



Reducing Bias for Evidence-Based Decision Making in Design

Matthias Standfest^(✉)

Archilyse AG, Flüelastrasse 31B, 8047 Zürich, Switzerland
standfest@archilyse.com

Abstract. This paper presents a strategy to help designers confronted with a massive flow of new technologies, exhaustive regulations, and design requirements. It summarises the existing proposals for AI-driven design development strategies, and lists frequent pitfalls like the focus on local optima or the lack of backpropagation. It identifies the main source of bias in generative design as a lack of detail and contextual complexity. The paper introduces an augmented AI process for preparing real-world data and its meta information to be used in design processes (BIM, GIS, external statistics including information such as rental price or spatial cognition). This evidence-based approach for deriving verifiable fitness functions presents a way to create holistic designs that reflect the complexity of today's built environment by using a posteriori and unbiased statistical models to substitute existing speculative categorisations. Hence, it allows avoiding naïve and overfitted solutions, and it inverts the dominant paradigm of automated generation and manual curation.

1 Old Wine in New Bottles: Design Development in the AI Era

Thomas Kuhn¹ is frequently cited on the effects of a Human-in-the-loop (HIL) approach to cover the complexity of true data-driven design creation (Carrier 2006). His research highlights the importance of subjective elements in the comparative evaluation of theories as well as its negative effects. According to him, the accumulation of failures in processes with significant subjective contributions can be attributed to the lack of empirical evidence and the resulting difficulty of clear decision-making in such matters.

This unparalleled level of complexity (Vrachliotis 2012, p. 163) comes with a high degree of responsibility and unavoidable cognitive overload (Matthews et al. 2020), while existing problem-solving strategies reveal themselves as inherently flawed: they are mainly based on human experience which is gathered over time. This process is

¹ According to Thomas Kuhn, competing paradigms are frequently incommensurable. That is, they are competing and irreconcilable accounts of reality. Thus, our comprehension of science can never rely wholly upon “objectivity” alone. Science must account for subjective perspectives as well, since all objective conclusions are ultimately founded upon the subjective conditioning/worldview of its researchers and participants.

based on the assumption that environmental requirements are not changing too fast – an assumption which obviously outlived its validity (Alexander 1978; Joedicke 1976; Stiny and Mitchell 1978; Vitruv 2015). As a result, designs struggle with an increasing imbalance between workload and creative freedom while still needing to adapt to technological requirements rather than the other way around. This imbalance is further nourished by the fact that the real estate sector is one of the least digital sectors in western economy (Kane et al. 2015). Therefore, coordination of projects always involves considerable additional efforts in communication. So, how can one reduce the cognitive overload for designers while empowering them with evidence-based knowledge to help them design better?

1.1 Current AI-Driven Design Development Strategies

Currently, the dominant paradigm in computer-aided architectural design (CAAD) is to improve the design space exploration of generative designs (Hester et al. 2018; Reisinger et al. 2021). To reduce the cognitive load many researchers try to arrange the design space in a two-dimensional grid. Unsupervised clustering or self-organising maps are used frequently, as are the concepts of Pareto fronts. Because this process is complex and difficult, ready-made “Volume Solvers” are increasingly popular. Dominant players (like spacemaker, skyline, testfit, architectures [sic!], proving ground, engrain, archistar, sitesolve, metabuild, digital blue foam, etc.) provide Pareto fronts designers can pick from. This comparative methodology is a common tool to iteratively improve competing designs.

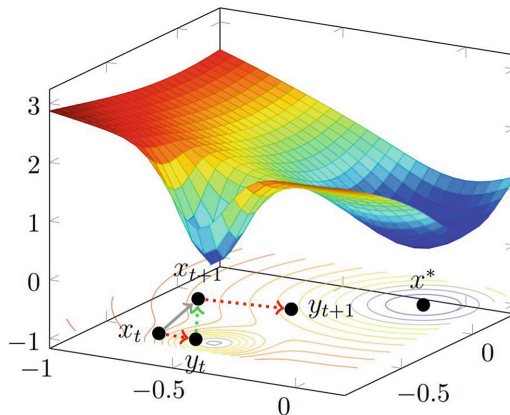


Fig. 1. Stochastic gradient descent (SDG). Image courtesy by Kleinberg et al. for an alternative view: When does SGD escape local minima? ICML 2018/02/17.

