Design and Artificial Intelligence 1 (2025) 100030

Contents lists available at ScienceDirect

Design and Artificial Intelligence

journal homepage: https://www.keaipublishing.com/en/journals/design-and-artificial-intelligence/

Full Length Article

Trailblazing specific generative models (SGMs) for early-stage engineering design concepts ☆

Chijioke C. Obieke [a], Jelena Milisavljevic-Syed [b], Pingfei Jiang [c], John Bridgeman [d], Ji Han [c,*]

[a] School of Mechanical and Aerospace Engineering, Queen's University Belfast, Belfast BT9 5AH, United Kingdom
[b] Sustainable Manufacturing Systems Centre, School of Aerospace, Transportation and Manufacturing (SATM), Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom
[c] INDEX, Business School, University of Exeter Tintagel House, 92 Albert Embankment, London SE1 7TY, United Kingdom
[d] Department of Civil and Environmental Engineering, School of Engineering, University of Liverpool, Brownlow Hill, Liverpool, L69 3GH, United Kingdom

ARTICLE INFO

Keywords:
AI collaboration
Specific generative models
Concept generation
Early-stage design

ABSTRACT

Recent advancements in large language models (LLMs) have the potential to revolutionise engineering design practices. LLM-powered tools such as ChatGPT and DeepSeek have shown significant promise in enhancing various aspects of engineering design. They leverage emergent data-driven technologies to deliver powerful capabilities that challenge the long-held belief that generating new concepts and automating engineering design activities are beyond the reach of artificial intelligence (AI). The capabilities demonstrated by these technologies are gradually overcoming initial resistance and fostering a more receptive attitude towards their adoption. However, a substantial gap remains between the generic capabilities of AI technologies and their specific application in engineering design. This study aims to leverage a combination of AI technologies to support early-stage engineering design activities. Specifically, we created a generative Markov chain model adjusted as a specific generative model (SGM) to support the complex and time-consuming conceptual early-stage design activities. The SGM does not require extensive and expensive data training like the generic LLM, but produces comparably more natural concepts. The SGM used in this study effectively addresses the privacy concerns associated with generic LLMs. This pioneering research on SGM is cost-effective and holds promise for applications beyond engineering design, including fields such as medicine.

1. Introduction

At the early stage of engineering design, design engineers engage in conceptual activities crucial to the design project's success. These activities, which include conceptualising a new engineering design problem and solution, are known to be iterative, challenging, and time-consuming [1–3]. Despite the importance and demanding nature of early-stage engineering design activities, there is limited research specifically supporting them with AI technologies. LLMs represent the state-of-the-art in AI, with potential applications across various fields. They offer various capabilities, including generative, text refinement, language translation, and question answering capabilities [4]. The generative capabilities of LLMs make them particularly desirable in early-stage engineering design, where success is defined by the creation of new and useful concepts. However, specialised applications of LLMs are constrained by their generic nature, substantial training resource demands, and privacy issues.

This reveals a significant gap between the broad, generic capabilities of LLMs (such as text generation, refinement, and image generation) and their practical application in addressing the unique challenges of early-stage engineering design. For instance, LLMs excel at generating text-based outputs but often lack the domain-specific knowledge required to produce technically accurate engineering concepts, such as novel mechanical designs or optimised product configuration concepts. Additionally, their inability to seamlessly integrate engineering design workflows limits their effectiveness in supporting iterative conceptual design tasks [5,6]. Consequently, it is crucial to investigate specific generative models (SGMs) tailored for engineering design applications to address the constraints of generic LLMs. Research on how design engineers can collaborate with AI to enhance engineering design practices is becoming increasingly important [7].

The concept of the SGM presented in this study represents a novel approach that harnesses the advanced generative capabilities of LLMs for specialised applications within specific fields. This approach involves in-

☆ Peer review under responsibility of Editorial Board of Design and Artificial Intelligence

* Corresponding author.
E-mail addresses: c.obieke@qub.ac.uk (C.C. Obieke), j.milisavljevicsyed@cranfield.ac.uk (J. Milisavljevic-Syed), p.jiang@exeter.ac.uk (P. Jiang), john.bridgeman@liverpool.ac.uk (J. Bridgeman), j.han2@exeter.ac.uk (J. Han).

https://doi.org/10.1016/j.daai.2025.100030
Received 31 July 2025; Received in revised form 17 September 2025; Accepted 27 September 2025
Available online 9 October 2025
3050-7413/Copyright © 2025 The Authors. Publishing services by Elsevier B.V. on behalf of Zhejiang University Press Co., Ltd. and KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

tegrating these capabilities with other computational or AI technologies, offering significant advantages over the use of generic LLMs with limited specialised applications. As mentioned, LLMs are typically general-purpose or multimodal [8], creating a notable gap in research focused on AI models tailored to the specific needs of specialised domains. Users are often compelled to adapt general-purpose LLMs, which are not optimised for particular fields or applications [9].

As a step towards SGMs in engineering design, this study aims to leverage a combination of AI technologies to enhance conceptual activities in early-stage engineering design. Across various fields, AI technologies are being explored for new applications to support challenging tasks. It is suggested that the potential of these technologies could be enhanced by combining them or reinforcing them with other technologies [10]. This study uses an adjusted generative Markov chain (GMC) or Markovian model as an SGM to support the engineering design textual concept generation. The adjustment of the GMC model includes text refinement with a generative pre-trained transformer (GPT) model [11]. A GPT model is a deep learning LLM trained in advance for use in various applications, so that the required training is not performed from scratch for such applications [12,13]. A Markovian model, on the other hand, is any model that exhibits a Markov chain property [14]. It can be adjusted as a GMC model to generate a new text based on example inputs and does not require extensive and expensive data training and retraining like the LLM [15]. The GMC model had been tested in a case study to successfully frame design problems, in which the solutions were considered of societal relevance by experienced and novice design engineers [15]. The case study results necessitated leveraging an LLM capability to create an effective SGM to support early-stage engineering design.

The next section reviews the early-stage design activities, including challenges, benefits, and computational support. Section 3 presents the SGM as a combination of a GMC model and text refinement features of an LLM. The results of the SGM performance evaluation are presented in Section 4, including its comparison to Copilot using ChatGPT, DeepSeek, and Grok. The discussion in Section 5 is on the study results, including how the SGM improves on privacy issues compared to LLMs. The conclusion of the study is given in Section 6.

2. Challenges and strategies in identifying engineering design concepts

2.1. Challenges and benefits of new design concepts

An engineering design concept can be effectively expressed through both text and images. A concise design brief or project title can describe an engineering problem or solution textually, while images often illustrate or inspire potential solutions. Combining text and images ensures comprehensive and inspirational support for new design concepts. Conceptualising and identifying problems and solutions are distinct yet complementary early-stage design cognitive activities, as the identification of a problem triggers a solution-seeking process [16]. Identifying design problems and ideas on how to solve them are challenging activities for design engineers and engineering design firms [17]. Design engineers are expected to establish (bring into existence or awareness) a new design problem of societal value and develop solution concepts for the problem [18,19]. Conceptualising and identifying new design problems is recognised as a particularly challenging activity [17]. As innovators of next-generation technologies, design engineers are expected to meet and overcome this challenge [20–23]. The benefits of successfully doing so are invaluable, with significant positive societal impact, as illustrated in Fig. 1.

The two foundational activities in early-stage engineering design are illustrated in Fig. 1. One involves conceptualising or identifying an unknown or existing problem. The other entails creating an unknown solution or identifying an existing solution for a given design problem. The term “unknown” is used in a global sense, meaning “new” on a broader

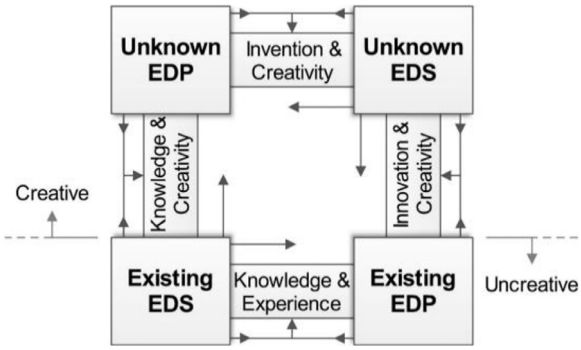

Fig. 1. The foundational activities in engineering design.

scale. For instance, a “new” house may not be globally new, as it could have been built years ago, but it is locally new to the owner. Therefore, an unknown problem or solution requires discovery or conceptualisation and is one that society has limited or no knowledge of. Conceptualising an unknown problem, as illustrated in Fig. 1, is crucial for developing unknown solutions that can lead to inventions. An invention is defined as a concept that experts in the field find unusual [24]. It seeks the truly new (an unknown problem or solution), which is time-consuming and offers no guarantee of a return on invested time, resources, and assets [3,25]. However, while not all inventions are driven by financial benefits [26], they can yield significant economic advantages [27,28]. Without conceptualising or identifying a new problem, the likelihood of obtaining a new solution necessary for invention is reduced [29,30].

Although a new design solution can address an existing problem, such combinations typically result in innovation, especially when the new solution offers a superior alternative to a prior one. Innovation often involves altering established methods [31]. Sometimes, an existing design solution reaches a point where further modifications are impractical without a radical breakthrough. As illustrated in Fig. 1, all activities involving something unknown or new necessitate creativity. Solving an existing problem with an existing solution falls into the uncreative region, as it relies on knowledge and experience rather than creativity. Identifying truly novel ideas is widely recognised as a complex and time-consuming process [1]. Conceptualising a new design problem requires divergent thinking. As reported, “generating ideas through divergent thinking is more difficult than evaluating and selecting ideas through convergent thinking; the ability to think divergently is essential to creativity and is regarded as a reliable predictor of creativity” [32]. Divergent thinking can lead to cognitive fatigue—cerebral exhaustion within 30 min of performing a challenging task, resulting in no tangible or creative output [33,34]. Cognitive fatigue can induce physiological issues such as creative burnout, frustration, tiredness, or withdrawal from a task [35–39].

