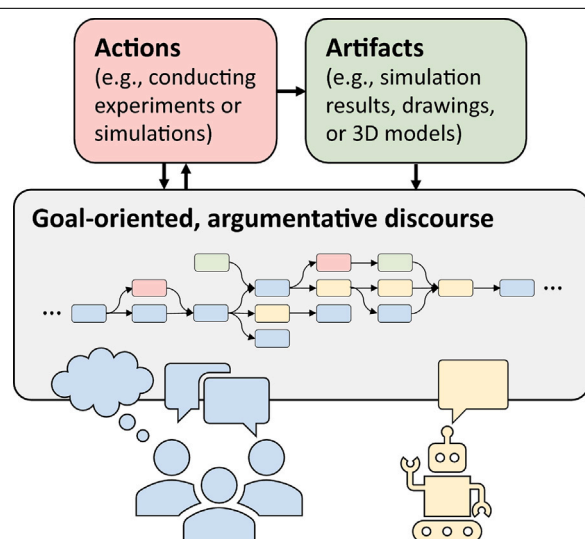


Perspective

Opportunities for large language models and discourse in engineering design

Jan Göpfert^{a,b}, Jann M. Weinand^{a,*}, Patrick Kuckertz^a, Detlef Stolten^{a,b}^a Institute of Energy and Climate Research – Techno-economic Systems Analysis (IEK-3), Forschungszentrum Jülich GmbH, Jülich, 52425, Germany^b Chair for Fuel Cells, RWTH Aachen University, Faculty of Mechanical Engineering, Aachen, 52062, Germany

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Product development process
 Conceptual design
 Design methodology
 Design generation
 Natural language processing
 Foundation models
 Multi-modal models

ABSTRACT

In recent years, large language models have achieved breakthroughs on a wide range of benchmarks in natural language processing and continue to increase in performance. Recently, the advances of large language models have raised interest outside the natural language processing community and could have a large impact on daily life. In this paper, we pose the question: How will large language models and other foundation models shape the future product development process? We provide the reader with an overview of the subject by summarizing both recent advances in natural language processing and the use of information technology in the engineering design process. We argue that discourse should be regarded as the core of engineering design processes, and therefore should be represented in a digital artifact. On this basis, we describe how foundation models such as large language models could contribute to the design discourse by automating parts thereof that involve creativity and reasoning, and were previously reserved for humans. We describe how simulations, experiments, topology optimizations, and other process steps can be integrated into a machine-actionable, discourse-centric design process. As an example, we present a design discourse on the optimization of wind turbine blades. Finally, we outline the future research that will be necessary for the implementation of the conceptualized framework.

* Corresponding author.

E-mail address: j.weinand@fz-juelich.de (J.M. Weinand).<https://doi.org/10.1016/j.egyai.2024.100383>

Received 18 March 2024; Received in revised form 30 May 2024; Accepted 2 June 2024

Available online 4 June 2024

2666-5468/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

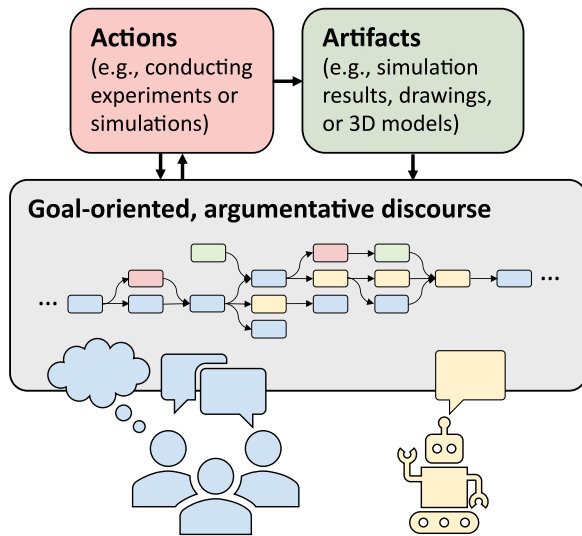


Fig. 1. The engineering design process depicted as a goal-oriented, argumentative discourse formed by inter- and intrapersonal communication in which machines can participate. As part of this discourse, external actions are invoked that in turn inform it.

1. Introduction

Engineering design can accelerate the transition to climate-neutral energy systems and help mitigate climate change. Power generation, transformation, transmission, and end-use all involve the use of technologies. Increasing their energy efficiency reduces greenhouse gas emissions. Design that lowers the costs of green technologies helps increasing their market penetration. In addition, improving other attributes can accelerate adoption and reduce negative environmental impacts.

