Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 136 (2025) 354–359

www.elsevier.com/locate/procedia

35th CIRP Design 2025

Training a Foundation Model in Engineering Design Understanding

Haluk Akay[a*], Antonio J. Capezza[b], Billy W. Hoogendoorn[b], Maryna Henrysson[c]

[a] *Faculty of Mechanical Engineering, Delft University of Technology, Delft, Netherlands*
[b] *Department of Fiber and Polymer Technology, KTH Royal Institute of Technology, Stockholm, Sweden*
[c] *Department of Energy Technology, KTH Royal Institute of Technology, Stockholm, Sweden*

* Corresponding author. *E-mail address:* h.j.akay@tudelft.nl

Abstract

Across industry, applications involving Artificial Intelligence are shifting from task-specific to general purpose foundation models able to perform a diverse set of previously unseen functions with minimal instruction or additional training. To develop such a foundation model for engineering design, training must be completed at a meaningful scale on artifacts of prior product development, which can be multimodal and sparsely annotated. This work presents a sequence learning framework for training a foundation model on contextual relationships between function, form, and fabrication in engineering design. This learning method is demonstrated with a case study in absorbent product design.

Keywords: Foundation Model, Design Representation, Artificial Intelligence

1. Introduction

Data-driven methods collectively referred to as Artificial Intelligence (AI) continue to transform a wide range of industrial applications due to novel Machine Learning architectures for recognizing patterns in complex data, widespread digitalization of data, availability of powerful computational hardware, and user-friendly application programming interfaces (APIs) to support implementation.

In engineering design, as in many fields, the initial tasks to be integrated with AI were focused problems with well-defined system boundaries. These include AI-based predictive tasks such as forecasting electric loads on a factory floor by learning from machine tool, building, and weather features [1] or predicting differences between 3D design models and fabricated components to establish benchmarks for tolerancing [2]. AI-based image recognition has been applied to quality control, such as classification of lithium-ion batteries to identify defects [3] and identification of process deviations in additively manufactured parts [4]. AI-based statistical language modeling has also been applied to textually described information retrieval in the context of engineering design [5].

The introduction of the Transformer neural network architecture [6] by Google in 2017 enabled a shift in general language understanding resulting in the development of Turing test-passing [7] generative conversational models such as ChatGPT. Since then, demand for integrated computational models capable of conversation, reasoning, and universal knowledge retrieval has grown. These criteria require models to not just map inputs to outputs for a well-defined task, but to acquire a deep understanding of the structure and semantics of a field in order to perform any requested function with minimal instruction or domain-specific training. Emulating the function of internet search engine, the information technology sector has developed so-called *foundation models* such as Google's Gemini [8], OpenAI's GPT [9], and Meta's Llama [10] meant

10.1016/j.procir.2025.08.062

for application to a diverse range of use cases within general information retrieval.

Despite reporting high scores on general language understanding benchmarks, such foundation models exhibit limitations when performing quantitative tasks. Given the bulk of their training is web-mined corpora from the internet, these foundation models tend to demonstrate better performance when having to estimate likelihood of words in a sequence compared to computing simple arithmetic. Foundation models are trained to behave as an average human internet user rather than a source of truth [11].

In this work, we examine how to train a foundation model in the general understanding of engineering design. Like existing foundation models, one for design would need to be conversant and able to perform a wide range of tasks. However, in contrast to those models, it would require technical rigor which comes from a deep understanding of the meaning and structure of requirements, solutions, and fabrication processes which constitute the engineering design process.

We approach this task from the perspective of modeling the design process itself as a form of language; a sequence of functionality, geometric form, and production. We develop a framework for training such a model, constructing the necessary dataset for this training, and demonstrate it through a case study on the design of absorbent products.

2. Related Work

Here, prior work towards the development of a foundation model for engineering design is reviewed, considering both theoretical frameworks and efforts to implement a working computational tool.