2.2. Computational support at early-stage engineering design

In engineering design, computational approaches significantly support various designer actions [40]. This is primarily due to the limitations of human rationality and the narrow focus of human attention, which fundamentally influence the design process. Computers enable us to manage more information than previously possible [41]. The need for creativity at the early stages of engineering design drives many computational cognition support tools. Despite creativity being essential for successful engineering design, some design engineers believe they are not creative, cannot achieve creativity, or find creativity challenging [42,43]. Creativity or cognition support tools enhance users’ ability to make discoveries or inventions from the early stages of information gathering, including identifying new design problems [44]. For example, the Combinator supports ideation at the early design stage by juxtaposing or superimposing unrelated image pairs to produce unfamiliar

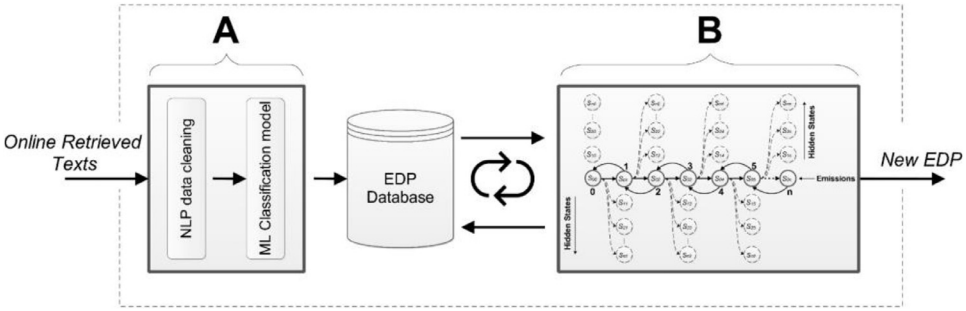

Fig. 2. A generative Markov chain (GMC) model for framing design problems.

solution concepts, inspiring design engineers [45]. The Collaborative Ideation Partner (CIP) is another computational support tool used in the early design stage [46]. CIP allows a design engineer to specify a design task and make an initial design sketch on one-half of its graphical user interface (GUI). It then displays conceptually similar images on the other half of its GUI, inspiring the engineer to add new features to the initial design sketch. Idea-Inspire 4.0 is a web-based support tool that enables design engineers to use nature's inspirations, such as the motions of biological entities, to solve engineering design problems [47]. This tool employs design-by-analogy, a powerful method for developing ideas using natural inspiration [48].

2.3. Emergent technologies to support design concept generation

Recent advancements in LLMs have significantly expanded opportunities to support challenging conceptual activities in the early stages of design. Prominent LLMs include the GPT, Prometheus, and Gemini models. The OpenAI GPT model powers AI tools such as ChatGPT and DALL-E [49], while the Microsoft Prometheus model, in combination with GPT, powers the Copilot AI tool [50]. The Google Gemini model powers the Gemini AI tool [51]. The GPT model is regarded as a state-of-the-art technology in natural language processing (NLP), capable of generating high-quality text from natural prompts [52]. One notable application of GPT models is the generation of new datasets, including text and images [53]. These models can create realistic text and images from prompts or natural language descriptions [54]. Studies have demonstrated the GPT model's proficiency in generating images with specified elements and producing images that approach human standards [55,56]. Additionally, the GPT model consistently generates conversational texts that are grammatically correct and suitable for essays, assignments, and online tests [57,58]. The capabilities of the GPT model can be integrated into other computational technologies via an application programming interface (API) [59,60].

2.4. Optimising a GMC model with an LLM

A Markov chain is a probabilistic model characterised by the Markov property, where the next state of a sequence depends solely on the current state, not on past states [14]. Consequently, a Markov chain model is unidirectional, moving only forward. In contrast, an LLM is autoregressive, determining the next state in a sequence based on past states [61–64]. A Markovian model can be generative with some adjustments and without extensive data training. For instance, a GMC can create new design problems using given examples without requiring large datasets or costly data training [15]. This GMC model, illustrated in Fig. 2, leverages the Markov property to reduce model size and computing resources and increase computing speed. In specific applications, small model sizes and low computational power consumption are desirable [65]. The Markov property relies on forward determinism (using the present condition to predict future events), which is considered more important than backward determinism (or autoregression) in deterministic chaos theory [66].

The Markovian model shown in Fig. 2 is data-driven and comprises two systems: *System A* and *System B*. In System A, online-retrieved texts are classified using machine learning (ML) techniques. These texts, which are one-line briefs or engineering design project titles, are manually labelled into two categories: those describing an engineering design problem (EDP) and those that do not. This classification is necessary since not all retrieved data clearly describe an EDP. The labelled data categories train the ML model to classify new texts. Thus, System A is a discriminative ML model that accepts input data and predicts the subgroup to which the data belong [67]. The continuous collection of these texts constitutes big data due to their increasing availability, variety, and speed of generation and processing [68,69]. Texts classified as EDPs in System A are stored in the EDP Database and form the input to System B, which is a Markovian model. In System B, the Markov property is utilised to process input text from System A, producing a distinct yet comparable output. The text is tokenised into words, each representing a Markov state. A new EDP is generated by randomly selecting an initial word and constructing a sequence where each subsequent word depends solely on the current word. Implemented in Python, this method maintains the natural linguistic flow to mimic coherent EDP generation. Since the goal is to frame a new EDP, repeating an existing EDP is undesirable. Therefore, a duplication recognition search, indicated with the feedback loop in Fig. 2, ensures that a new EDP from the GMC model does not duplicate any in the EDP Database. While the GMC model can generate new and useful outputs without data training, the readability of the outputs remains inconsistent.

An LLM can optimise the readability of the GMC model output, and the LLM used in this study is the GPT model. While there are alternatives for optimising GMC model outputs, as discussed in Section 3.2, the GPT model currently offers the best solution. Additionally, the GPT model saved training time and resources that would be required to train a deep learning model from scratch for this study. The only feature of the LLM needed for this task is text refinement, and a small GPT model can be specifically trained for this with less data and without further retraining, unlike LLMs. The model in Fig. 2 is realised as a web-based support tool, Pro-Explora, and has been empirically validated to support generating new and useful EDPs [15]. Combining the strengths of GMC and LLM presents a leverageable opportunity, translatable to supporting challenging concept generation activities in the early stages of design.

This section highlights the availability of computational tools for early-stage design concept generation. While some tools target design solutions, others are too general to effectively support early-stage design activities. Despite the critical role of new EDPs in innovation and invention (Section 2.1), there is a lack of specific computational support for conceptualising and identifying new EDPs [70,71]. Typically, the conceptualisation of a new EDP and its solution occurs simultaneously, indicating the need for integrated computational support. However, no comprehensive model or tool currently exists to support these activities

Table 1
Case study participants' details.

Participant	Gender			Academic qualification (obtained/in view)			Experience (years)	
	Male	Female	Total	B	MS	PhD	≤ 3	> 3
Novice	12	7	19	5	11	3	19	0
Experienced	23	1	24	10	9	5	0	24
Total	35	8	43	15	20	8	19	24

Note: B, Bachelors; MS, Masters.

fully. Although such a tool would facilitate rapid concept generation, research in this area is insufficient. This study addresses these research gaps.

3. Methodology

This section presents a method for an SGM to support early-stage design activities. A user experience case study was conducted to verify the performance of the GMC model in Fig. 2. Based on the case study feedback, the GMC model is optimised using LLMs, resulting in the SGM (GMC + LLM) model. The SGM is evaluated for readability and similarity of its output compared to the GMC model and human output, respectively. Additionally, ChatGPT, DeepSeek, and Grok were used to compare the outputs of the SGM with the Copilot outputs.

3.1. User-experience case study

Table 1 presents the demographic profiles of the 43 participants involved in the case study to verify the usefulness of the GMC model output. The model shown in Fig. 2 has been translated into a web-based tool, Pro-Explora V1 [15], enabling the case study to be conducted online. This allows participants to explore the GMC model's capabilities. As part of the case study, participants were asked to provide feedback on the tool upon completion. They were automatically redirected to a feedback page, where they shared their opinions to aid in improvements. All responses were collected online in text format and unstructured.

The case study was advertised through channels targeting professionals and students (Year 2 and above). All participants listed in Table 1 have engineering design-related backgrounds from various countries, including Canada, France, India, Russia, Singapore, the United Kingdom, and the United States. Participants with less than three years of work experience are classified as "Novice," while those with three or more years are classified as "Experienced." For the Novice group, the mean (M) work experience is 1.5 years, with a standard deviation (SD) of 0.7 years and a range (R) of 2 years. For the Experienced group, the mean work experience is 8 years, with an SD of 3.8 years and an R of 17 years.

An inclusion criterion of 0–40 years of work experience is applied to ensure that participants can easily perform computer-based activities. The GUI of the web tool for the case study is shown in Fig. 3, with an example of a logged-in user, Engineer. The two optional settings for using the tool are: 1) "Select explore domain" and 2) "Select number of problems to explore." The "Select explore domain" setting offers four options: "Engineering design product," "Engineering design research," "Engineering design machine intelligence," and "Engineering design cross-domain"—the latter having the largest database, combining the other three domains.

The case study web tool's "Select number of problems to explore" setting, illustrated in Fig. 3, ranges from 1 to 5, with 1 as the default. The "List of explored domain(s)" section of the GUI displays the history of all explored domains. Activating the Explore/ReExplore button initiates a new EDP-framing process based on the selected settings. Fig. 3 showcases an example of three EDPs framed by Pro-Explora, which can be selected and saved for future review. Pro-Explora can frame over 100 unique EDPs per minute, a feat unattainable by the human brain without support due to cognitive limitations, as discussed in Section 2.1. Participants utilised Pro-Explora V1 to frame five EDPs in 10 min, excluding time spent on other activities such as completing demographic questionnaires, reading instructions, and watching instructional podcasts. In total, 43 participants framed 215 design problems using Pro-Explora V1. Each participant rated the perceived usefulness of the design problems on a Likert scale of 1 (lowest) to 10 (highest), with an average rating of 7.74 out of 10.