Large language models (LLMs) have transformed the field of natural language processing (NLP) and increasingly have an impact outside of academia. LLMs already dominate almost every benchmark in natural language understanding (e.g., [1]) and current research focuses on extending them by means of other modalities such as images, videos, or sensor signals [2,3]. Many research fields, such as medicine [4] and chemistry [5], are discussing the future implications of these models for their field. The engineering sciences are a knowledge-intensive domain, which is likely to experience great progress via the adaptation of recent methods developed in the NLP community. In this paper, we argue that foundation models such as LLMs can be used for creative reasoning tasks in the engineering design process, complementing and integrating existing computational methods such as topology optimization.

First, we provide engineers with a summary of the recent advances in NLP and outline which aspects of engineering design have been digitized thus far (Section 2). In Section 3, we place goal-oriented, argumentative discourse at the center of the product development process (see Fig. 1) and propose making the reasoning steps explicit in the form of a new digital artifact. On this basis, we describe how LLMs and multi-modal foundation models can assist in the design discourse (Section 4) and outline interesting directions of future research (Section 5). The presented ideas are transferable to other contexts in which creativity and reasoning play an important role, such as scientific discovery in general.

2. Background

As this article bridges multiple domains, not all of which may be familiar to most readers, we provide a thorough background for LLMs and foundation models as well as the digitization of engineering design.

2.1. LLMs and foundation models

In its early years, NLP applications relied on hard-coded rules. When the body of digitally-available text increased statistical methods became more prominent. Roughly from 2013 onward, NLP began to be dominated by machine-learning methods, in particular deep learning ones. [6]

Building on the distributional hypothesis, which states that words tend to have similar meanings if they occur in similar contexts [7], words (and other text units) have been represented in dense multi-dimensional vector representations using self-supervised learning [8–10]. These word embeddings allow for the similarity between words to be calculated based on vector distances, and as they constitute a rich feature, they have been a popular component of many NLP pipelines. However, these word embeddings have been static; that is, they are context-independent (e.g., the word ‘seal’ has the same vector representation regardless of whether it refers to the animal or machine element). Contextual word embeddings based on language models resulted in significant improvements over a wide range of NLP benchmarks [11,12].

In 2017, the transformer architecture was proposed [13], and has been the dominant neural network architecture for language models ever since, replacing previous approaches based on feed-forward models, convolutional neural networks, recurrent neural networks, and long short-term memory networks. The transformer architecture was originally proposed in the context of machine translation and thus consists of an encoder and a decoder. Both comprise a stack of transformer blocks, each of which consists of a multi-head self-attention mechanism and a fully connected feed-forward network. In the decoder, the multi-head attention is masked and followed by an attention layer that attends to the output of the encoder. Compared to recurrent and convolutional layers, the self-attention layers are less computationally complex (for sequence lengths that are smaller than the embedding dimension) and do not suffer from the vanishing gradient problem over long sequences. Additionally, the self-attention mechanism is highly parallelizable as opposed to the sequential nature of recurrent layers. [13]

Various large language models have been proposed utilizing either the encoder (e.g., the BERT family of models [12,14]), decoder (e.g., the GPT [15–17], BLOOM [18] or LLaMA [19] family of models), or both (e.g., the T5 and UL2 model family [20–22]) of the transformer architecture. Language models are typically trained on predicting the next token or masked tokens in a sequence in which a token can be a word, character, or a sub-word unit, with most models using sub-word-tokenization. Because models can be trained on this objective in a self-supervised setting, large unlabeled corpora (such as Wikipedia, book corpora, and Common Crawl data) can be used for training. The computational advantages of the transformer architecture further aid increasing model sizes and training data volumes.

Increasing the amount of data, parameters, and computation, several emergent abilities of LLMs have been discovered [23]. Sufficiently large models are capable of in-context learning [17] and chain-of-thought reasoning [24]. Previously requiring fine-tuning on downstream tasks (e.g., named entity recognition or question answering), LLMs can now yield decent performance on new tasks by merely including a task description and few examples in the input [17]. Whether these abilities emerge suddenly in a sharp transition at a certain scale or in a gradual and predictable way is still subject of scientific debate [25]. With instruction fine-tuning and reinforcement learning from human feedback, the generated responses to a prompt aligned more with the user’s intent, removing the need for careful prompt selection [26,27]. To solve tasks beyond the capabilities of LLMs in isolation, they have been trained to use tools for which the input and output can be represented as text, such as a calculator or Python console [28]. Further work on the agent-like behavior of LLMs combines reasoning and acting capabilities [29] or add self-reflection capacities [30].

The NLP community saw great advances in a wide range of benchmarks using LLMs. In addition to models operating on textual input alone, recently, multi-modal models that also process other modalities such as images, videos, and/or sensor-signals have become a focus of research [2,3]. Abstracting the LLM concept to other modalities, training procedures, and so forth, the term foundation model was coined [31].

