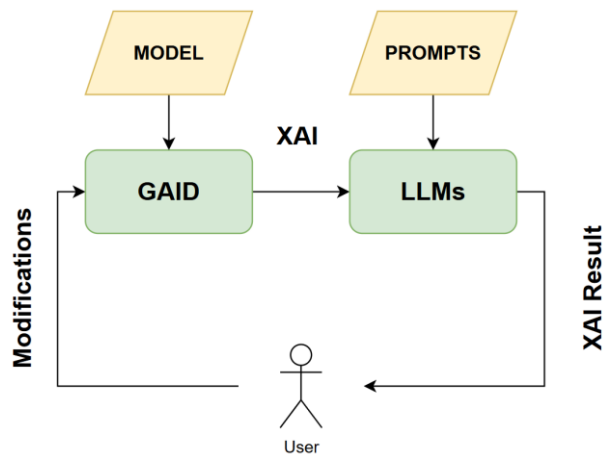


Figure 4: Research Methodology

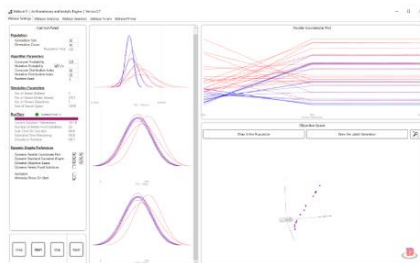
Key Areas in the Screenshot:

1. Control Panel:

- **Population:** The "Generation Size" is 20, meaning the algorithm is working with 20 solutions per generation.
- **Algorithm Parameters:** The algorithm is set to use crossover, mutation, and mutation distribution probabilities. These parameters control how the population of solutions evolves over time.
- **Simulation Parameters:** Details about the number of values, objectives, and size of the search space for the optimization process.

2. Runtime:

- **Current Solution / Generation:** This shows the current generation (19) and total number of generations (9), indicating the progress of the optimization.
- **Simulation Runtime:** The optimization process has been running for about 1 second.



Hey there. This is a screenshot of the Wallacei plugin result. Please explain it for me.

Figure 5: ChatGPT response to test its ability to identify Wallacei

How Wallacei Decides on Optimum Results:

The optimization algorithm in Wallacei uses **evolutionary principles** to explore the solution space. The algorithm evaluates each solution by calculating its performance based on your defined objectives (maximize volume, minimize area).

- The **Pareto front** represents the best trade-offs between volume and area. Solutions on the Pareto front cannot be improved in one objective without degrading the other.
- As the generations evolve, Wallacei uses **crossover and mutation operators** to combine solutions and introduce diversity, guiding the population towards optimal trade-offs.
- The **Pareto front solutions** are considered optimal because they represent the best compromises between the two conflicting objectives (volume and area).

How to Understand the Algorithm's Decisions:

1. **Examine the Pareto Front:** Look at where the **solutions are clustered** in the **Objective Space** plot. Solutions that lie closer to the upper-right corner of the plot (higher volume and lower area) are likely optimal solutions according to the algorithm.
2. **Check Convergence:** If the **standard deviation graphs** show decreasing variability, the population is converging, meaning the algorithm is focusing on the most promising solutions.
3. **Trace Evolution in the Parallel Coordinate Plot:** Follow the lines in the **Parallel Coordinate Plot** to see how solutions evolve across generations. Solutions that consistently improve in both volume and area would be considered optimal. You can analyze if any specific solutions are improving in one objective while degrading in the other.

Figure 6: ChatGPT's explanations for algorithm's function

The optimal solution you're referring to, with the following details:

- $1/\text{volume} = 0.000284$
- Area = 235
- Height = 30
- Upper circle dimension = 5.67
- Lower circle dimension = 6.54

appears in **Generation 9, Individual 19**, clearly visible in the provided screenshot.

Figure 7: ChatGPT response to locating the optimal solution

Key Insights:

1. **Height (Data Input 1):**
 - Changes in **height** appear to **have a significant impact on the volume (FV1)**, as increasing the height leads to higher volume values in the solutions. Height is also likely contributing to the overall size of the cylinder, which would directly influence the volume.
 - The **volume (FV1)** shows notable fluctuations in response to changes in height, while **area (FV2)** is not as greatly affected by height variations alone.
2. **Radii (Data Input 2 and 3):**
 - The **radii (upper and lower circles)** primarily affect the **area (FV2)**. Changes in these values show more significant effects on the **surface area**, as increasing the radii increases the cross-sectional area of the cylinder, which leads to a larger surface area.
 - **Area (FV2)** is much more sensitive to changes in the **radii** compared to the volume.

Conclusion:

- **Height** (data input 1) seems to have a **greater influence on maximizing the volume (FV1)**. This is because height directly scales the cylinder's volume.
- **Radii** (data inputs 2 and 3) are more influential in controlling the **area (FV2)** of the cylinder, as expected for cylindrical shapes, where the area is heavily determined by the size of the cross-sections.

Thus, from the screenshots and the observed patterns, **height** (data input 1) appears to be more significant for affecting the **volume (FV1)**, while **radii** (data inputs 2 and 3) have a larger effect on **area (FV2)**.