To indicate the variabilities of the 215 EDPs framed by Pro-Explora V1 in the case study, each EDP was vectorised into numerical arrays using the term frequency–inverse document frequency (TF–IDF) method. TF–IDF, widely used in studies, determines the frequency of a word in a document and assigns a weight based on its importance [72]. Consequently, individual words or tokens in a sentence have different weight values, summing to various averages. This method is ideal for observing variabilities in the GMC model output, as discussed in Section 5. TF–IDF is the product of term frequency (TF) and inverse document frequency (IDF), as shown in Eq. 1 [72–74].

$$\text{TF} - \text{IDF} = \text{TF}(t, d) \cdot \text{IDF}(t, D) \tag{1}$$

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{2}$$

$$\text{IDF}(t, D) = \log\left(\frac{D}{\sum_{t \in D} f_{t,D}}\right) \; (d \in D : t \in d) \tag{3}$$

where $f_{t,d}$ is the number of times a token (t) appears in a document (d) in a corpus; $\sum_{t' \in d} f_{t,d'}$ is the total tokens in d; D is the total documents in the corpus containing d; and $\sum_{t \in D} f_{t,D}$ is the number of d containing t in D. Note that the logarithm in Eq. 3 is base ten. The TfidfVectorizer library from the Python sklearn NLP package efficiently applies TF–IDF and is used in this study. During the TfidfVectorizer process, the corpus is vectorised to obtain TF–IDF scores [75].

3.2. Case study-based optimised GMC model

The participants' feedback from the case study indicates that an integrated tool supporting the identification of design problems and solutions would be highly beneficial. One participant noted, "Pro-Explora seems a very good tool to find problems and work on their solutions." Currently, there is a scarcity of engineering design tools or SGMs that facilitate both the identification of design problems and the exploration of solution concepts. Another participant remarked, "I have seen AI-based design solutions before, but never AI problem identification. It sometimes generates quite 'inhuman' sentences, which can be difficult to read, but overall, it is quite effective." This feedback underscores the readability challenges associated with the design problems generated by Pro-Explora. The collective feedback from all participants was utilised to optimise the GMC model into an SGM for early-stage engineering design, as illustrated in Fig. 4.

Participants' feedback is crucial to this study. Owing to the study's small scale, manual coding was employed to analyse the feedback, ensuring thorough interpretation. This approach, though more time-consuming and labour-intensive, offers deeper contextual understanding compared to software-based methods, avoiding unnecessary complexity and providing a more authentic interpretation [76,77]. At least ten

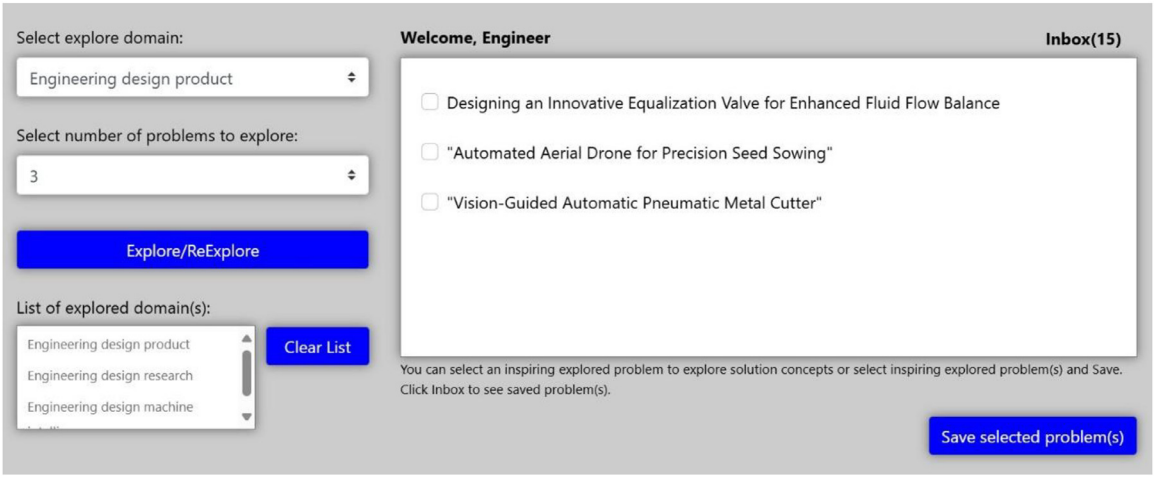

Fig. 3. Pro-Explora V1 graphical user interface.

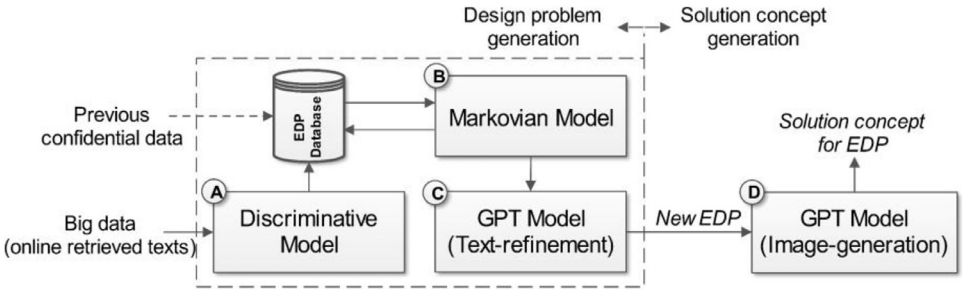

Fig. 4. A specific generative model (SGM) for early-stage design concept generation.

participants frequently noted the need to refine the tool's output and integrate a feature for exploring new EDPs and their solutions. Other comments highlighted minor technical issues with the tool. However, all feedback was overwhelmingly positive.

Fig. 4 shows an SGM (GMC + LLM) comprising four systems: System A, System B, System C, and System D, which support the identification of a new EDP and its solution concepts. As mentioned in Section 2.4, the LLM used in this study is the GPT model as indicated in Fig. 4. Systems A, B, and C form the core of the SGM, facilitating the identification of a new EDP, while System D is optional and aids in exploring solution concepts for the EDP. In Fig. 4, Systems A and B represent Systems A and B in the GMC model shown in Fig. 2. As detailed in Section 2.4, the input data for System A is retrieved online using Scrapy, a Python data extraction tool. These online sources include journals and conference databases, with the target texts being the titles of previous engineering design projects that describe the EDP [78,79]. Confidential company records, medical information, or musical lyrics can be stored directly in the database or processed by System A for classification, if required (see Section 2.4). The Markovian model prevents the entire data from being passed directly to an LLM for training, thereby preventing direct data logging. It also acts as an encryptor, passing previous data to the LLM without directly referencing the original data, as the output from System B is unique and different from the input. This is discussed further in Section 5.

Next, the three major optimisations of the GMC model (Fig. 2) following its case study and the alternatives to the GPT model used for the SGM are discussed. These optimisations include readability, complementary solution concept exploration, and database self-reinforcement.

Optimisation 1: Pro-Explora V1 output readability

The readability of System B in Fig. 2 is enhanced in Fig. 4 using a GPT-3.5 model, which refines incoherent sentences, as discussed in Section 2.3. The output of the Markovian model (System B, Fig. 4) is improved in System C by instructing the GPT-3.5 model to reword the output. Rewording, which involves expressing the same concept differently, enhances both comprehension and conceptualisation [80,81]. To maintain the original concept generated by the Markovian model, the "temperature" parameter of the GPT-3.5 model is set to a low value of 0.2. The temperature controls the randomness or dissimilarity of the reworded output compared to the original [82], with values ranging from 0 to 2.

Optimisation 2: solution concept exploration for GMC model-framed EDP

Based on participants' feedback from the Pro-Explora use case study (Section 3.1), an image generative model is integrated into the GMC model (Fig. 2) to create the SGM (Fig. 4), to enable exploring solution concepts for the EDPs generated by the GMC model. The reworded EDP from System C (Fig. 4) is used as a prompt for System D, which includes the GPT image model, DALL-E 3. The style parameter is set to "natural" to generate photorealistic images that facilitate design ideation. Additionally, the *n* parameter is set to 2 to produce two images per prompt. According to concept–knowledge theory, creative design emerges from aligning new concepts with relevant knowledge [83]. To support this, each EDP framed by the Markovian model in the SGM is mapped to engineering design solution (EDS) concepts, as shown in Fig. 5. This approach provides comprehensive support for early-stage engineering design activities, aiding in both identifying design problems and mapping them to potential solutions.

Optimisation 3: self-reinforcement of the GMC model database

The EDPs framed by the GMC model and deemed useful are added to the database for self-reinforcement. For instance, EDPs that participants selected and rated highly for usefulness are included in the database. This is in addition to the continuous texts retrieved from online sources

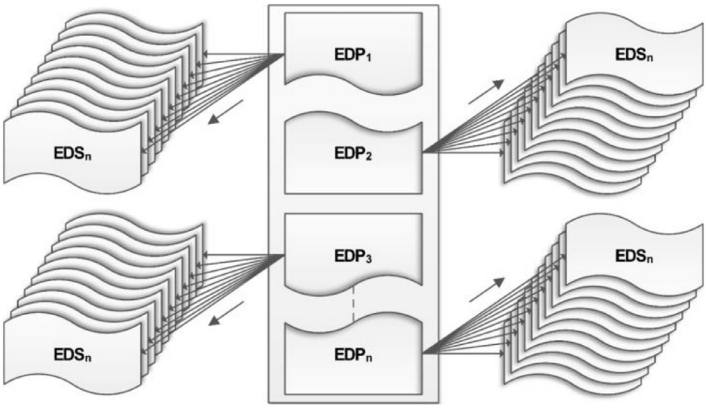

Fig. 5. Mapping an engineering design problem (EDP) to infinite engineering design solution (EDS) concepts.

(Section 2.4). This self-reinforcement technique is beneficial because it adds EDPs to the relevant domain databases, increasing the model's output variability and saving time and resources needed to retrieve and classify text online within the discriminative model (System A, Figs. 2 and 4).