2.1. Design representation methods

Prior to the advent of widespread digitalization, practitioners of engineering design worked to develop a universal framework for guiding better and repeatable engineering design. Even without a computational implementation, these illuminate the structure and semantics of design into core axioms and are a useful basis for modeling the language of engineering design.

The Theory of Inventive Problem-Solving (TRIZ) [12] distills engineering design into a set of 40 principles for invention. As an example, one principle is “universality” where one physical component can serve multiple functions in order to reduce the number of total parts. This set of principles strives to provide a structured repository of elemental design solutions, but attempting to address every design problem becomes challenging as societal needs evolve over time.

Axiomatic Design [13] condenses engineering design into just two principles relating to (1) maintaining functional independence by avoiding the same requirement being coupled to multiple solutions, and (2) minimizing complexity as measured by density of information content. Axiomatic Design also provides a framework for hierarchically representing functional (what), physical (how), and process domains as a tree-structure which is easily adapted for computational purposes.

Principles of product design and development [14] represents engineering design as a sequential process. By representing the process as an iterating sequence, design is extended to the early stage of identifying societal needs as well as to the late stage of the product lifecycle to usage and re-use or end-of-life.

These frameworks seek to define structure and semantics for engineering problem-solving, which is a critical step in developing a general foundation model of design. However, for a field as complex as engineering design, it is difficult to construct a rule-based system applicable to any unseen case. For this reason, we explore how to train a learning-based system on such design understanding.

2.2. Training computational models on design understanding

With the availability of foundation AI models for general usage, research has been conducted integrating engineering design knowledge into the intelligence of the AI model. For a case study in vehicle maintenance, a language model was fine-tuned on a dataset of aircraft maintenance record logbooks to integrate a hierarchically structured aircraft design ontology with an AI [15]. The fine-tuned model was able to recommend semantically similar maintenance actions to the ground truth labels.

In the domain of additive manufacturing, a knowledge base of best practices accessible by a user-friendly interface was connected to Computer Aided Design (CAD) software to embed both automated geometric modification and feedback for quality control into the design process [16]. Despite reporting long runtimes interrupting the design workflow, this concept demonstrated how a structured repository of established principles could be utilized in real time to guide design decision-making.

In the space of representing design knowledge for computational understanding, our previous work has sought to discover latent problem-solving structure by extracting functional requirements from documentation [17]. We applied language model-based question answering to semantically identify valuable design information buried in unstructured texts without relying on keyword matching. We also applied language model-based chat completion to extract undocumented design information conversationally from human experts [18]. Both these methods are integrated into the dataset construction system in this presented work and are applied to construct the dataset used in the case study.

3. Methodology

In this section we present a framework for training a foundation model in general engineering design understanding. First, the assumptions for modeling design as a sequence are introduced. Next, the training routine is presented. Finally, an integrated method for collecting training examples from multimodal and undocumented sources of design information is overviewed.

3.1. Sequence modeling of engineering design

The objective is to develop a training routine that will allow a foundation model to learn the semantics and structure of engineering design. Thus far, existing foundation models are trained in general understanding of the semantics of natural language. For this purpose, language is modeled as a sequence, with the meaning of each word highly influenced by context being the words preceding and following. This is the basis of the cloze test in language understanding [19]. Given a word sequence, a language model attempts to estimate the probability of any given word w_t occupying a position t in the sequence given neighboring context. For such a context window of size 1, this probability can be expressed as:

$$p(w_t | w_{t-1}, w_{t+1})$$

The language model gains the ability to accurately compute such a probability distribution over all words in a vocabulary by learning a multidimensional feature space within which words occupy a position represented by a vector. When a machine learning model is trained, the loss used for parameter updates corresponds to a measure of vector similarity between the actual word in the training example and the maximum likelihood prediction given by the model. This system of working through an existing sequence and attempting to predict words based on context is known as *masked language modeling*. This is an efficient training routine because otherwise unlabeled language data with no additional annotations can serve as a labeled dataset for supervised machine learning simply by masking an item in the sequence and then revealing the unmasked identity as the ground truth to guide a parameter update during learning.