2.2. Computational engineering design

Today, the digitization of engineering design is well advanced. In the past, technical drawing was performed on drawing boards until software for computer-aided design (CAD) was developed in the second half of the twentieth century, which is generally adopted today. Currently, we utilize finite element analysis, topology optimization, design-support tools for additive manufacturing, and more. With model-based systems engineering and digital twins, the product development process became centered around digital models. Virtual and augmented reality enables visualization and interaction with designs. Due to the breadth of the field, only a brief overview of advances in computational engineering design can be given here, focusing on design generation, design strategy learning, and NLP (in particular LLMs) for engineering design.

Generative adversarial networks, feedforward neural networks, variational autoencoders, as well as reinforcement learning systems have been used in design-related generation tasks such as topology optimization or shape synthesis based on visual modalities (e.g., images, voxels, point clouds, etc.) [32]. Other work focuses on learning design strategies. Given a state in solving a truss design problem, Raina et al. [33] predicts what actions humans perform next. Gyory et al. [34,35] analyze real time data of design teams to suggest measures from a predefined list if the communication or action frequency appears to be too low.

Lexical databases [36–38] and stopword lists [39] for technological vocabulary and jargon have been proposed, as have engineering-related ontologies [40–42]. With ontologies come knowledge graphs. However, there has been a lack of specialized engineering knowledge graphs thus far [43]. Only recently, Siddharth et al. [44] build a knowledge graph using patent claims. With a trend towards industry 4.0 and digital twins, the subject of the semantic representation of technological knowledge will probably be increasingly addressed in the future. NLP methods, which have been less applied in the engineering sciences compared to the biomedical and material ones, have become increasingly popular recently. In design research, NLP has been applied to requirements extraction, ontology construction, patent analysis, and more [45].

Using foundation models such as LLMs or pre-trained multi-modal models in the engineering design process is a recent and unexplored topic. Several studies have experimented with using LLMs to provide designers with inspirational stimuli for ideation. In two explorative studies, Zhu and Luo [46,47] prompted GPT-2 and -3 to generate design concepts (text-to-text) based on the description of either a concept, problem, or analogy in both a fine-tuning and few-shot learning setting. Similarly, Zhu et al. [48] fine-tune GPT-3 for bio-inspired design concept generation. Ma et al. [49,50] compare LLM-generated design solutions with crowdsourced ones and find that the human-generated solutions are more novel and diverse. Tian et al. [51] classify and experiment with different prompts for conceptual design. Jadhav and Farimani [52] combine an LLM and FEM to iteratively optimize truss structures. Ataei et al. [53] enhance the requirements elicitation process by simulating users via LLMs. Doris et al. [54] present DesignQA, a benchmark for evaluating multi-modal LLMs on the tasks of extracting rules in requirement documents, referencing elements in CAD renderings and drawings based on requirement documents, as well as checking rule compliance. Evaluation results for GPT-4 highlight limitations of current foundation models in understanding complex technical documents. Other work has focus on design concept

evaluation combining a pre-trained language model (BERT) and image models in a multi-modal one [55,56]. Song et al. [57] provided an extensive overview of multi-modal machine learning for engineering design. They outline possible applications, but focus on lower level tasks such as text-to-shape or shape-to-text synthesis.

Orthogonal to the prior work, we concentrate on the design process itself as a complex, iterative, and dynamic reasoning process and situate recent advances in NLP and machine learning in a superordinate framework.

3. Depicting the design process as a goal-oriented, argumentative discourse

We have digital artifacts for shapes, assembly processes, stress distributions, flow patterns, and more. Until now, however, computer-aided engineering, as practiced in industry, has not included the creative and argumentative process of the product development process itself. In the following, we argue that this process could be digitized and partially automatized next, and outline how this can be achieved.

Many steps in the product development process are performed using computation and are not based on human thought alone. However, humans are needed to integrate these computational processes, be they calculations, simulations, or optimizations, into a meaningful superordinate product development process. Human thought and world knowledge is required to reduce the solution space in advance and come up with original ideas that have not been modeled to be computationally accessible before. For example, when a bicycle is designed, the starting point is not a blank slate but an idea of how a bicycle looks and how it has worked well for over a century. If a standardized aerodynamic tube shape across bicycle manufactures is proposed, it is unlikely that this idea originated from a numerical optimization. Instead, background knowledge and the ability to think and reason are used. Solving engineering problems requires an argumentative discourse. As such, argumentation is inherent to the product development process. Experiments and calculations, etc., inform the discourse to provide necessary information. Nevertheless, argumentation is rarely given a lot of attention, perhaps because it is hidden, as it is typically not made explicit in a readable or visual representation.