Alternatives for GMC model Optimisation 1: Gramformer, a Python library for improving text readability [84], could enhance the GMC model output. However, its results are not as sophisticated as those produced by the GPT-3.5 LLM used in the SGM in this study. Alternatively, a small language model (SLM) can be specifically trained for text refinement and integrated into the SGM [84]. Unlike an LLM, an SLM can run on local drives rather than the cloud, requiring fewer resources for training in text refinement applications. Additionally, using an SLM ensures superior privacy for this and similar SGM applications, as discussed in Section 5.

Alternatives for GMC model Optimisation 2: As an alternative to using the GPT image model, Google Images can be retrieved to inspire potential solutions for GMC model-generated problems. This approach is utilised in the CIP computational support tool discussed in Section 2.2. ConceptNet and Google image retrieval can also be employed to explore solution concepts for GMC model design problems [85,86]. ConceptNet is a semantic network that provides extensive background knowledge for applications working with natural language text [87]. However, these alternatives have limitations: They are not generative and do not create new images like a GPT image model. ConceptNet accepts a noun as a prompt, while a GPT image model uses a sentence description. Unlike Google Images, GPT model images are not subject to copyright issues, as their prior existence is not traceable to a specific source. This allows them to be freely used as part of a design document.

3.3. Performance verification for GMC model and SGM outputs

An evaluation is conducted to compare the performance of the GMC model and the SGM (or GMC LLM-optimised) model using the Pro-Explora V2 tool. The evaluation also tests the performance of the SGM against Copilot using popular AI tools—ChatGPT, DeepSeek, and Grok. The focus is on the readability of the outputs from the GMC model and SGM, as highlighted in the case study feedback for the GMC model (Section 3.1). This evaluation aims to demonstrate that an SGM can offer a cost-effective and data-secure approach for specific applications. An interface of Pro-Explora V2 displaying generated EDPs is shown in Fig. 6, while another interface with solution concepts is shown in Fig. 7. The tool is accessible at https://www.explorefoss.com.

Figs. 6 and 7 reflect the SGM illustrated in Fig. 4. Systems A, B, and C in Fig. 4 facilitate the output shown in Fig. 6, while System D facilitates the output in Fig. 7. The interface in Fig. 7 is activated by selecting any Pro-Explora V2 framed EDP in Fig. 6 and clicking the "Explore solution concepts" button, as explained in Section 3.1. For example, in Fig. 6, the first displayed EDP is selected, and the "Explore solution concepts" button is clicked to display the solution concepts shown in Fig. 7. This capability is based on the participants' needs and was unavailable before the case study. Clicking the "ReGenerate concepts" button generates a new set of two EDS image concepts for the displayed design problem. Clicking "Back to problem-exploring" returns the user to the interface shown in Fig. 6. The "More Design Tasks" button can be used for additional design activities not covered in this study. The outputs of the GMC model were compared with those of the SGM using the standard readability metrics presented in Table 2.

Table 2
Readability metrics for comparing outputs of the generative Markov chain (GMC) model and specific generative model (SGM).

Metric	Metric scale	Metric scale interpretation
Flesch reading ease (FRE)	0–100	The higher the score, the easier to read and understand a given text. A score of 0–30 indicates that the text is difficult to read and meant to be understood by professionals or graduates [88].
Flesch–Kincaid grade level (FGL)	0–18	The higher the score, the higher the academic level required to understand a text. A score above 12 indicates that the text is for college or graduate reading levels based on the United States educational system [88,89].
Gunning fog index (GFI)	0–20	The higher the score, the higher the academic level required to understand a text. A score of 13–17 indicates that the text is for college-level understanding, while a score above 17 indicates postgraduate-level understanding based on the United States educational system [90,91].

The metrics in Table 2 are standard readability metrics, obtained using Eqs. 4–6. Using en_core_web_lg, a Python open-source library for NLP, the cosine similarities between the EDPs framed by the GMC model and the SGM were determined and compared to human-framed EDPs [92]. The cosine similarity ranges from 0 to 1, with 1 indicating 100 % similarity. In Eqs. 4–5 [93], ASL represents the average sentence length for an assessed text, and ASW denotes the average syllables per word in the text.

$$\mathrm{FRE} = 206.835 - (1.015 \times \mathrm{ASL}) - (84.6 \times \mathrm{ASW}) \tag{4}$$

$$\mathrm{FGL} = (0.4 \times \mathrm{ASL}) + (12 \times \mathrm{ASW}) - 15 \tag{5}$$

$$\mathrm{GFI} = 0.4 \times (\mathrm{ASL} + \text{hard words}) \tag{6}$$

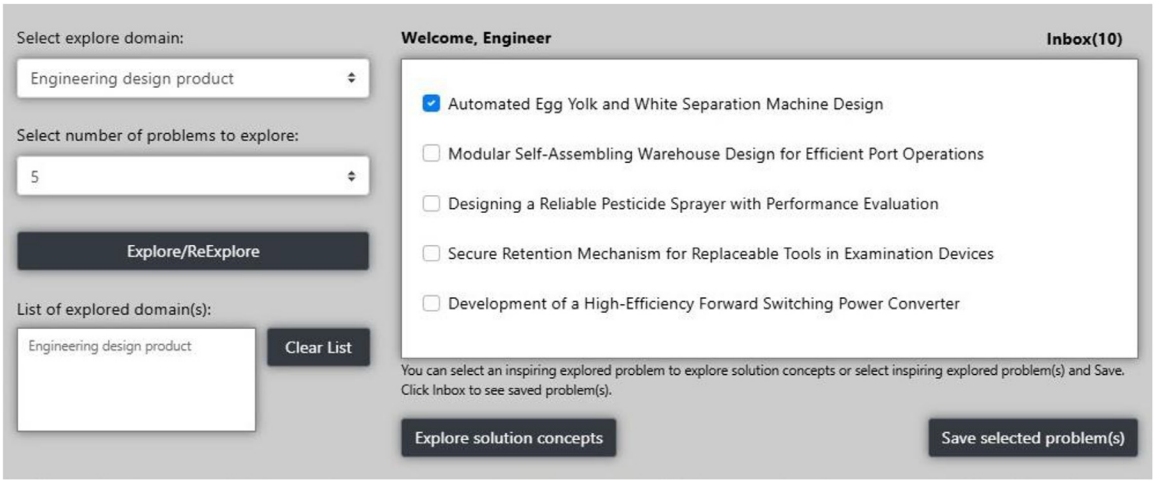

Fig. 6. Graphical user interface (GUI) for Pro-Explora V2.

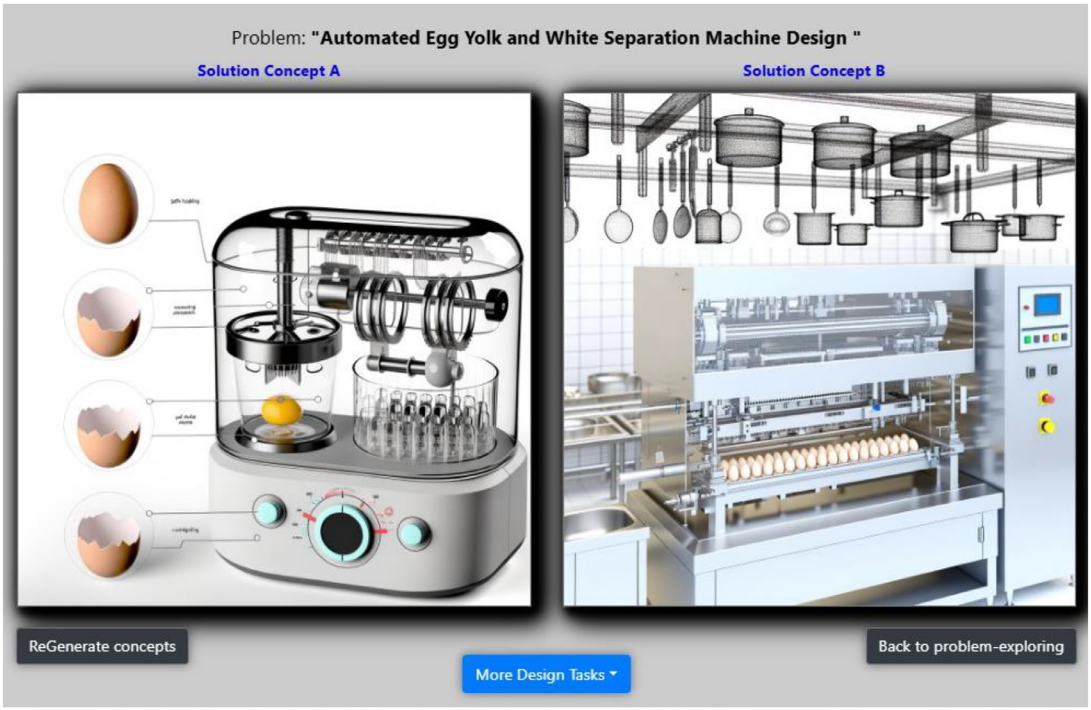

Fig. 7. Pro-Explora V2 solution concept GUI.

4. Results

This section presents the results of the study, evaluating the objective of employing integrated AI technologies to facilitate early-stage engineering design activities. It specifically examines the variability of the GMC model outputs as a generative tool, the efficacy of optimising the GMC model as an SGM, the similarity between SGM outputs and human-generated designs, and the comparability of SGM outputs—produced without extensive training—to those of LLMs trained with substantial data and computational resources.

4.1. Variability in GMC model output

Fig. 8 presents a box plot illustrating the variability in the EDPs framed by the GMC model for all participants in the case study discussed in Section 3.1. Box plots are widely used to display variability [94]. The box plot in Fig. 8 provides insights into the variability in EDPs for each participant's five EDPs and across the 43 participants. It highlights outliers and the spread of TF–IDF scores for each participant based on the range and interquartile range (IQR). Comparing the spreads in Fig. 8 helps observe the variability of the generated EDPs across all participants.

4.2. SGM performance evaluation

Table 3 presents the results of the readability evaluation for the outputs of the GMC model and the SGM. Details of the standard metrics used are provided in Section 3.3. To meet the 100-word minimum requirement for readability assessment using the metrics in Table 2 [95], the EDPs of three participants were merged, resulting in 14 groups.