While architectural design is often represented as a linear optimisation process (Bre et al. 2016), it is not. Architectural design is more related to a stochastic gradient descent process [Fig. 1] in a hyper-dimensional solution space (De et al. 2020). Every design is a sample in this solution space, and the level of detail is the amount of dimensions to probe in this space. Of course both, the amount of samples and of dimensions, correlate with the probability of finding the global optimum. The hypothesis is that algorithms can be used to generate an abundance of geometric variants which are derived from gradually changing the underlying numeric input parameters. While the number of versions generated this way indeed might be infinite, they do not cover the whole design space but merely a subset which is reflecting the specified input-output mapping. So rather than approaching a local or global optimum of the solution space by iterative re-design, as has been done so far, now the generative design methods predefine a solution space which can be searched instead. Such searches are mostly implemented as fitness functions for multi-criteria optimisations (MCO) and reproduce the best set of input parameters for the given parametric design. Often these results get falsely interpreted as global optima for the given design task but given the limitations of the investigated solution space (which can be read as bias) these claims are clearly exaggerated. If one only analyses a single subsection of the solution space, albeit very thoroughly (which is exactly what is done when using a simplified shape grammar approach), one will most likely not find a global optimum. To mitigate the pitfalls of the local optima, domain experts have already come up with their own “stochastic” process: the design competitions where topologically very different setups can be compared.

Additionally, the synthesis-analysis loop of design is strongly affected by the analysis tools where errors in the analysis are typically propagated into the design. And prominent MCO strategies for building layouts (Sangani 2021) lack both scope of analysis and scope of design space. The oversimplifications they are built upon introduce additional bias and reduce the necessary design diversity even further. Most current MCO tools used in design development are not leveraging true alternative concepts. Also, they strip designers of both creative and interpretive influence. MCO are trojan horses claiming to reduce the HIL issue just by adding new bias. This is caused by insufficient linking of complex subjective parameters to the geometry of the design studies at the very beginning. Current volume optimisers are merely overfitting without taking necessary information into consideration. But designers should be able to comparatively evaluate the alternative theories with as little bias as possible. Simply put, the machine should rank or select, and the architect should design rather than the other way around. This challenges current strategies, but it is much more in line with the historic developments of other engineering disciplines.

1.2 Issues with False Design Automation

This false understanding of better design development through automation causes many issues:

- a) **Focus on the early stage.** Currently, most effort is put on early-stage design and analysis (e.g., spacemaker).