Fig. 1. Design represented as a sequence of function, form, and fabrication

For the domain of engineering design, this masked sequence modeling routine is relevant for two key reasons. First, design data is rarely labeled consistently, so such a sequence modeling framework can transform the learning process into a supervised task with no further manual intervention required, which is a significant challenge to constructing datasets at scale. Second, design, like language, can also be modeled as a sequence, as overviewed in Section 2.1. At a high level, this design sequence consists of identifying a societal need, followed by the mapping to a functional requirement, followed by embodiment through a physical design parameter, finally produced through a process variable. For the purposes of this work, we can simplify this sequence to a triplet representing function, form, and fabrication, or functional requirements (FRs), design parameters (DPs), and process variables (PVs), to borrow terminology from Axiomatic Design theory. To adapt masked sequence modeling to design, we can mask a random portion of this sequence as illustrated in Figure 1. Next, the core learning task of the model can be formalized as estimating the probability of the identity of the masked design element, given the context, expressed as:

$$p(DP_t | FR_{t-1}, PV_{t+1})$$

By training on a dataset of sequences of these design elements, a model can learn from context the semantic meaning of function, form, and fabrication.

3.2. Training Framework

The following framework is presented for training a foundation model on general engineering design understanding through masked sequence modeling. In order to facilitate compatibility with existing design practices and minimize introduction of novel terminology, the overarching representation scheme is based on theory from established Axiomatic Design principles [13]. The objective of this training is to learn from prior sequences of functions, physical solutions, and fabrication processes the meaning of each of such design elements as they exist within these design sequences.

This input for this routine is the hierarchical tree representation described in Axiomatic Design of functional requirements, design parameters, and process variables. The first step involves restructuring the training data into a sequence by flattening each tree. Next, the three resulting sequences are concatenated to form a longer sequence upon which the masked prediction task will be performed. At random, 10 – 20% of the design elements in the sequence have their identities masked; this ratio being based on typical procedures for masked language modeling [20].

The baseline model and parameters utilized for training can be either an existing pre-trained foundation model or an un-trained model with an architecture appropriate for sequence modeling such as a recurrent neural network (RNN) or a transformer. This model is then implemented to predict each masked design sequence element by taking the non-masked elements as inputs, and outputs a prediction. After the prediction is made, the masked elements are unmasked, and the predictions are compared to the actual labels, and a measurement of accuracy (or loss) is computed and used as feedback to update the model weights. This training routine is illustrated in Figure 2 and described in a pseudo-code Table 1.

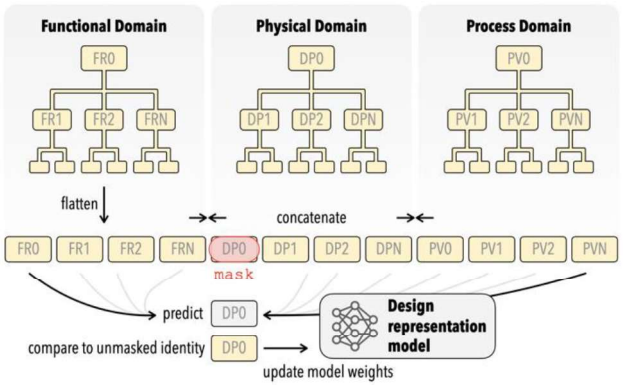

Fig. 2. Training framework integrating masked sequence modeling with hierarchical representations of the design process

Table 1. Masked sequence modeling for learning semantics of design

Data preparation	1. Extract structured tree representations of functional (FR), physical (DP), and process (PV) domains from design documentation
	2. Flatten FR, DP, PV trees into three lists
	3. Concatenate lists into a single sequence
Sequence learning	4. Mask 10% - 20% of items in the sequence
	5. Take unmasked items as input context window
	6. With sequence modeling architecture, compute probability distribution for masked identity
	7. Measure similarity score between maximum likelihood identity and ground truth identity to compute loss
	8. Update model weights
	9. Repeat training process with another sequence

After training is completed, the model should have learned a collection of weights for representing the total vocabulary of design elements as embedded feature vectors.