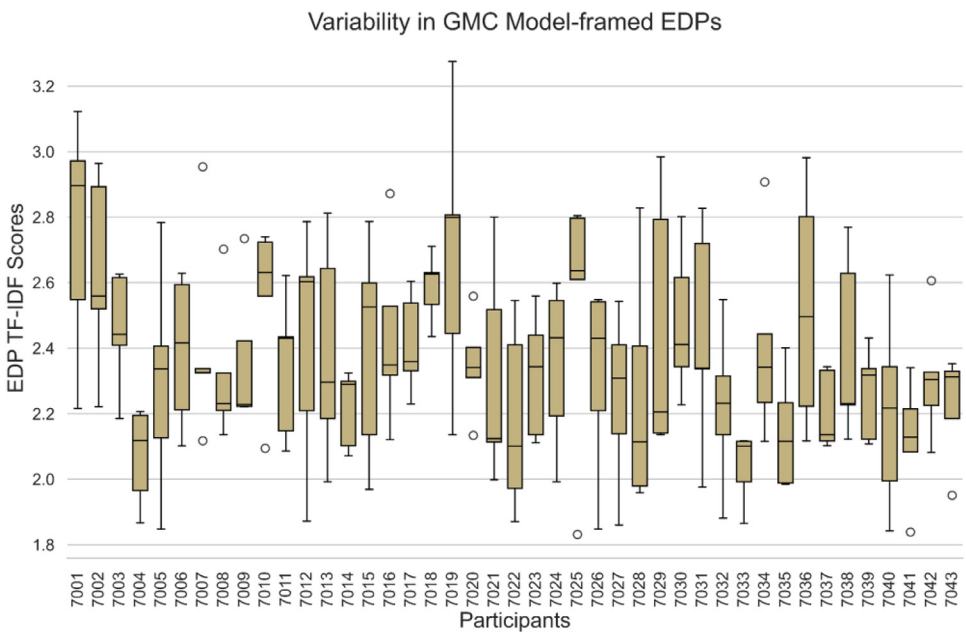

Fig. 8. Variability in participants' GMC model-framed EDPs.

Table 3
Performance evaluation metrics for GMC model and SGM outputs.

Group	Participants Merged	Number of Words		Output Readability Scores					
		GMC Model	SGM Model	GMC Model			SGM Model		
				FRE	FGL	GFI	FRE ↑	FGL ↓	GFI ↓
1	7001, 7002, 7003	130	159	-10.4	16.8	14.8	18.7	13.3	13.3
2	7004, 7005, 7006	103	134	32.9	10.3	8.9	41.1	9.6	9.4
3	7007, 7008, 7009	112	134	23.8	11.7	9.3	28.6	11.4	9.7
4	7010, 7011, 7012	114	144	2.2	14.8	14.5	24.1	12.2	12.7
5	7013, 7014, 7015	103	135	11.5	13.3	12.7	28.6	11.4	10.0
6	7016, 7017, 7018	124	132	10.7	13.7	14.8	19.1	12.7	12.6
7	7019, 7020, 7021	114	140	19.5	12.4	10.6	36.8	10.3	10.7
8	7022, 7023, 7024	107	136	30.3	10.7	10.5	42.6	9.5	9.7
9	7025, 7026, 7027	110	130	22.3	11.9	11.1	32.6	10.7	9.8
10	7028, 7029, 7030	113	130	13.2	13.2	12.0	15.4	13.2	13.3
11	7031, 7032, 7033	102	132	9.9	13.5	12.9	38.4	10.0	10.4
12	7034, 7035, 7036	118	143	23.6	11.8	12.7	28.6	11.5	12.5
13	7037, 7038, 7039	109	136	12.1	13.3	13.5	25.6	11.8	13.0
14	7040, 7041, 7042, 7043	137	162	18.1	12.5	12.2	34.8	10.3	10.7

FRE: Flesch reading ease, FGL: Flesch–Kincaid grade level, GFI: Gunning fog index.

Group 14 in Table 3 includes four participants due to the remainder when 43 participants are divided into 13 groups. The readability scores for the EDPs generated by participants using the GMC model are shown in Table 3. These EDPs were then refined using the GPT-3.5 model, and the group readability scores are recorded in the SGM column in Table 3. The number of words in the EDPs increased slightly due to the GPT-3.5 model refinement. Regression analysis of the metrics for the GMC model and SGM shows Flesch reading ease (FRE) ($r = 0.67$, $p = 0.008$), Flesch–Kincaid grade level (FGL) ($r = 0.74$, $p = 0.003$), and Gunning fog index (GFI) ($r = 0.77$, $p = 0.001$), with p-values < 0.05 indicating statistical significance. As shown in Table 3, based on the standard definitions in Table 2, improved readability is evidenced by a relative increase in FRE scores and a decrease in FGL and GFI scores.

4.3. Cosine similarities

Table 4 presents the cosine similarity results for EDPs framed by humans and those generated by the GMC model and SGM. EDP1–EDP5 are conceived by design engineers without computational support. EDP6–EDP10 are randomly selected from participants' generated EDPs, with

Table 4
Comparisons of EDPs framed by human, GMC model, and SGM.

	EDP6 (EDP6')	EDP7 (EDP7')	EDP8 (EDP8')	EDP9 (EDP9')	EDP10 (EDP10')	Q*
EDP1	0.8411 (0.6870)	0.9250 (0.8937)	0.7520 (0.7500)	0.8653 (0.8308)	0.8642 (0.8501)	0.3095
EDP2	0.6418 (0.7315)	0.7692 (0.7744)	0.6602 (0.7929)	0.7330 (0.7988)	0.6581 (0.6503)	0.3600
EDP3	0.6369 (0.5392)	0.7922 (0.7808)	0.6022 (0.6991)	0.7031 (0.7502)	0.6744 (0.7180)	0.2487
EDP4	0.6892 (0.8015)	0.7935 (0.8162)	0.6636 (0.6307)	0.7400 (0.7349)	0.7252 (0.7658)	0.2890
EDP5	0.6853 (0.6330)	0.8035 (0.7706)	0.5920 (0.5853)	0.7347 (0.7377)	0.7024 (0.7002)	0.1249

Q*: "Tell me and I forget, teach me and I remember, involve me and I learn." [96]

Table 5
Comparison of Pro-Explora output and Copilot output.

	Source	Generated Design Concept	Virtual Concept Evaluation		
			ChatGPT	DeepSeek	Grok
1	Copilot	Smart Irrigation System for Efficient Water Management	Human	Human	AI
	Pro-Explora	Developing Advanced Downhole Technologies for Efficient Mixed-Flow Irrigation Systems	AI	AI	Human
2	Pro-Explora	Development of a Heat-Resistant Robotic Fire Extinguishing System	Human	AI	Human
	Copilot	Renewable Energy-Powered Autonomous Drone for Surveillance	AI	Human	AI
3	Copilot	Design of a Portable Solar Charging Station	Human	Human	Human
	Pro-Explora	Innovative Design for Sustainable Energy Systems	AI	AI	AI
4	Pro-Explora	Design of a Portable Cyclone Separator for Atmospheric Water Extraction	AI	AI	Human
	Copilot	Automated Greenhouse Monitoring and Control System	Human	Human	AI
5	Copilot	IoT-Based Smart Home Energy Management System	Human	AI	AI
	Pro-Explora	Compact Robotic System for Automated Glass Surface Cleaning	AI	Human	Human
6	Pro-Explora	Developing a Self-Balancing Support Mechanism for Elderly Care Facilities	AI	AI	Human
	Copilot	Development of a Wearable Health Monitoring Device	Human	Human	AI
7	Copilot	Design and Implementation of a Smart Traffic Light System	Human	Human	AI
	Pro-Explora	Intelligent Vehicle Safety System for Accident Prevention	AI	AI	Human
8	Pro-Explora	Development of an Autonomous Energy-Generating Vehicle	AI	AI	AI
	Copilot	Energy-Efficient HVAC System for Sustainable Buildings	Human	Human	Human
9	Copilot	Robotic Arm for Precision Manufacturing Applications	AI	AI	AI
	Pro-Explora	Adjustable-Height Manual Metal Bending Machine Design	Human	Human	Human
10	Pro-Explora	Designing an Accessible Vehicle for Safe Patient Transfer	Human	AI	Human
	Copilot	Wireless Sensor Network for Environmental Monitoring	AI	Human	AI

their cosine similarities to EDP1–EDP5 shown in Table 4 without parentheses. EDP6–EDP10 are reworded using the GPT-3.5 model and indicated as EDP6'–EDP10' in Table 4, with their cosine similarities to EDP1–EDP5 shown in parentheses. Additionally, the similarity between the naturally framed EDPs and a non-EDP, represented by a quote (Q*), is included for further insight and interpretation.

4.4. SGM and Copilot outputs comparison with ChatGPT, DeepSeek, and Grok

As presented in Table 5, Pro-Explora was compared with ChatGPT, DeepSeek, and Grok for early-stage engineering design applications. Copilot was prompted to generate ten EDPs. Similarly, Pro-Explora, powered by the SGM depicted in Fig. 4, generated ten EDPs. The EDPs generated by Copilot and Pro-Explora were paired and presented in Table 5. Each pair of EDPs was then evaluated by Chatgpt, DeepSeek, and Grok with the prompt: "Identify which of the following topics is AI-generated and which is human-generated." Note that none of the EDPs were human-generated; the prompt was designed to prevent bias. To avoid the AI tools detecting a pattern in the EDPs, the order of presentation of the EDPs generated by Copilot and Pro-Explora was alternated. As discussed in Section 5, an SGM could be a cost-effective approach for certain applications.

5. Discussion

5.1. Uniqueness of the GMC model outputs

A generative model is characterised by its output variability or deterministic chaos while following logical rules. The box plot in Fig. 8 illustrates the variability in the EDPs generated by the GMC model during the case study reported in Section 3.1. As shown in Fig. 8, there are few or no overlaps in the median and IQR of the TF–IDF, indicating variability in the EDPs. The TF–IDF values of the five EDPs across 43 participants show an average IQR of 0.28 (range: 0.01–0.65, spread: 0.64), suggesting moderate to high variability and significant data dispersion. The average range of TF–IDF values is 0.63 (range: 0.24–1.14, spread: 0.90), also indicating variability [97]. The variability in the outputs of the GMC model is also relative to the input data. A generative model is non-deterministic, meaning the input and output data are distinct. The GMC model in Fig. 2 includes a duplication recognition search to ensure that a generated EDP has no duplicates in the database, maintaining the variability of the output data relative to the input data.