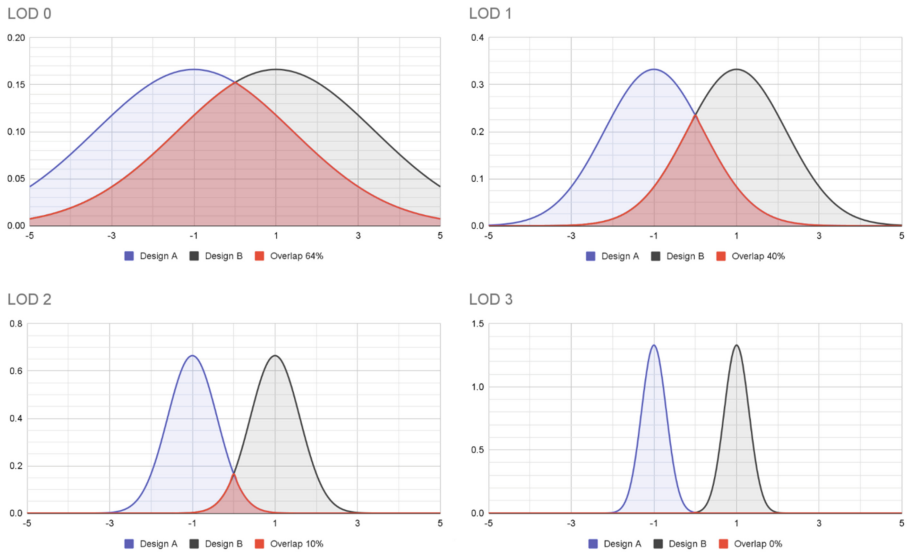


Fig. 2. Level of development in decision making. This conceptual illustration highlights the effect of comparing probability distributions of different standard deviations. This is due to the fact that in low levels of detail the final performance for each simulation dimension can only be predicted with low accuracy, while with higher levels of detail the accuracy of the performance prediction with simulation increases. The overlapping areas are those, where it cannot be decided which of the designs is finally going to perform “better”. This uncertainty area (the overlap) is smaller, the higher the level of detail of the underlying model, which of course is limited according to the MacLeamy Curve. Image courtesy by the author.

The MacLeamy Curve [Fig. 3] tells us that changes applied soon cost less money. The lower the level of development (LOD) of a design, the cheaper it is to apply changes. [Fig. 2].

MacLeamy Curve

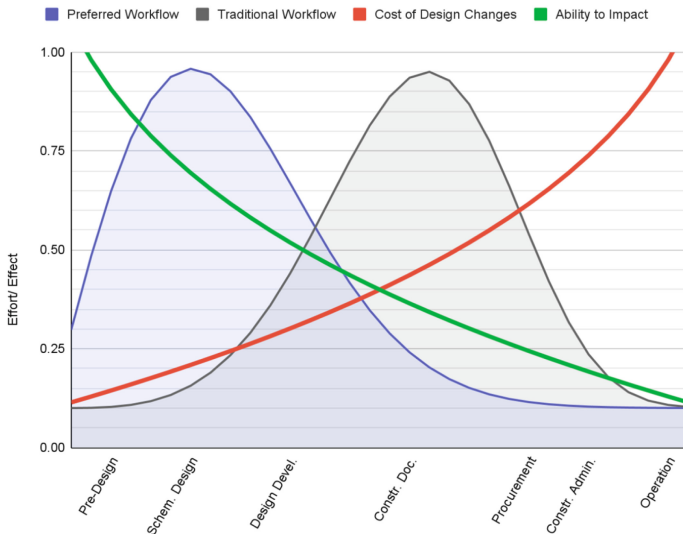


Fig. 3. MacLeamy Curve. Image courtesy by the author.

But low LOD also means there is no clear distinction between alternate designs. On the contrary – selecting designs based on low LODs is prone to error because too much can change [Fig. 2]. Any simulation based on these low granularities tends to be misleading. Additionally, many simulations do not consider enough contextual information like the built environment.

- b) **Lack of holistic analysis.** The majority of simulations in architecture come with interfaces to the frequently used design environments, middleware like Speckle or plugin ecosystems. This way, simulating many physical or geometrical aspects of designs becomes a commodity. And all these deterministic simulations share a materialistic origin – their scope hardly contains social or psychological evaluations (complex relationships between architecture and humans like rental prices or social and environmental impacts). The disregard of such factors leads to paradoxical results: even the most energy-efficient building is hardly sustainable, if it is vacant for years because the local market needs were not met properly. This perfectly illustrates Kuhn’s initially mentioned critique when subjective, difficult-to-measure parameters are ignored (Carrier 2006). Clearly this drift towards materialistic simulations has a reason: psychological or social implications are much harder to measure. At least with *a priori* models (Halevy et al. 2009).

Only data-intensive statistical modelling of contextual features has the potential to reduce bias with *a posteriori* models to predict such psychological and social effects. This mitigates the negative effects of incomplete evaluation scopes the domain still struggles with. On the other hand, ignoring complex data results in misleading optimisations. So the lack of a large data set for holistic modelling is evident.

- c) **Feature weights in multi-dimensional space.** Not all features are scaled and distributed equally. This can cause irregularities in analysis and design. To identify strengths and weaknesses, it is required to normalise the parameters of a dataset. Otherwise, features with outliers are overrepresented. Such a normalisation can only be done if a sufficient amount of real-world data is provided for benchmarking.