Having described that a goal-driven, argumentative discourse is at the core of the design process, we argue that it should be represented as a digital artifact. Humans communicate, argue, and reason using natural language. Hence, the argumentative discourse can be largely represented in textual form. Many parts of the design process, however, cannot be represented as text. Therefore, we distinguish the design discourse from external actions (such as performing an experiment or simulation) and other engineering artifacts (such as a drawing, or 3D model). We formulate that external actions are invoked from within the design discourse and in turn inform the discourse, either directly or indirectly, by yielding other engineering artifacts that inform the design discourse (see Fig. 1).

Representing the argumentative discourse as a digital artifact would improve the documentation of the design process. Instead of only archiving the results of process steps (e.g., CAD files or the results of simulation runs), the reasoning process is documented and hence archivable. For a past development process to be efficiently used for the development of a new product generation, past decisions and alternatives must be accessible. Having the reasoning process explicitly documented makes past design decisions traceable. Furthermore, making the reasoning process explicit could improve collective reasoning and therefore collaborative design. Finally, it would allow for machines to participate in the reasoning process, which the next section covers.

4. Foundation models as interlocutors in the design discourse

It is apparent that LLMs and other foundation models will (and already do) reshape how we code, write, and search for information.

Prompt: We're looking to increase the energy efficiency of our next-generation wind turbine blades. Attached are the current design specifications, material constraints, and performance data. Please propose modifications to optimize blade efficiency.

Thought: What factors influence blade efficiency?

Thought: Blade shape, material choice, and surface texture are key.

Action: Confirm: Are we prioritizing efficiency over cost?

Prompt: Yes, but keep costs within reasonable limits.

Thought: What are the constraints?

Action: Review material constraints and design specifications.

Thought: The blades must withstand high winds and variable weather conditions.

Action: Research on advanced composites suitable for harsh weather.

⋮

Thought: How can blade shape be optimized for increased efficiency?

Action: Study biomimicry examples for aerodynamic optimization.

Thought: Incorporating serrations on trailing edges can reduce noise and improve lift.

Action: Model different serration patterns.

⋮

Thought: Can surface texture improve performance?

Thought: Sharks' skin denticles reduce drag in water; similar textures might reduce air resistance.

Action: Run simulations on blade sections with micro-textured surfaces.

Thought: Evaluate manufacturing feasibility of micro-textured surfaces.

Action: Research manufacturing techniques for detailed surface textures.

⋮

Prompt: Can you visualize the modified blade design with these optimizations?

⋮

■ = Human ■ = Machine

Fig. 2. An example of a design discourse on wind turbine blade optimization. The example was generated by prompting GPT-4 to translate an example of a design discourse on the conceptual design of bicycles (see our original preprint [58]) to that of wind turbine blades and has not been post-processed. The example is intended to clarify the ideas presented here using a simple method and is not designed to provide practical value to actual wind turbine engineers.

Beyond that, we believe that they can be directly involved in the creative and argumentative design process, and so complement generative design and topology optimization tools that are already able to automatize large parts of design exploration and optimization.

An example of a design discourse on wind turbine blade optimization involving humans and machines is depicted in Fig. 2. Besides illustrating a design discourse, the example highlights the ability of today's LLMs to provide creative stimuli. The discourse was generated using an LLM and suggests the incorporation of serrations on the trailing edges of wind turbine blades to reduce noise and improve lift, as well as textured blade surfaces to further increase aerodynamic efficiency.

In the previous section, we argued that explicitly representing the argumentative discourse of the design process as a digital artifact results in many advantages. One of these is the potential for machines to participate in the reasoning process. LLMs and related multi-modal models have several characteristics that suggest that they can be successfully applied for this purpose.

LLMs and related multi-modal models can input and output natural language. This facilitates their integration into the argumentative discourse, which to a large extent is conducted in natural language anyway, be it in intra- or interpersonal communication. World knowledge is important both to interact with humans, where a certain common ground is required, and for reducing the solution space and coming up with creative solutions. Preliminary evidence suggests that LLMs possess rich representations of the world despite being trained on simple objectives [59]. It is evident that humans are constrained by their knowledge when designing by analogy or biomimicry. In contrast, LLMs can accumulate wide-ranging knowledge during training and provide valuable creative stimuli, for example, by generating cross-domain analogies [60]. Temperature is a parameter used in LLM text generation to control the probability distribution of tokens (thus affecting the diversity of outputs), which is often associated with creativity. However, the correlation is less clear in practice [50,61,62].

LLMs of sufficient size display strong results on various tasks involving reasoning and are able to perform step-by-step reasoning [24].

Multi-modal models can operate on various forms of design representations, which is important, because throughout the product development process, designers utilize different types of representations of their designs (i.e., text, tables, sketches, and 3D models) [57]. Finally, many tasks in the engineering design process cannot be solved by means of pure thought but require specialized engineering software and databases (i.e., CAD and simulation software, patent and material databases). Recent work shows that LLMs can be trained to interact with APIs (including the decision of when to call which API with which arguments) in a self-supervised setting [28].