5.2. Comparing the performances of the GMC model and SGM

The results in Table 3 demonstrate a statistically significant improvement in the readability of outputs generated by the SGM compared to the GMC model. Specifically, rewording GMC model outputs resulted in metric score differences with p-values less than 0.05 across all metrics. The range of FRE scores for the reworded EDPs is approximately 15–43, an improvement from the previous range of 0–33. The improved score range is considered adequate for design engineers (both novice and experienced), college graduates, and college students [91,98]. Based on the standards in Table 2, higher FRE scores and lower FGL and GFI scores indicate improved readability. This condition is met in most cases in Table 3. However, for Groups 2, 3, 7, and 10, some FGL and GFI metric scores of the SGM outputs are slightly above or equal to those of the GMC model.

Although not considered in the interpretation of metric scores in Table 2, the negative FRE score for the GMC model in Group 1 can be treated as zero [99]. Based on Eq. 4, a high ASL or ASW value can result in a negative FRE score, indicating more difficult readability. The optimisation of the GMC model outputs slightly increases the number of words in each EDP. Table 3 shows that the maximum difference in word count between the GMC model and the SGM occurs in Group 5 (153 vs. 103, a difference of 32 words). Given that there are at least 15 sentences in the merged EDPs, this difference translates to an average of approximately two words per sentence, suggesting minimal variation in sentence lengths after rewording.

5.3. The appropriateness of grouping the participants' EDPs for evaluation

The grouped EDPs in Table 3 comprise a minimum of 15 sentences and 102 words each. These groups serve as a basis for evaluating the readability of individual EDPs generated by the GMC model and SGM. Although the grouped EDPs may exhibit some disconnection in flow, this does not undermine the validity of the readability assessments conducted in this study. Eqs. 4–6 address parameters such as ASL, ASW, and the number of hard words [100,101], without factoring in text coherence [101]. Consequently, the readability scores for the grouped EDPs apply to individual EDPs, providing insights into the ease with which the text can be read and comprehended [93].

5.4. Comparing the similarities of the GMC model and SGM outputs to human output

The cosine assessment results in Table 4 indicate that the natural context of EDPs generated by the GMC model is preserved following optimisation with a GPT model. The similarity scores for GMC model EDPs and SGM outputs are consistently close across all comparisons with EDPs (EDP1–EDP5) created by design engineers. The trained Python model used for analysis can interpret texts with varying degrees of similarity. This is demonstrated by comparing EDPs framed by design engineers with a quote (Q^*), as shown in Table 4. The similarity scores for the quote (non-EDP) range from 12 % to 36 %, whereas the scores for the GMC model and SGM range from 54 % to 93 %, closely aligning with a "natural" EDP. The use of a low temperature parameter, as detailed in Optimisation 1 (Section 3.2), facilitated this natural alignment. Therefore, optimising the GMC model with a GPT-3.5 model significantly enhances the GMC model's output while preserving its natural context.

5.5. Evaluating SGM and Copilot outputs with ChatGPT, DeepSeek, and Grok

As presented in Section 4.4, the EDPs generated by Pro-Explora, an SGM, are comparable to those generated by Copilot, an LLM. To prevent bias, the two categories of EDPs were presented to the LLM AI tools—ChatGPT, DeepSeek, and Grok—with a prompt to believe that one was AI-generated and the other human-generated. For each prompt across the AI tools, the order (Copilot first and Pro-Explora second) of the EDPs was alternated to prevent the AI tools from detecting a pattern in the order of presentation. As the evaluation results in Table 5 show, some EDPs generated by Copilot and Pro-Explora were equally evaluated as human-generated. This highlights that in certain applications, an SGM might be a more cost-effective option. It should be noted that the evaluation of a concept as human-generated by the AI tools does not necessarily mean that it is better than the AI-generated concept. Based on the reasons provided by the AI tools, some concepts were judged to be too advanced to be generated by a human.

Beyond cost-effectiveness, the SGM proposed in this study addresses the growing privacy concerns with LLMs [102]. The outputs from the GMC model in Fig. 2 are not traceable to the original data used to generate them. This serves as a form of encryption for the data before they are passed to the GPT model for refinement, as the refinement process could cause data logging at the LLM service provider's end. Hence, the GMC model can be combined with an LLM for text refinement. However, for higher privacy, a small language model could be trained specifically for text refinement and incorporated into the SGM (Section 3.2), eliminating any communication with the cloud-based LLM. This training could be one-off without repeated retraining, ensuring that confidential data supporting design concept generation would not be passed to an LLM.

A significant amount of digital data may still not be in the public domain due to privacy concerns. These data could be usefully applied by developing an SGM, helping businesses derive value from their data without sacrificing privacy or confidentiality. The future research challenge would be to identify more methods of developing an appropriate SGM to suit a given application. For example, the SGM used in this study is applicable in the medical field for drug discovery. It could also be applied in the music industry for generating new lyrics without compromising the data logging of confidential lyrics to an LLM.

5.6. Practical and theoretical implications

5.6.1. Practical implications

This study has direct and practical implications for engineering design practitioners. The GMC model supports design engineers in exploring, conceptualising, and identifying new EDPs whose solutions would be of societal relevance. When a design engineer observes or perceives an EDP, there is a rapid shift in thought towards possible solution concepts for the EDP. Feedback from case study participants suggests this. For example, a participant remarked, "It is great to have some help, the thin[g] I liked the most is that I could feel how mi [sic] mind started to get inspired and thinking of solving some of the problems generated by Pro-Explora." Hence, the GMC model is further optimised to incorporate an aspect of solution concept exploration for its framed EDPs, leveraging the GPT-3 image model. The SGM presented in this study provides comprehensive support to design engineers at the early-stage design. Current AI tools, such as ChatGPT, Copilot, DeepSeek, and Grok, are not specifically created to support early-stage engineering design. There remains a need to create SGMs for engineering design applications, as done in this study.

An experimental test in this study shows that two solution concepts are consistently returned when the GPT-3 image model is prompted with a framed EDP. All the returned images contain features related to the idea expressed in the EDP prompt (see Fig. 7), corroborating study findings about the image model as presented in Section 2.2. The intent of the SGM in this study is to specifically support the challenging process of conceptualising and identifying new EDPs. Some of the participants' feedback highlights the need for this sort of specialised support. Remarks include, "This is a very interesting topic, I have seen AI-based design solutions before—but never AI problem identification," "A wide spectrum of engineering design problems," and "amazing and refreshing as a design engineer to reignite our thought process."

In about 10 min, based on their experiences, participants used Pro-Explora to frame at least five useful EDPs and rated them highly, con-

sidering the usefulness of their solutions to society. Whilst some of the EDPs framed may have a related existing solution, it prompts design engineers to think freely, divergently, and stochastically about new EDPs they could solve. It helps design engineers break the mental constriction in initiating radical and inventive thinking. Initiating such thinking naturally without support is challenging [17]. The process could lead to cognitive fatigue and other associated challenges, as presented in Section 2.1. Using the SGM through the Pro-Explora tool also eliminates the need for expertise in prompt engineering required for the effective use of AI tools. Pro-Explora does not require knowledge of prompt engineering, as users just need to click to achieve the same result as using prompts. The prompting is effectively handled by the backend for users, simplifying the process of using the power of GPT.

5.6.2. Theoretical implications

Besides the practical implications, this study also has theoretical implications, providing opportunities for further research. For example, identifying design problems computationally was previously thought to be impossible [103]. As the first comprehensive model and tool to support early-stage design activities, this study offers insights that can inspire further research. In sequence generation, the SGM presented in this study demonstrates that a GMC model can be optimised to match the capability of industrial LLMs without data training. A GPT model is used in this study to optimise the GMC model, and the constraints that limit the use of other alternatives/models are highlighted (Section 3.2). Research into reducing or eliminating these constraints would create more optimisation alternatives for the GMC model, increasing its applicability in various areas.

The synergy of the GMC and GPT models in this study produced an SGM with greater applicability in supporting early-stage design activities than the separate models. With relatively fewer data [15], the GMC model can generate useful concepts that only need modest rewording to improve their readability, as indicated by the results in Tables 3 and 4. It should be noted that a generative model could “hallucinate” or generate nonsensical text [104]. For example, the GPT-3.5 model that powers text conversations in ChatGPT sometimes writes “plausible-sounding but incorrect or nonsensical answers” [58]. However, although the SGM can sometimes produce incoherent text, this sort of combination is necessary to discover uncommon concepts for invention or radical innovation [105].

6. Conclusion

This study introduces the SGM as a cost-effective approach to utilising specific capabilities of LLMs for tailored design applications. We combined a GMC model with a GPT model to create an SGM to support early-stage design activities. The GMC model, integrated within the Pro-Explora tool, was verified in a case study to frame unique EDPs deemed useful by both experienced and novice design engineers. Guided by case study feedback, we utilised a GPT text model and a GPT image model to optimise the GMC model, resulting in the SGM. The GPT text model enhances the readability of the GMC model-framed EDPs, while the image model aids in exploring solution concepts for these EDPs. Standard readability metrics and cosine similarity were employed to evaluate the performance of the SGM. The results indicate a statistically significant readability improvement for the SGM compared to the GMC model alone. Additionally, the model demonstrates comparable output with Copilot for EDP generation when evaluated using ChatGPT, DeepSeek, and Grok. For the intended SGM application, the model offers superior data privacy compared to AI tools that directly use LLMs, mitigating the issue of data logging. The SGM in this study enables design engineers to rapidly generate comprehensive design concepts within 10 min, a feat that would be challenging without support due to cognitive fatigue and other associated challenges.