It can thus be stated that the Pareto fronts generated by current generative design tools are far from optimal, even misleading at worst. To mitigate this, focusing on more details, a larger scope, and normalised scales is recommended. The method we propose allows for improving these parameters at a novel level of cost-effectiveness.

2 Method

Two extremes of the application at hand can be illustrated: either one creates evidence-based design processes which empower designers and strengthen the creativity of the designers or one is reducing their roles to become mere curators of deterministic solution sets. The resulting hypothesis is that offering an HIL-friendly AI solution could help designers to regain decision power. Also in the context of explainable and collaborative processes AI augmentation shows potential for efficiency and transparency increase in data homogenisation compared to black box ML projects. In these regards, the data pipeline developed is able to support real estate decision makers with validated and reliable benchmarks. Methodologically, these steps are the basis for better data in an automated design development:

- a) **Homogenisation process via augmented AI process.** The most important step is to convert any plan from any file (including 2D raster and vector files) into structured BIM data (IFC 4.0 LOD 200) with an economically efficient augmented AI process. A HIL-strategy is commonly used in industrial applications to improve the accuracy of object recognition when confidence levels of ML predictors are low. At its core, this process is focusing on the validation of the geometrical information with multiple sources of truth, like provided room lists or governmental GIS data regarding the building hull geometry. The resulting IFC files of the reconstructed Digital Twins serve as a validated single source of truth (SOT) for the subsequent analysis. [Fig. 4].

Objects and areas of these IFC files are not only labelled and attributed according to widespread standards, they are modelled with similar generative processes. This homogeneous data structure is important to avoid geometric/semantic bias in subsequent statistical analysis.

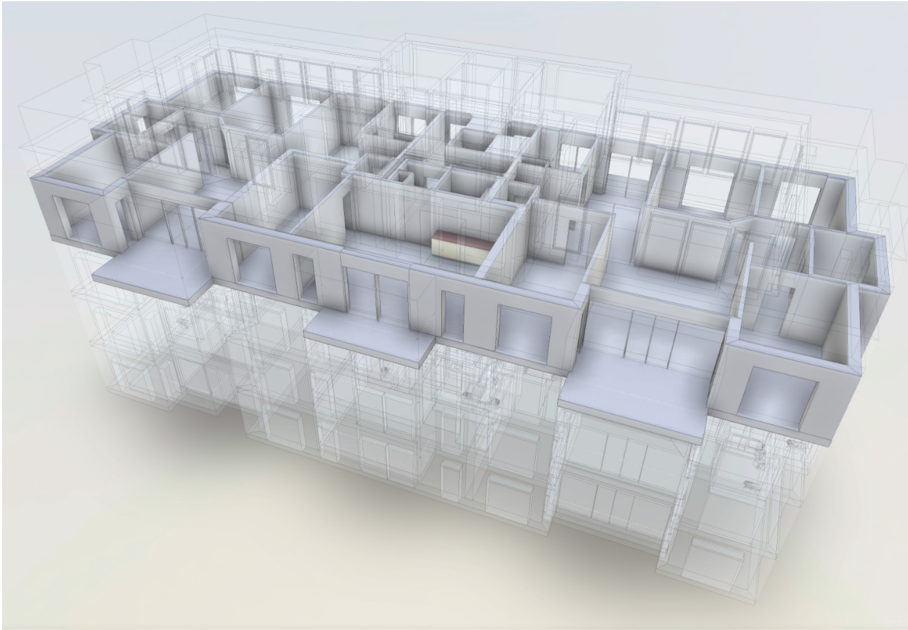


Fig. 4. IFC file rendered in autodesk online viewer. Image courtesy by Archilyse.