3.3. Extracting multimodal design elements

A dataset must be constructed for the design repository used in the training framework. Here we describe a system to aggregate such elements in structured form from multimodal artifacts including textual, graphical, and undocumented knowledge.

The first mode of design data artifacts considered is documentation of prior designed products and systems. To extract a structured representation of functional information as needed to input to the training framework in Section 3.2, recursive question answering [17] is the proposed method. In this routine, a language model is given a document as context and prompted to initialize a functional decomposition by identifying the highest-level functional requirement (FR_0). Next, this obtained answer is used to inform a more detailed question mapping FR_0 to the physical domain to identify *how* it is addressed by the highest-level design parameter DP_0. Subsequently, both these elements are used to inform a more detailed prompt extracting the requirements needed for the solution DP_0 to address the problem FR_0. This results in a decomposition of the highest-level *what-how* pair into a set of sub-requirements. The process recurses until termination conditions are met.

The next mode of design data considered is undocumented forms of knowledge which may only exist in the memory of expert engineers. For this mode, the interactive chat completion module of a Push-Pull Digital Thread for manufacturing systems [18] is the proposed method. In this method, a similar extractive functional decomposition process is performed but now in a conversational format to flexibly pursue domains of design and process knowledge which connect to knowledge of the interviewee.

The final mode of data considered encompasses design processes being presently conducted. To avoid requiring additional data acquisition systems, a computational solution to transform a passive real-time documentation system, such as a closed-circuit video record, into structured functional information is used. Unedited video footage of a fabrication process is segmented into frames. In regular intervals, an image frame is passed to a multimodal language model accompanied by a prompt asking the model to infer, from the image, a process variable relating to a functional requirement previously extracted from documentation or expert knowledge sources. The algorithm is a nested loop, performing this task for each image frame and for each functional requirement.

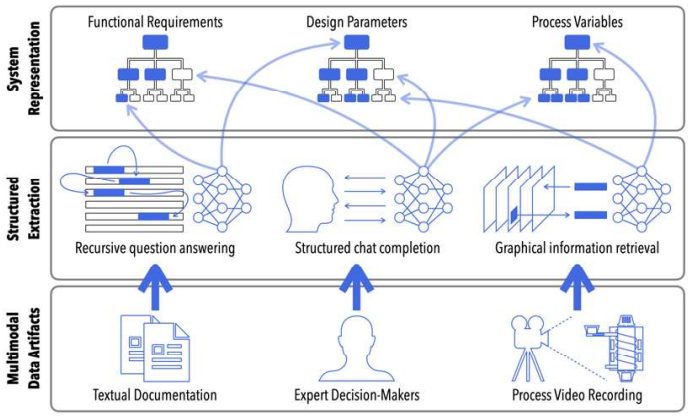

Fig. 3. System for extracting design information from multimodal artifacts

If the aggregated results of the three modes of design data extraction result in sufficient size of data points, K-means clustering methods can be implemented to organize the hierarchy according to functional similarity. The integration of the structured extractions is illustrated in in Figure 3.

4. Case Study

This framework for training a foundation model in engineering design understanding is demonstrated with a case study in the design of absorbent products. Bio-based alternatives to single-use petroleum-derived plastics can be manufactured using byproducts of the agro-food industry. In this case, absorbent materials intended for use in sanitary products are produced using gluten powder discarded when wheat is processed into flour [21]. The gluten is processed using a heated extrusion machine to form a porous absorbent material.