Figure 8: ChatGPT's response to prompt 3

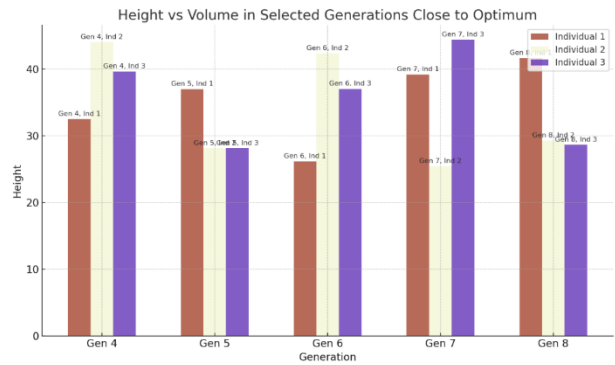


Figure 9: Initial and incorrect sensitivity analysis of the results for the defined optimization problem

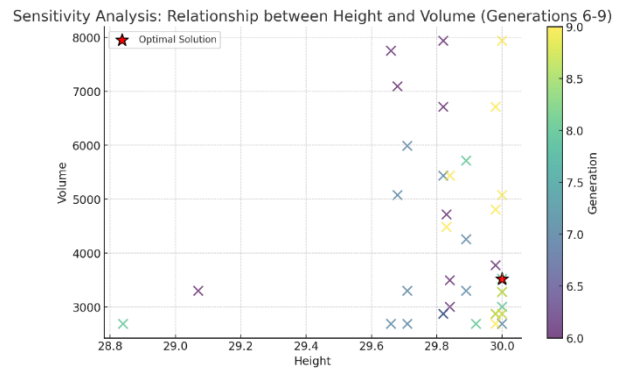


Figure 10: The final and correct visual content produced by ChatGPT

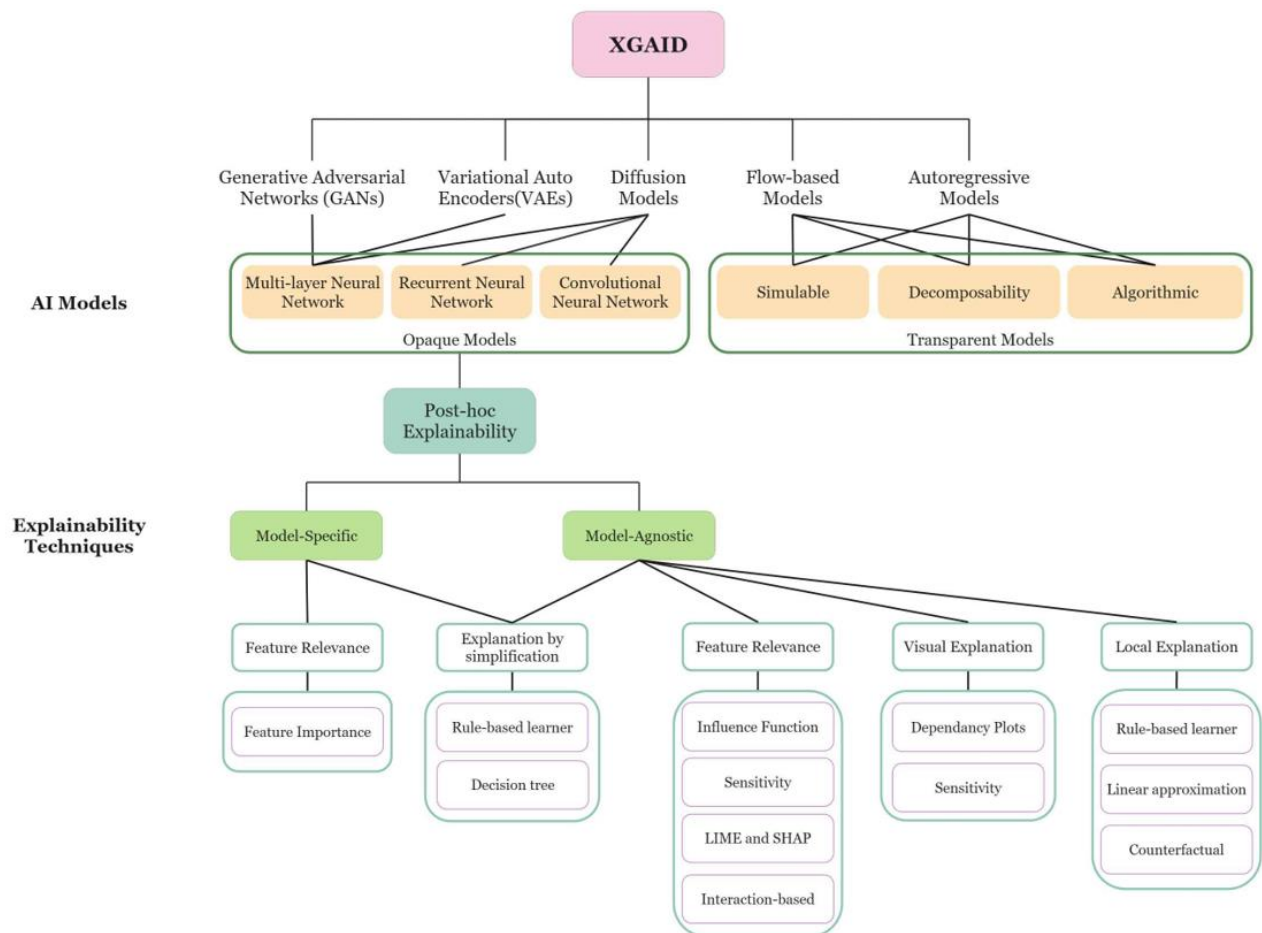


Figure 11: XGAID Taxonomy