Practically, this study contributes to the research field by addressing the knowledge gap in computationally identifying useful EDPs, previously considered impossible. This advancement facilitates rapid concept generation at the early-stage engineering design, accelerating innovation and invention. Theoretically, this study opens new avenues for research. It demonstrates that the GMC model can be optimised to match the capabilities of LLMs without extensive data training. Although a GPT model was used in this study to optimise the GMC model, the constraints limiting the use of other models are highlighted to inspire further research into improving these alternatives.

This study is limited by its reliance on GPT models offered for a fee. Future research will focus on training an SLM specifically for the SGM, potentially involving the development of a bespoke text rewording model from scratch.

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.

CRediT authorship contribution statement

Chijioke C. Obieke: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jelena Milisavljevic-Syed:** Supervision, Conceptualization. **Pingfei Jiang:** Investigation, Conceptualization. **John Bridgeman:** Supervision, Conceptualization. **Ji Han:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization.

References

[1] E. Whelan, S. Parise, J. De Valk, R. Aalbers, Creating employee networks that deliver open innovation, MIT Sloan Manag. Rev. 53 (2011) 37.
[2] S.P. Green, Plagiarism, norms, and the limits of theft law: some observations on the use of criminal sanctions in enforcing intellectual property rights, Hastings LJ 54 (2002) 167.
[3] L.S. Young, T.L. Harper, Innovation versus invention: serendipity, chance, and experience favoring prepared minds, Accel. New Food Prod. Des. Dev. (2017) 117–126.
[4] D. Bajaj, A. Goel, S. Gupta, H. Batra, MUCE: a multilingual use case model extractor using GPT-3, Int. J. Inf. Technol. 14 (2022) 1543–1554.
[5] J. Zhou, J.D. Camba, The status, evolution, and future challenges of multimodal large language models (LLMs) in parametric CAD, Expert Syst. Appl. (2025) 127520.
[6] Special Issue, Large language models in design and manufacturing, J. Comput. Inf. Sci. Eng. (2025) 25.
[7] C.C. Obieke, J. Bridgeman, J. Han, A framework of AI collaboration in engineering design (AICED), Proc. Des. Soc. 5 (2025) 91–100.
[8] L.M. Schulze Buschoff, E. Akata, M. Bethge, E. Schulz, Visual cognition in multimodal large language models, Nat. Mach. Intell. (2025) 1–11.
[9] L. Wang, Z. Wan, C. Ni, Q. Song, Y. Li, E. Clayton, B. Malin, Z. Yin, Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review, J. Med. Internet Res. 26 (2024) e22769.
[10] P. Thomas, D. Nicholas, The fourth industrial revolution: shaping new era, J. Int. Aff. 72 (2018) 17–22.
[11] F. Liu, T. Jin, J.S. Lee, Automatic readability assessment for sentences: neural, hybrid and large language models, Lang. Resour. Eval. (2025) 1–32.
[12] K. Haynes, H. Shirazi, I. Ray, Lightweight URL-based phishing detection using natural language processing transformers for mobile devices, Procedia Comput. Sci. 191 (2021) 127–134.
[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) 30.
[14] A. Elfeki, M. Dekking, A Markov chain model for subsurface characterization: theory and applications, Math. Geol. 33 (2001) 569–589.
[15] C.C. Obieke, J. Milisavljevic-Syed, A. Silva, J. Han, A computational approach to identifying engineering design problems, J. Mech. Des. 145 (2023) 041406.
[16] A. Darwish, Optimality in Translation, Writescope Publishers, 2008.
[17] C.C. Obieke, J. Milisavljevic-Syed, J. Han, A study of the early-stage engineering design activities in practice, Proc. Des. Soc. 3 (2023) 405–414.
[18] E. Belanger, J. Moller, J. She, Challenges to engineering design teamwork in a remote learning environment, Educ. Sci. 12 (2022) 741.
[19] G. Molecke, A.-C. Pache, How do we know when social innovation works? A review and contingencymodel of social impact assessment, Handbook of Inclusive Innovation, (2019) 83–105.
[20] A. Olewnik, R. Yerrick, A. Simmons, Y. Lee, B. Stuhlmiller, Defining open-ended problem solving through problem typology framework, Int. J. Eng. Pedagogy (IJEP) (2020) 10.