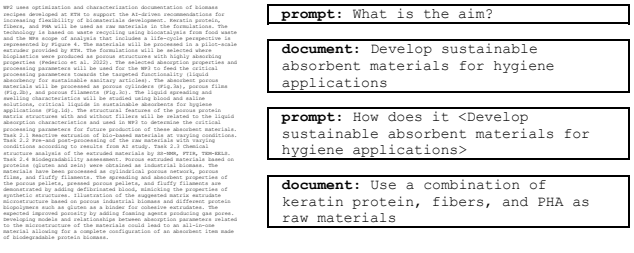

Fig. 4. Question-answering used to extract functional requirements from text

4.1. Dataset construction

To construct the training and validation datasets for the case study, three information sources were considered corresponding to the system for multimodal design element extraction in Section 3.3. The first source is a textual project description. The second is a human subject-matter expert

experienced with the design problem and fabrication process. The third is a sequence of image artifacts from a video of the fabrication process. From each of these sources, functional requirements, design parameters, and process variables were extracted to construct example sequences for learning the semantics of design. The textual information source was processed using recursive question answering and yielded 23 FRs and 1 DP, an example of which shown in Figure 4. The interactive chat completion with the human subject-matter expert, an example of which shown in Figure 5, yielded 2 FRs, 10 DPs, and 4 PVs.

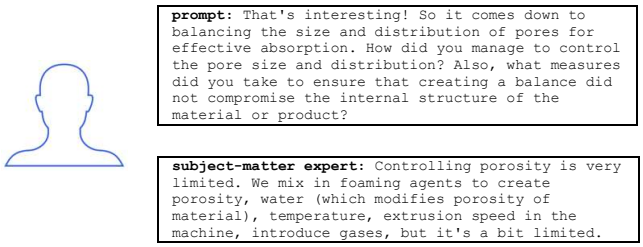

Fig. 5. Generative chat-completion used to interact with subject-matter expert

The fabrication video inference yielded 23 PVs. These artifacts were extracted by taking a video of the extrusion process, and then segmenting the video shot at 30 frames per second into a sequence of images taken as every 300th frame from the video (each image corresponding to a snapshot every 10 seconds). Given the video was of the process domain, the key goal was to extract process variables based on the functional requirements previously extracted from the textual artifacts. For this purpose, a multimodal input was provided to the Gemini-Pro-Vision model first including the image frame and next including a prompt to identify process variables related to a particular functional requirement. The image frames were iterated through once for each high-level functional requirement. An example of the prompt and outputs for frame number 6600 corresponding to midway (3 minutes and 40 seconds) into the process is shown in Figure 6. The resulting constructed dataset includes design elements constituting functional-form-fabrication triplets for the case study, which can be used for training.

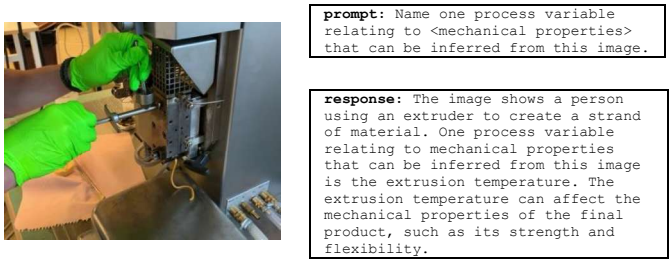

Fig. 6. Process variables extracted from image frames of a fabrication video using a multimodal language-vision model

4.2. Training

The training procedure was performed as outlined in the Table 1 pseudo-code. Google Gemini Application Programming Interface (API) was used to process the training data as JSON Lines object format. The training took place over

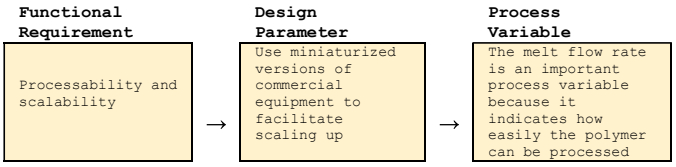

Fig. 7. Example function-form-fabrication triplet from training data

20 epochs. A learning rate of 1.0 was used, representing the rate at which new training data overrides previously seen examples. An adapter size of 4 relating to the number of trainable parameters was used. The masked sequence learning routine involved presenting function – form – fabrication triplets, such as the example shown in Figure 7, with one element randomly masked to the model, and instructing the model to make a prediction as to its masked identity. The loss and fraction of correction predictions measured over the training steps are illustrated in Figure 8.