- b) **Data enrichment.** Once the validated digital twin of a building is created, it has to be positioned within a GIS model of the environmental geometry. [Fig. 5]. This way, semantically annotated geometry containing building hulls, topographic information, and additional map layers of a 20 km x 20 km area is fused with the annotated geometry of the digital twin located at its centroid. This contextual information of the built environment is gathered from commercially available data sources, open government data sources and open-source repositories. The resulting data has a high level of detail regarding interior and exterior geometry. This is a prerequisite for detailed spatial analysis – and mandatory if bias known from purely building-hull driven models needs to be avoided.
- c) **Data densification.** After generating a comprehensive semantically annotated geometric representation of the digital twin and its environment, different simulations to calculate spatial qualities are applied. [Fig. 6]. Using a hex-grid in a 25 cm resolution to define the location of the observation points has shown a high number of benefits. For every point the following simulations are calculated: semantic spherical viewshed (how much of which label can be seen?), 3D Isovist, traffic noise (at different times of the day), natural light (including atmospheric illumination and direct sunlight every 2 h throughout the year), accessibility (multiple centrality metrics like betweenness or closeness). In addition to these 50 simulations per observation point, 25 discrete features for every room are computed (like area type, net area, perimeter length, largest inscribed rectangle, furnishability, etc.) and additional 68 features per

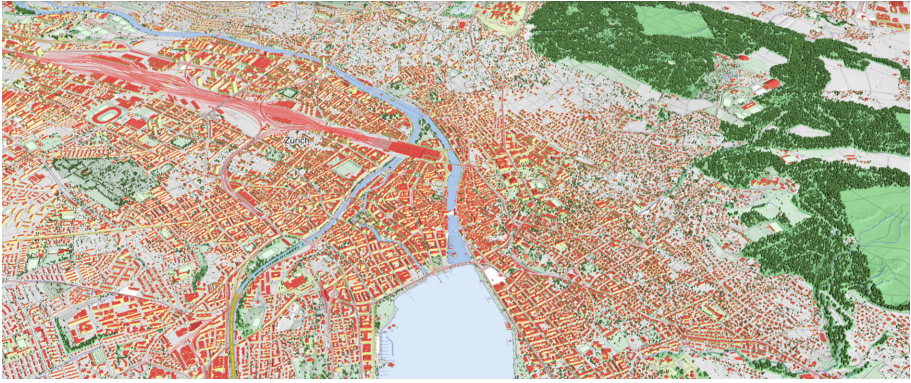


Fig. 5. Screenshot showing the topography of Switzerland, using open-gov geo data. Image courtesy of SwissTopo.

floor level (projected wall areas, number of doors, surface of bathroom walls, etc.). For individual rooms it has proven to be useful to also aggregate the values of the observation points into seven figure summaries (STD, Min., Mean, Max., P20, P50, P80). Covering a multitude of simulations allows for reducing the bias in the data set drastically.

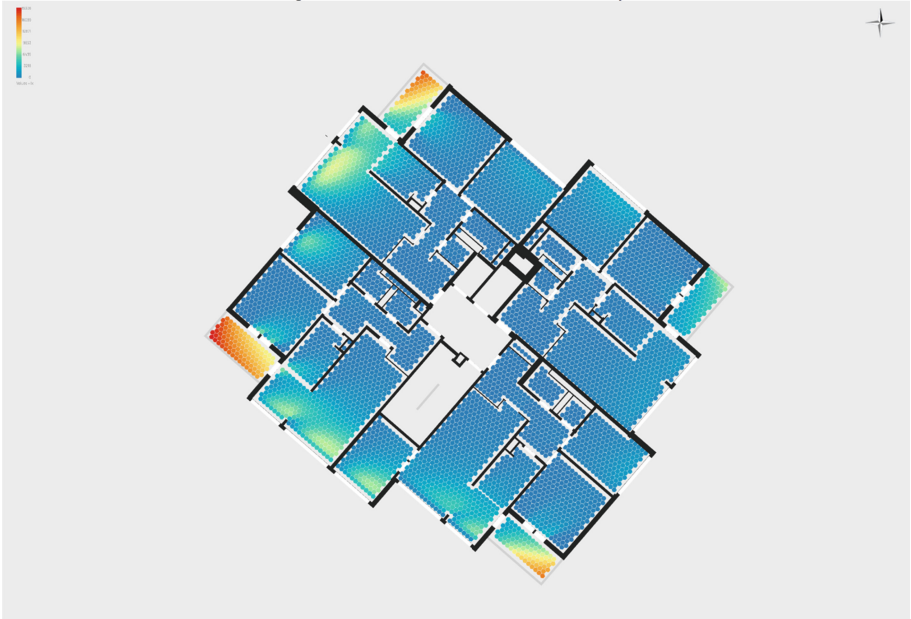


Fig. 6. High-resolution heatmap. Image courtesy by Archilyse.