Despite the promising characteristics and capabilities of LLMs, there are various risks associated with human-AI collaboration, due to shortcomings of today's LLMs and human biases in the use of AI. LLMs misinterpret user intents, inherent human biases from their training data, "hallucinate" facts, and are prone to wrong reasoning. To mitigate resulting risks, humans should follow, understand, and question the provided reasoning. Yet, humans tend to over-rely on algorithmic advice [63,64].

The role of humans in the design process is unlikely to change fundamentally. After all, they hold the design goal, so they must guide the design process, provide requirements, set priorities, and understand and evaluate the information provided by the machine. While in the early years of LLMs careful prompt selection was crucial, this became less so with approaches such as reinforcement learning from human feedback. However, LLM-generated design solutions are still highly sensitive to the prompt [50]. We suspect that this sensitivity can be reduced by fine-tuning LLMs on task-specific data, integrating them into purpose-built frameworks, and clarification requests.

We conclude that these models are fundamentally applicable to the purpose of assisting in the engineering design discourse. Nevertheless, a single call to a model will not be of great utility; instead, the models must be embedded within a framework to solve complex engineering tasks. In the next section, we outline aspects of such a framework and simultaneously highlight promising future research directions.

5. Recommendations for future research

Many puzzle pieces for the outlined transformation of the engineering design process are at hand. However, several aspects, which have not been sufficiently researched, are likely to be necessary for a successful implementation of the proposed concept, namely:

Formalizing the engineering design discourse within a framework of stages and components. Although approaches to interpret how outputs are formed in deep neural network models (such as probing [65]) exist, in practice, LLMs largely constitute black boxes. To increase the interpretability and accuracy of LLMs, approaches such as scratch-pad [66] or chain-of-thought prompting [24] steer the models towards the generation of intermediate steps. Complementary to the aforementioned approaches of generating intermediate steps, models can be guided to imitate certain patterns of thinking or to follow a logical flow by embedding calls to LLMs into a framework with a predefined causal structure. In such frameworks "querying a language model becomes a computational primitive" [67]. For example, answering questions in steps adhering to formal logic yields interpretable reasoning traces and reduces the "hallucination" of facts [67]. Similarly, formalizing the engineering design discourse within a framework of stages and recurring components will help safeguard the models and increases the detail at which reasoning processes can be documented and verified. Furthermore, it contributes to the clarity of the discourse and makes LLMs more controllable, in that humans have more points of intervention, allowing them to provide more fine-grained guidance. For defining such a framework, the extensive literature on design research can assist (e.g., [68,69]).

Creating machine-actionable interfaces for engineering software. We noted that using external tools via APIs with LLMs is already an active area of research. The tools invoked in the design discourse could include specialized engineering software such as topology optimization. However, for today's LLMs, the tools are required to provide a textual interface; that is, both inputs and outputs are represented as text [28]. Therefore, the interfaces of specialized engineering software should be adapted according to the models' requirements. Note that the requirements for an interface are bound to change with multi-modal models.

Learning representations better corresponding to skills required for engineering. Good spatial imagination, solid engineering knowledge, and high heuristic competence are skills that correlate with higher quality engineering design solutions [68]. For foundation models to fulfill a supportive role in the engineering design process, the learned representations should in turn correlate to a high degree with these skills. Therefore, the proportion of technical documents (such as patents or specialist books) in the pre-training corpus can be increased. However, although LLMs compress large amounts of knowledge in their weights, a representation corresponding to good spatial imagination and causal understanding of physical processes is unlikely to be learned from text alone. With a trend towards multi-modal models, incorporating images, videos, and other modalities, we see this problem as likely to be addressed in the near future. Additionally, strengthening the reasoning and creative capabilities of LLMs is a promising and growing research direction.

Specifying evaluation metrics and datasets. Prior work has expressed the need for design-specific metrics for evaluating deep generative models in design synthesis [46,49,57,70]. Similarly, design-specific metrics are likely to be required to evaluate the performance of frameworks to assist in the argumentative design discourse. In cases where ground truth is not required, more complex assessments, for which there is not yet a calculable metric, but for which human evaluators are required, such as the assessment of usefulness and feasibility [49], could possibly be approximated by foundation models in the future. Furthermore, as participation in the argumentative discourse of engineering design is a new task, there is a need for public datasets to evaluate and compare models against.