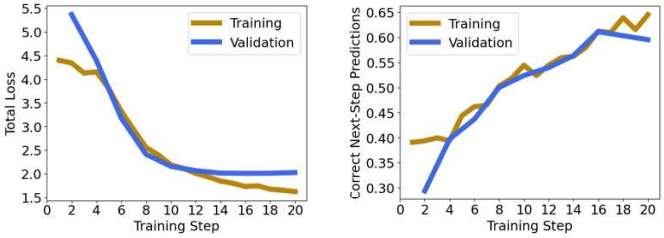

Fig. 8. Model performance through progress of training

5. Results and Discussion

Two metrics can be primarily used to quantitatively evaluate the results. The first is a metric of understanding the structure of the design process. This relates to the ability of the trained model to identify the type of design element missing from the triplet. Each training triplet contained a functional requirement, a design parameter, and a process variable. The ability to know if, given function and form, that fabrication information was missing relates to the model's general understanding of the structure of the design process. The model correctly identified the general type of masked design element at a rate of 100%.

The second metric relates to the semantic design understanding, or comprehension of meaning, of the model. Through statistical natural language processing, textual data is represented as vectors in a high dimensional space where position corresponds to language meaning. We can estimate the functional similarity between the predicted design element $\mathbf{FR_p}$ and the ground truth $\mathbf{FR_t}$ as the cosine distance between these two vectors. This is computed as the dot product divided by vector length, as expressed below.

$$functional\ similarity \approx \frac{\boldsymbol{FR_p} \cdot \boldsymbol{FR_t}}{\|\boldsymbol{FR_p}\| \, \|\boldsymbol{FR_t}\|}$$

On average, over the validation dataset, the average functional similarity of the trained model predictions was 67.2%. The ability of the trained model to consistently identify the category of the masked design domain suggests

effectiveness in learning structure of the design process. Semantic similarity as a metric of functional similarity suggests opportunities to improve performance.

The presented method illustrates a generalized training framework intended for application to engineering design. A limitation of the case study was the small scale at which data was extracted. This was likely a factor in performance and limits the applicability of the trained model beyond this study. However, the training framework is such that through future community-sourced effort to expand the repository, others may share more data to scale available design knowledge.

Another limitation of this work relates to the methods for extracting functional requirements, physical design parameters, and process variables. This framework only includes textual, graphical, and undocumented knowledge. However, design data can exist in other modes such as tabulated and 3D models. Functional information must be extracted from all forms of multimodal data. Also, methods for latent representations of design sequences beyond natural language descriptions must be developed. Representing with text requires a statistical language model to embed their meaning into vector space, allowing an opportunity for information loss in the process.

The proposed framework is based on design representations from Axiomatic Design theory. However, the methods could be adapted to other heuristic-based methods such as TRIZ for integration into existing design practices. Many design theories share an emphasis on language descriptions of function and graphic representations of form, presenting an opportunity for future work to develop compatibility with a wider range of design methodologies through sequence learning for collective societal benefit.

Finally, the ethical and pedagogical aspects of developing a so-called foundation model in the field of engineering design must be discussed. The benefits of training an AI-based model on aggregated knowledge include leveraging prior collective problem-solving information to design for novel engineering challenges. However, drawbacks include subverted usage of such a model as justification for ill-conceived designs as well as a shortcut for decision-making. Prior to widespread deployment of such models in engineering design, it is advisable to develop a code of best practices for utilizing AI to make critical decisions.