- d) **Data normalisation (benchmarking).** Having access to a large data set of built apartments (more than 7 million m²), this information is used to normalise the results of the simulations. This way, all the values aggregated per room have been converted from raw values into percentiles. Resulting in values between 0 and 1, indicating how many percent of the data set have lower or higher values regarding the respective figure. This allows for intuitive outlier detection (everything smaller than 0.1 or larger 0.9), interpretation of significant strengths and weaknesses and, of course, identification of average values close to the expectation (median). [Fig. 7].

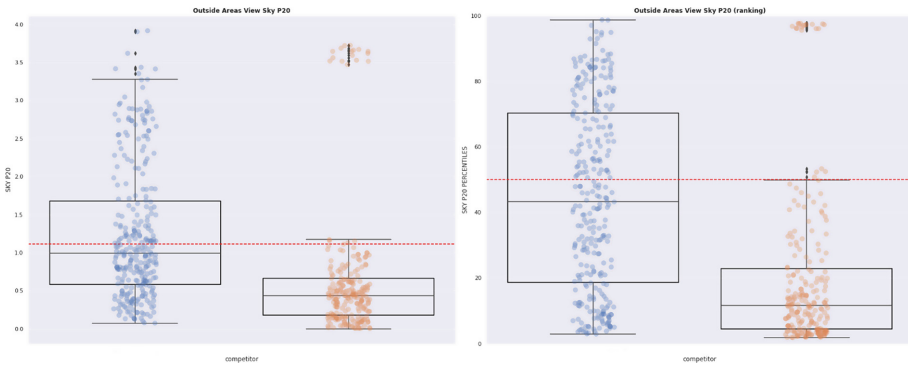


Fig. 7. The minimal skyview of all private outside areas of two competing designs (blue and orange). On the left in steradian, and on the right as their respective percentile rank. Image courtesy by Archilyse.

3 Interpolation/Application

The data generated can be used in two ways: Either directly as features for established processes (like building masses for cost estimation), or via supervised learning to enhance statistical models. In both cases, the processes benefit from a standardised data input and normalised feature vectors.

- a) **Estimation of costs.** Applied to building cost estimations (CE), the provided data can be used to increase speed and precision of the process. Building masses are labelled and constructed in a homogeneous way, so comparing different architectural designs is not impacted by the different drawing styles of the different architects. Usually, BIM-based CE suffers from labelling mistakes caused by the complex UX of off-the-shelf CAAD environments. Consequently, the costly and tedious process of quality control for building mass extractions can be replaced by the presented method. Additionally, the level of detail and feature aggregation (e.g., m² of bathroom surfaces for the cost of tilings) is a magnitude higher than the usually used purely m³-based CE approaches.