6. Conclusion

We propose conceiving the design process as a goal-oriented, argumentative discourse on which foundation models can operate. Making the reasoning steps explicit in a new digital artifact of the product development process could lead to improved documentation and increased collaboration, and makes certain forms of machine assistance possible in the first place. We describe how LLMs and multi-modal foundation models can assist in the design discourse and outline interesting directions of research, including structuring of the design discourse in stages and components, the provision of machine-actionable interfaces for specialized engineering software, the development of foundation models with learned representations that better correspond to the skills required in the design process, and the creation and publication of common metrics and datasets as a community effort. We do not intend to describe a theoretical framework that is far from ever being implemented in practice, but believe that the product development process is about to change, with the aspects described in this article being part of its future.

CRedit authorship contribution statement

Jan Göpfert: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Jann M. Weinand:** Writing – review & editing, Visualization, Supervision, Project administration. **Patrick Kuckertz:** Writing – review & editing, Supervision. **Detlef Stolten:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors would like to thank the German Federal Government, the German state governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) – project number: 442146713. Furthermore, this work was supported by the Helmholtz Association, Germany under the program “Energy System Design”.

References

- [1] Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, et al. SuperGLUE: A stickier benchmark for general-purpose language understanding systems. In: *Advances in neural information processing systems*. 32, Vancouver, BC, Canada: Curran Associates, Inc.; 2019, p. 3261–75.
- [2] Gan Z, Li L, Li C, Wang L, Liu Z, Gao J. Vision-language pre-training: Basics, recent advances, and future trends. *Found Trends Comput Graph Vis* 2022;14(3–4):163–352. <http://dx.doi.org/10.1561/0600000105>, Publisher: Now Publishers, Inc.
- [3] Driess D, Xia F, Sajjadi MSM, Lynch C, Chowdhery A, Ichter B, et al. PaLM-E: An embodied multimodal language model. 2023, [arXiv:2303.03378](https://arxiv.org/abs/2303.03378).
- [4] Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. *Nature* 2023;616(7956):259–65. <http://dx.doi.org/10.1038/s41586-023-05881-4>.
- [5] M. Hocky G, D. White A. Natural language processing models that automate programming will transform chemistry research and teaching. *Digital Discov* 2022;1(2):79–83. <http://dx.doi.org/10.1039/D1DD00009H>, Publisher: Royal Society of Chemistry.
- [6] Manning CD. Human language understanding & reasoning. *Daedalus* 2022;151(2):127–38. http://dx.doi.org/10.1162/daed_a_01905.
- [7] Harris ZS. Distributional structure. *WORD* 1954;10(2–3):146–62. <http://dx.doi.org/10.1080/00437956.1954.11659520>.
- [8] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. 2013, [arXiv:1301.3781](https://arxiv.org/abs/1301.3781).
- [9] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013, [arXiv:1310.4546](https://arxiv.org/abs/1310.4546).
- [10] Pennington J, Socher R, Manning C. GloVe: Global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing*. Doha, Qatar: Association for Computational Linguistics; 2014, p. 1532–43. <http://dx.doi.org/10.3115/v1/D14-1162>.
- [11] Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. 2018, [arXiv:1802.05365](https://arxiv.org/abs/1802.05365).
- [12] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. Minneapolis, Minnesota: Association for Computational Linguistics; 2019, p. 4171–86. <http://dx.doi.org/10.18653/v1/N19-1423>.
- [13] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in neural information processing systems*, vol. 30, Long Beach, CA, USA: Curran Associates, Inc.; 2017, p. 5998–6008.
- [14] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. 2019, [arXiv:1907.11692](https://arxiv.org/abs/1907.11692).
- [15] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018.
- [16] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. 2019.
- [17] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. In: *Advances in neural information processing systems*, vol. 33, virtual: Curran Associates, Inc.; 2020, p. 1877–901.
- [18] Workshop B, Scao TL, Fan A, Akiki C, Pavlick E, Ilić S, et al. BLOOM: A 176b-parameter open-access multilingual language model. 2023, [arXiv:2211.05100](https://arxiv.org/abs/2211.05100).
- [19] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, et al. LLaMA: Open and efficient foundation language models. 2023, [arXiv:2302.13971](https://arxiv.org/abs/2302.13971).
- [20] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res* 2020;21(140):1–67.
- [21] Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. 2022, [arXiv:2210.11416](https://arxiv.org/abs/2210.11416).