6. Conclusion

As AI applications shift from dedicated well-defined tasks to general purpose knowledge sources, the need for such a foundation model in engineering design to leverage prior problem-solving experience for future decision-making. In this work we present a framework for training such a model on the structure and semantics of design. We also present a method for constructing the required datasets for such training. We demonstrate this on a case study of absorbent product design and share initial results on a small scale for the trained model. Future directions of this work will focus on validating the recommendations provided by the model for unseen design tasks in order to establish metrics for reliability, as well as performing training on larger scale repositories of prior examples across varied domains of engineering design.

Acknowledgments: This work was supported by a Digital Futures Postdoc Fellowship (Data-driven design for climate action), the KTH Sustainability Office (Environment and sustainability without boundaries grant, and the Swedish Research Council (BioRESorb project).

References

[1] Walther, J., Spanier, D., Panten, N., Abele, E., 2019. Very short-term load forecasting on factory level – A machine learning approach. Procedia CIRP 80, 705–710.
[2] Zhu, Z., Anwer, N., Huang, Q., & Mathieu, L. 2018. Machine learning in tolerancing for additive manufacturing. CIRP annals, 67(1), 157-160.
[3] Huber, J., Tammer, C., Krotil, S., Waidmann, S., Hao, X., Seidel, C., Reinhart, G., 2016. Method for Classification of Battery Separator Defects Using Optical Inspection. Procedia CIRP 57, 585–590.
[4] Caggiano, A., Zhang, J., Alfieri, V., Caiazzo, F., Gao, R., & Teti, R. 2019. Machine learning-based image processing for on-line defect recognition in additive manufacturing. CIRP annals, 68(1), 451-454.
[5] Gammack, J., Akay, H., Ceylan, C., & Kim, S. G. 2022. Semantic knowledge management system for design documentation with heterogeneous data using machine learning. Procedia CIRP, 109, 95-100.
[6] Vaswani, A., Shazeer, H., Parmar, N., Uszko, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 5998–6008.
[7] Jones, C., & Bergen, B. (2024). Does GPT-4 pass the Turing test?. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1) (pp. 5183-5210).
[8] Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J. B., Yu, J., & Ahn, J. 2023. Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805.
[9] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., et al. 2020. Language Models Are Few-Shot Learners. Advances in Neural Information Processing Systems 33: 1877–1901.
[10] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., et al. 2023. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
[11] Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 5185-5198).
[12] Altshuller, G. S., & Shapiro, R. B. 1956. Psychology of inventive creativity. Issues of Psychology, 6, 37-49.
[13] Suh, N. P. 1997. Design of systems. CIRP Annals, 46(1), 75-80.
[14] Ulrich, K. T., & Eppinger, S. D. (2016). Product design and development. McGraw-hill.
[15] Wang, P., Karigiannis, J., & Gao, R. X. (2024). Ontology-integrated tuning of large language model for intelligent maintenance. CIRP annals, 73(1), 361-364.
[16] Ellsel, C., & Stark, R. (2024). A knowledge-driven, integrated design support tool for additive manufacturing. Proceedings of the Design Society, 4, 1747-1756.
[17] Akay, H., Yang, M., & Kim, S. G. 2021. Automating design requirement extraction from text with deep learning. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 85390, p. V03BT03A035). American Society of Mechanical Engineers.
[18] Akay, H., Lee, S. H., & Kim, S. G. 2023. Push-pull digital thread for digital transformation of manufacturing systems. CIRP annals, 72(1), 401-404.
[19] Taylor, W. L. 1953. Cloze procedure: A new tool for measuring readability. Journalism quarterly, 30(4), 415-433.
[20] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT* (Vol. 1, p. 2).
[21] Jugé, A., Moreno-Villafranca, J., Perez-Puyana, V. M., Jiménez-Rosado, M., Sabino, M., & Capezza, A. J. (2023). Porous thermoformed protein bioblends as degradable absorbent alternatives in sanitary materials. ACS Applied Polymer Materials, 5(9), 6976-6989.