- b) **Estimation of revenue.** State-of-the-art methods for automated valuation models (AVM) are based on some flavour of hedonic regression. Traditionally, categorical input related to certain architectural qualities is provided on a 5-point scale (low, below average, average, above average, high). Both the number of features as well as their resolution are limited to human perception. The presented method ensures a holistic scope of highly accurate features regarding spatial qualities that are outperforming both precision and accuracy of existing manual processes by far. This drastically improved dataspace has reduced noise and higher density levels. AVMs based on these provided features have shown a reduction in the respective prediction error by more than 50%.
- c) **Estimation of sustainability.** For questions related to sustainability, it is important to have accurate building masses and areas, but it is even more important to have data about physical exposure of the individual areas. Calculating the grey energy needed for a design is strongly correlated with the amount of materials used. And using digital twins for thermal analysis is superior to a building-hull-only approach since the internal arrangement is impacting the overall thermal behaviour. Additionally, having information about solar exposure of the different rooms can provide further insight on ways to optimise heating and artificial light. And the generated solar profile of the rooms can be used for predictive heating control too. Furthermore, vacancy rates are a waste of grey energy. So, reducing vacancy is improving the energy footprint of buildings drastically. By using real-life energy consumption of buildings and the digital twin strategy presented above, some stakeholders have even been able to successfully deploy statistical models to derive the thermal resistance coefficients of the different building materials – and hence the proper thermal model of the building.
- d) **Analysis for judging competitions.** For choosing from a set of competing designs, especially in real-world scenarios with high investment risks, stakeholders involved tend to hire a high number of experts for design quality assessment. The provided features, including meta-models like cost and revenue estimations, help to drastically reduce the communication overhead and to increase both speed and accuracy of this process. Detailed comparative analysis is significantly more objective when compared to its conventional counterpart. Reducing the bias in this step is directly reducing the planning and decision-making risk involved otherwise.
- e) **Aggregating into fitness functions for better comparative analysis.** When applied in generative design loops, the generated benchmarks and simulation results provide a much more accurate indication of architectural qualities for fitness functions. Multi-criteria optimization (MCO) using the provided benchmarks and meta-models can generate significantly improved design proposals.

4 Conclusion

A process for built and to-be-built architecture was presented that allows for drastically increased accuracy in related decision-making. A novel approach was presented to include features previously subjectively decided upon. The dominant paradigm of partial optimisation was challenged. The dominant paradigm of automated generation and manual curation was inverted.

Acknowledgements. We thank Archilyse AG Zürich who provided insight and expertise that greatly assisted the research. We thank Anja Mutschler for her comments that greatly improved the manuscript.

References

- Alexander, C.: *A Pattern Language. Towns, Buildings, Construction*. Oxford University Press, Oxford (1978)
- Bre, F., et al.: Residential building design optimisation using sensitivity analysis and genetic algorithm. *Energy and Buildings* **133**, 853–866 (2016)
- Carrier, M.: Ziel und struktur der methodologischen theorien. *Deut Z Philos* **54**(3), 387–400 (2006)
- De, S., Maute, K., Doostan, A.: Bi-fidelity stochastic gradient descent for structural optimization under uncertainty. *Comput. Mech.* **66**(4), 745–771 (2020)
- Halevy, A., et al.: The unreasonable effectiveness of data. *IEEE Intell. Syst.* **24**, 8–12 (2009)
- Hester, J., et al.: Building design-space exploration through quasi-optimization of life cycle impacts and costs. *Build. Environ.* **144**, 34–44 (2018)
- Joedicke, J.: *Angewandte Entwurfsmethodik für Architekten*. Karl Krämer, Stuttgart (1976)
- Kane, G.C., et al.: *Strategy, not Technology, drives Digital Transformation*. MIT Sloan Management Review and Deloitte University Press, *Becoming a Digitally Mature Enterprise* (2015)
- Matthews, G., et al.: Stress, skilled performance, and expertise: overload and beyond. In: Ward, P., et al. (eds.) *The Oxford handbook of expertise*, pp. 490–524. Oxford University Press (2020)
- Reisinger, J., Knoll, M., Kovacic, I.: Design space exploration for flexibility assessment and decision making support in integrated industrial building design. *Optim. Eng.* **22**(3), 1693–1725 (2021)
- Sangani, P.: Know your architects. *10 Firms That Use Artificial Intelligence Technologies for AEC. Re-Thinking the Future* (2021). <https://www.re-thinkingthefuture.com/know-your-architects/a3670-10-firms-that-use-artificial-intelligence-technologies-for-aec/>. Accessed 01 Apr 22
- Stiny, G., Mitchell, W.J.: *The palladian grammar*. *Environ. Planning B* **5**, 5–18 (1978)
- Vitruv, 2015: *Zehn Bücher über Architektur*. *De Architectura Libri Decem*. Wiesbaden: Marixverlag
- Vrachliotis, G.: *Geregelte Verhältnisse: Architektur und technisches Denken in der Epoche der Kybernetik*. Springer, Vienna (2012)