- [22] Tay Y, Dehghani M, Tran VQ, Garcia X, Wei J, Wang X, et al. UL2: Unifying language learning paradigms. 2023, [arXiv:2205.05131](https://arxiv.org/abs/2205.05131).
- [23] Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. 2022, [arXiv:2206.07682](https://arxiv.org/abs/2206.07682).
- [24] Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. 2023, [arXiv:2201.11903](https://arxiv.org/abs/2201.11903).
- [25] Schaeffer R, Miranda B, Koyejo S. Are emergent abilities of large language models a mirage?. 2023, [arXiv:2304.15004](https://arxiv.org/abs/2304.15004).
- [26] Wei J, Bosma M, Zhao VY, Guu K, Yu AW, Lester B, et al. Finetuned language models are zero-shot learners. 2022, [arXiv:2109.01652](https://arxiv.org/abs/2109.01652).
- [27] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. *Adv Neural Inf Process Syst* 2022;35:27730–44.
- [28] Schick T, Dwivedi-Yu J, Dessi R, Raileanu R, Lomeli M, Zettlemoyer L, et al. Toolformer: Language models can teach themselves to use tools. 2023, [arXiv:2302.04761](https://arxiv.org/abs/2302.04761).
- [29] Yao S, Zhao J, Yu D, Du N, Shafraan I, Narasimhan K, et al. React: Synergizing reasoning and acting in language models. 2023, [arXiv:2210.03629](https://arxiv.org/abs/2210.03629).
- [30] Shinn N, Labash B, Gopinath A. Reflexion: an autonomous agent with dynamic memory and self-reflection. 2023, [arXiv:2303.11366](https://arxiv.org/abs/2303.11366).
- [31] Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. 2022, [arXiv:2108.07258](https://arxiv.org/abs/2108.07258).
- [32] Regenwetter L, Nobari AH, Ahmed F. Deep generative models in engineering design: A review. *J Mech Des* 2022;144(071704). <http://dx.doi.org/10.1115/1.4053859>.
- [33] Raina A, Cagan J, McComb C. Design strategy network: A deep hierarchical framework to represent generative design strategies in complex action spaces. *J Mech Des* 2021;144(021404). <http://dx.doi.org/10.1115/1.4052566>.
- [34] Gyory JT, Soria Zurita NF, Martin J, Balon C, McComb C, Kotovsky K, et al. Human versus artificial intelligence: A data-driven approach to real-time process management during complex engineering design. *J Mech Des* 2021;144(2). <http://dx.doi.org/10.1115/1.4052488>.
- [35] Gyory JT, Kotovsky K, McComb C, Cagan J. Comparing the impacts on team behaviors between artificial intelligence and human process management in interdisciplinary design teams. *J Mech Des* 2022;144(10). <http://dx.doi.org/10.1115/1.4054723>.
- [36] Sarica S, Luo J, Wood KL. TechNet: Technology semantic network based on patent data. *Expert Syst Appl* 2020;142:112995. <http://dx.doi.org/10.1016/j.eswa.2019.112995>.
- [37] Jang H, Jeong Y, Yoon B. TechWord: Development of a technology lexical database for structuring textual technology information based on natural language processing. *Expert Syst Appl* 2021;164:114042. <http://dx.doi.org/10.1016/j.eswa.2020.114042>.
- [38] Shi F, Chen L, Han J, Childs P. A data-driven text mining and semantic network analysis for design information retrieval. *J Mech Des* 2017;139(11). <http://dx.doi.org/10.1115/1.4037649>.
- [39] Sarica S, Luo J. Stopwords in technical language processing. In: Amancio DR, editor. *Plos One* 2021;16(8):e0254937. <http://dx.doi.org/10.1371/journal.pone.0254937>.
- [40] Morbach J, Wiesner A, Marquardt W. OntoCAPE—A (re)usable ontology for computer-aided process engineering. *Comput Chem Eng* 2009;33(10):1546–56. <http://dx.doi.org/10.1016/j.compchemeng.2009.01.019>.
- [41] Booshehri M, Emele L, Flügel S, Förster H, Frey J, Frey U, et al. Introducing the open energy ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy AI* 2021;5:100074. <http://dx.doi.org/10.1016/j.egyai.2021.100074>.
- [42] Sanfilippo EM, Kitamura Y, Young RIM. Formal ontologies in manufacturing. *Appl Ontol* 2019;14(2):119–25. <http://dx.doi.org/10.3233/AO-190209>.
- [43] Han J, Sarica S, Shi F, Luo J. Semantic networks for engineering design: A survey. *Proc Des Soc* 2021;1:2621–30. <http://dx.doi.org/10.1017/pds.2021.523>, Publisher: Cambridge University Press.
- [44] Siddharth L, Blessing LTM, Wood KL, Luo J. Engineering knowledge graph from patent database. *J Comput Inf Sci Eng* 2021;22(2). <http://dx.doi.org/10.1115/1.4052293>.
- [45] Siddharth L, Blessing L, Luo J. Natural language processing in-and-for design research. *Des Sci* 2022;8:e21. <http://dx.doi.org/10.1017/dsj.2022.16>.
- [46] Zhu Q, Luo J. Generative pre-trained transformer for design concept generation: An exploration. *Proc Des Soc* 2022;2:1825–34. <http://dx.doi.org/10.1017/pds.2022.185>, Publisher: Cambridge University Press.
- [47] Zhu Q, Luo J. Generative transformers for design concept generation. *J Comput Inf Sci Eng* 2023;23(4). <http://dx.doi.org/10.1115/1.4056220>.
- [48] Zhu Q, Zhang X, Luo J. Biologically inspired design concept generation using generative pre-trained transformers. *J Mech Des* 2023;145(041409). <http://dx.doi.org/10.1115/1.4056598>.
- [49] Ma K, Grandi D, McComb C, Goucher-Lambert K. Conceptual design generation using large language models. 2023, [arXiv:2306.01779](https://arxiv.org/abs/2306.01779).

- [50] Ma K, Grandi D, McComb C, Goucher-Lambert K. Exploring the capabilities of large language models for generating diverse design solutions. 2024, [arXiv:2405.02345](#).
- [51] Tian Y, Liu A, Dai Y, Nagato K, Nakao M. Systematic synthesis of design prompts for large language models in conceptual design. CIRP Ann 2024. <http://dx.doi.org/10.1016/j.cirp.2024.04.062>.
- [52] Jadhav Y, Farimani AB. Large language model agent as a mechanical designer. 2024, [arXiv:2404.17525](#).
- [53] Ataei M, Cheong H, Grandi D, Wang Y, Morris N, Tessier A. Elicitron: An LLM agent-based simulation framework for design requirements elicitation. 2024, [arXiv:2404.16045](#).
- [54] Doris AC, Grandi D, Tomich R, Alam MF, Cheong H, Ahmed F. DesignQA: A multimodal benchmark for evaluating large language models' understanding of engineering documentation. 2024, [arXiv:2404.07917](#).
- [55] Yuan C, Marion T, Moghaddam M. Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model. J Mech Des 2021;144(2). <http://dx.doi.org/10.1115/1.4052366>.
- [56] Song B, Miller S, Ahmed F. Attention-enhanced multimodal learning for conceptual design evaluations. J Mech Des 2023;145(4). <http://dx.doi.org/10.1115/1.4056669>.
- [57] Song B, Zhou R, Ahmed F. Multi-modal machine learning in engineering design: A review and future directions. 2023, [arXiv:2302.10909](#).
- [58] Göpfert J, Weinand JM, Kuckertz P, Stolten D. Opportunities for large language models and discourse in engineering design. 2023, [arXiv:2306.09169](#).
- [59] Li K, Hopkins AK, Bau D, Viégas F, Pfister H, Wattenberg M. Emergent world representations: Exploring a sequence model trained on a synthetic task. 2023, [arXiv:2210.13382](#).
- [60] Ding Z, Srinivasan A, Macneil S, Chan J. Fluid transformers and creative analogies: Exploring large language models' capacity for augmenting cross-domain analogical creativity. In: C&C '23, New York, NY, USA: Association for Computing Machinery; 2023, p. 489–505. <http://dx.doi.org/10.1145/3591196.3593516>.
- [61] Peeperkorn M, Kouwenhoven T, Brown D, Jordanous A. Is temperature the creativity parameter of large language models?. 2024, [arXiv:2405.00492](#).
- [62] Bhavya B, Xiong J, Zhai C. Analogy generation by prompting large language models: A case study of instructGPT. In: Shaikh S, Ferreira T, Stent A, editors. Proceedings of the 15th international conference on natural language generation. Waterville, Maine, USA and virtual meeting: Association for Computational Linguistics; 2022, p. 298–312. <http://dx.doi.org/10.18653/v1/2022.inlg-main.25>.
- [63] Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. Organ Behav Hum Decis Process 2019;151:90–103. <http://dx.doi.org/10.1016/j.obhdp.2018.12.005>.
- [64] Passi S, Vorvoreanu M. Overreliance on AI literature review. Technical report, 2022.
- [65] Belinkov Y. Probing classifiers: Promises, shortcomings, and advances. Comput Linguist 2022;48(1):207–19. http://dx.doi.org/10.1162/coli_a_00422.
- [66] Nye M, Andreassen AJ, Gur-Ari G, Michalewski H, Austin J, Bieber D, et al. Show your work: Scratchpads for intermediate computation with language models. 2021, [arXiv:2112.00114](#).
- [67] Creswell A, Shanahan M. Faithful reasoning using large language models. 2022, [arXiv:2208.14271](#).
- [68] Fricke G. Successful individual approaches in engineering design. Res Eng Des 1996;8(3):151–65. <http://dx.doi.org/10.1007/BF01608350>.
- [69] Pahl G, Beitz W, Feldhusen J, Grote K-H. Engineering design. London: Springer; 2007, <http://dx.doi.org/10.1007/978-1-84628-319-2>.
- [70] Regenwetter L, Srivastava A, Gutfreund D, Ahmed F. Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. 2023, [arXiv:2302.02913](#).