[21] J. Vale, K. Gordon, R. Kirkscey, J. Hill, Student and faculty perceptions of capstone purposes: what can engineering learn from other disciplines? in: Proceedings of the Canadian Engineering Education Association (CEEA), 2020.
[22] U.S. NAE - National Academy of EngineeringThe Engineer of 2020: Visions of Engineering in the New Century, National Academies Press, Washington, DC, 2004.
[23] U. Jorgensen, H. Lindegaard, S. Brodersen, P. Boelskifte, Foundations for a new type of design-engineers, in: DS 68-8: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 8: Design Education, Lyngby/Copenhagen, Denmark, 2011, pp. 275–286. 15.-19.08. 2011.
[24] K. Chuah, The 5-Second Inventor, Dog Ear Publishing, Indianapolis, 2009.
[25] UKIPO - Intellectual Property Office UK, Intellectual property basics, Crown copyright, Newport, 2021.
[26] S. Dirksen, A reconsideration of the physicians' Immunity statute, Duke L. Tech. Rev. 1 (2001) 1.
[27] M.G. Sipe, Patent Law's philosophical fault line, Wis. L. Rev. (2019) 1033.
[28] H. Zhang, L. Estabrooks, A. Perry, Bringing invention education into middle school science classrooms: a case study, Technol. Innov. 20 (2019) 235–250.
[29] UKIPO - Intellectual Property Office UK, IP Basics, Crown Copyright, Newport, 2017.
[30] NIPO - Norwegian Industrial Property Office, what can be patented? Patentstyret (2022).
[31] N. Jayaratna, B. Wood, Every innovation is a new thing but not every new thing is an innovation, Rev. Int. Proj. (2008) 13–25.
[32] M. Sun, M. Wang, R. Wegerif, J. Peng, How do students generate ideas together in scientific creativity tasks through computer-based mind mapping? Comput. Educ. 176 (2022) 104359.
[33] J. Van Cutsem, S. Marcora, K. De Pauw, S. Bailey, R. Meeusen, B. Roelands, The effects of mental fatigue on physical performance: a systematic review, Sports Med. 47 (2017) 1569–1588.
[34] M.R. Smith, A.J. Coutts, M. Merlini, D. Deprez, M. Lenoir, S.M. Marcora, Mental fatigue impairs soccer-specific physical and technical performance, Med. Sci. Sports Exerc. 48 (2016) 267–276.
[35] P.J. Hurley, Reconceptualizing ego depletion as transient cognitive fatigue, available at SSRN 3797263. https://doi.org/10.31234/osf.io/sy2mw, (2021).
[36] P. Stanton, Conscious Creativity: Look, Connect, Create, Leaping Hare Press, 2018.
[37] A. Ariza-Montes, J.M. Arjona-Fuentes, H. Han, R. Law, Employee responsibility and basic human values in the hospitality sector, Int. J. Hosp. Manag. 62 (2017) 78–87.
[38] S.M. Marcora, W. Staiano, V. Manning, Mental fatigue impairs physical performance in humans, J. Appl. Physiol. 106 (2009) 857–864.
[39] S. Southern, S. Domzalski, in: Developing Intuition: The Key to Creative Futures Research, Annual Meeting of the American Educational Research AssociationNew Orleans, Louisiana, 1984, p. 105.
[40] M. Bernal, J.R. Haymaker, C. Eastman, On the role of computational support for designers in action, Des. Stud. 41 (2015) 163–182.
[41] H.A. Simon, Problem forming, problem finding and problem solving in design, design and systems: general applications of methodology, transaction Publishers, N. B., 1995, pp. 245–257.
[42] T. Howard, S.J. Culley, E. Dekoninck, in: Creativity in the Engineering Design Process, DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 2007 28.-31.07. 2007pp. 329-330 (exec. Summ.), full paper no. DS342_P_493.
[43] S.R. Daly, E.A. Mosyjowski, C.M. Seifert, Teaching creativity in engineering courses, J. Eng. Educ. 103 (2014) 417–449.
[44] B. Shneiderman, Creativity support tools: accelerating discovery and innovation, Commun. ACM 50 (2007) 20–32.
[45] J. Han, F. Shi, L. Chen, P.R.N. Childs, The Combinator – a computer-based tool for creative idea generation based on a simulation approach, Des. Sci. 4 (2018) e11.
[46] J. Kim, M.L. Maher, The effect of AI-based inspiration on human design ideation, Int. J. Des. Creat. Innov. (2023) 1–18.
[47] A. Chakrabarti, P. Sarkar, B. Leelavathamma, B. Nataraju, A functional representation for aiding biomimetic and artificial inspiration of new ideas, Ai Edam 19 (2005) 113–132.
[48] J.S. Linsey, J. Laux, E. Clauss, K.L. Wood, A.B. Markman, in: Increasing Innovation: A Trilogy of Experiments Towards a Design-by-Analogy Method, International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2007, pp. 145–159.
[49] X. Yan, Y. Xiao, Y. Jin, Generative large language models explained [ai-explained], IEEE Comput. Intell. Mag. 19 (2024) 45–46.
[50] R. Al-Jarf, in: Translation of Medical Terms by AI: A Comparative Linguistic Study of Microsoft Copilot and Google Translate, International Conference on Artificial Intelligence and its Applications in the Age of Digital Transformation, Springer, 2024, pp. 220–235.
[51] M. Kamran, M. Faizan, S. Wang, B. Han, W.-Y. Wang, Generative AI and prompt engineering: transforming rockburst prediction in underground construction, buildings, 15 (2025) 1281.
[52] A. Grzybowski, K. Pawlikowska–Łagód, W.C. Lambert, A history of artificial intelligence, Clin. Dermatol. 42 (2024) 221–229.
[53] J.D. Weisz, M. Muller, J. He, S. Houde, Toward general design principles for generative AI applications, arXiv preprint arXiv:2301.05578, (2023).
[54] G. Brockman, I. Sutskever, OpenAI, Introducing OpenAI, 2015.
[55] G. Marcus, E. Davis, S. Aaronson, A very preliminary analysis of dall-e 2, arXiv preprint arXiv:2204.13807, (2022).
[56] L. Chen, L. Sun, J. Han, A comparison study of human and machine-generated creativity, J. Comput. Inf. Sci. Eng. 23 (2023) 051012.
[57] S. Milano, J.A. McGrane, S. Leonelli, Large language models challenge the future of higher education, Nat. Mach. Intell. (2023) 1–2.
[58] A. Glynn, Suspected undeclared use of artificial intelligence in the academic literature: an analysis of the Academ-AI dataset arXiv preprint arXiv:2411.15218, Nov 20, 2024.
[59] G. Caroppi, K. Västilä, P. Gualtieri, J. Järvelä, M. Giugni, P.M. Rowiński, Acoustic Doppler velocimetry (ADV) data on flow-vegetation interaction with natural-like and rigid model plants in hydraulic flumes, Data Br. 32 (2020) 106080.
[60] M.A. Jesson, J. Bridgeman, M. Sterling, Novel software developments for the automated post-processing of high volumes of velocity time-series, Adv. Eng. Softw. 89 (2015) 36–42.
[61] V. Gandhi, Brain-computer interfacing for assistive robotics: electroencephalograms, recurrent quantum neural networks, and user-centric graphical interfaces, academic press 2014.
[62] J. Hou, S. Xu, Near-real-time seismic human fatality information retrieval from social media with few-shot large-language models, in: Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022, pp. 1141–1147.
[63] L.D. Manocchio, S. Layeghy, W.W. Lo, G.K. Kulatilleke, M. Sarhan, M. Portmann, FlowTransformer: a transformer framework for flow-based network intrusion detection systems, arXiv preprint arXiv:2304.14746, (2023).
[64] A.J.G. Sison, M.T. Daza, R. Gozalo-Brizuela, E.C. Garrido-Merchán, ChatGPT: more than a weapon of mass deception, ethical challenges and responses from the human-centered artificial intelligence (HCAI) perspective, arXiv preprint arXiv:2304.11215, (2023).
[65] M. Gupta, P. Agrawal, Compression of deep learning models for text: a survey, ACM Trans. Knowl. Discov. Data (TKDD) 16 (2022) 1–55.
[66] G. Schurz, in: Kinds of Unpredictability in Deterministic Systems, Law and Prediction in the Light of Chaos Research, Springer, 1996, pp. 123–141.
[67] S. Dube, in: Unleashing the Power of Generation, An Intuitive Exploration of Artificial Intelligence: Theory and Applications of Deep Learning, Springer, 2021, pp. 121–139.
[68] H. Anandakumar, R. Arulmurugan, C.C. Onn, Big data analytics for sustainable computing, Mob. Netw. Appl. 24 (2019) 1751–1754.
[69] J.S. Johnson, S.B. Friend, H.S. Lee, Big data facilitation, utilization, and monetization: exploring the 3Vs in a new product development process, J. Prod. Innov. Manag. 34 (2017) 640–658.
[70] C. Obieke, J. Milisavljevic-Syed, J. Han, Supporting design problem-exploring with emergent technologies, Procedia CIRP 91 (2020) 373–381.
[71] C.C. Obieke, J. Milisavljevic-Syed, J. Han, Data-driven creativity: computational problem-exploring in engineering design, Proc. Des. Soc. 1 (2021) 831–840.
[72] A. Jalilifard, V.F. Caridá, A.F. Mansano, R.S. Cristo, F.P.C. da Fonseca, Semantic sensitive TF-IDF to determine word relevance in documents, advances in computing and network communications: proceedings of CoCoNet 2020, Volume 2, Springer 2021, pp. 327–337.
[73] W. Zhang, T. Yoshida, X. Tang, A comparative study of TF* IDF, LSI and multi-words for text classification, Expert Syst. Appl. 38 (2011) 2758–2765.
[74] T. Beysolow, Applied natural language processing with Python : implementing machine learning and deep learning algorithms for natural language processing /Taweh Beysolow II, Apress (2018).
[75] M.K. Anand Deshpande, Artificial Intelligence for Big Data: Complete Guide to Automating Big Data Solutions Using Artificial Intelligence Techniques, Packt Publishing, Birmingham, 2018.
[76] T. Basit, Manual or electronic? The role of coding in qualitative data analysis, Educ. Res. 45 (2003) 143–154.
[77] K.G. Isangula, S. Kelly, J. Wamoyi, Manual qualitative data coding using MS word for students and early career researchers in resource-constrained settings, Int. J. Qual. Methods 23 (2024) 16094069241299223.
[78] C.A. Langford, P.F. Pearce, Increasing visibility for your work: the importance of a well-written title, Am. Assoc. Nurse Pract. 31 (2019) 217–218.
[79] Y.F. Greenspan, A guide to teaching elementary science: ten easy steps, Brill. Sense (2016).
[80] N. Triki, Elaboration paradigms in PhD theses introductions, Deviation (s), (2014) 202–225.
[81] M. Haghverdi, L.R. Wiest, The effect of contextual and conceptual rewording on mathematical problem-solving performance, Math. Educ. (2016) 25.
[82] M. Zong, B. Krishnamachari, A survey on GPT-3, arXiv preprint arXiv:2212.00857, (2022).
[83] J.G. Michaeli, G. Hou, X. Hu, M. Hou, Initial investigation of analytic hierarchy process to teach creativity in design and engineering, (2014).
[84] R. Sreemathy, M. Turuk, S. Chaudhary, K. Lavate, A. Ushire, S. Khurana, Continuous word level sign language recognition using an expert system based on machine learning, Int. J. Cogn. Comput. Eng. 4 (2023) 170–178.
[85] J. Han, F. Shi, L. Chen, P. Childs, The analogy retriever–an idea generation tool, in: DS 87-4 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 4: Design Methods and Tools, Vancouver, Canada, 2017, p. 2017. 21-25.08011-020.
[86] S. Sarica, J. Han, J. Luo, Design representation as semantic networks, Comput. Ind. 144 (2023) 103791.
[87] R. Speer, C. Havasi, ConceptNet 5: a large semantic network for relational knowledge, in: The People's Web Meets NLP: Collaboratively Constructed Language Resources, 2013, pp. 161–176.
[88] L.A. Wood, O. Agbonlahor, M.M. Tomlinson, S. Kerstiens, K. Vincent, A.C. McLeish, K.L. Walker, J.L. Hart, Readability of online e-cigarette cessation information, Tob. Induc. Dis. (2022) 20.

[89] F. Murtaza, L. Shirreff, L.N. Huang, M. Jacobson, R. Jarcevic, M.K. Christakis, Quality and readability of online health information on menopausal hormone therapy in Canada: what are our patients reading? Menopause 29 (2022) 54–62.
[90] C. Bonner, P. Patel, M.A. Fajardo, R. Zhuang, L. Trevena, Online decision aids for primary cardiovascular disease prevention: systematic search, evaluation of quality and suitability for low health literacy patients, BMJ Open 9 (2019) e025173.
[91] O. Ahmadi, A. Wood, The readability and reliability of online information about adenoidectomy, J. Laryngol. Otol. 135 (2021) 976–980.
[92] A. Kulkarni, A. Shivananda, Natural language processing recipes : unlocking text data with machine learning and deep learning using Python /Akshay Kulkarni, Adarsha Shivananda (2021).
[93] S. Štajner, R. Evans, C. Orasan, R. Mitkov, What can readability measures really tell us about text complexity, Proceedings of workshop on natural language processing for improving textual accessibility, Citeseer, 2012, pp. 14–22.
[94] J. DaCosta, D. Bhatia, O. Crothers, J. Talks, Utilisation of optical coherence tomography and optical coherence tomography angiography to assess retinal neovascularisation in diabetic retinopathy, Eye 36 (2022) 827–834.
[95] C. Ferguson, M. Merga, S. Winn, Communications in the time of a pandemic: the readability of documents for public consumption, Aust. N. Z. J. Public Health 45 (2021) 116–121.
[96] M.E. Hoque, Tell me and I forget; teach me and I remember; involve me and I learn: navigation the way of teaching, EDRC J. Learn. Teach. 7 (2021) 30–42.
[97] D.S. Moore, G.P. McCabe, B.A. Craig, Introduction to the Practice of Statistics, Macmillan Higher Education, 2014.
[98] K.M. Davidson, C.J. Espie, C. Lammie, Conducting randomised controlled trials: finding better ways to explain research to people with anti-social personality disorder who have low literacy levels, Crim. Behav. Ment. Health 21 (2011) 265–278.
[99] H.Y. Baek, D.D. Cho, R.A. Jordan, E. Kuvvet, The differential effect of social disclosure on loan funding and loan repayment: evidence from fixed-rate peer-to-peer lending, Manag. Finance (2020).
[100] M. Smith, R. Taffler, Readability and understandability: different measures of the textual complexity of accounting narrative, Account. Audit. Account. J. 5 (1992) 0-0.
[101] B. Bruce, A. Rubin, K. Starr, Why readability formulas fail, IEEE Trans. Prof. Commun. (1981) 50–52.
[102] H. Kibriya, W.Z. Khan, A. Siddiqa, M.K. Khan, Privacy issues in large language models: a survey, Comput. Electr. Eng. 120 (2024) 109698.
[103] Y. Celik, Rules for doing and managing research project, 2019.
[104] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y.J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, ACM Comput. Surv. 55 (2023) 1–38.
[105] M.D. Cohen, J.G. March, J.P. Olsen, A garbage can model of organizational choice, Adm. Sci. Q. (1972) 1–25.

https://dspace.org/

Trailblazing specific generative models (SGMs) for early-stage engineering design concepts

Obieke, Chijioke C.; Milisavljevic-Syed, Jelena; Jiang, Pingfei; Bridgeman, John; Han, Ji

2025-09
Elsevier
https://doi.org/10.1016/j.daai.2025.100030
Obieke CC, Milisavljevic-Syed J, Jiang P, et al., (2025) Trailblazing specific generative models (SGMs) for early-stage engineering design concepts. Design and Artificial Intelligence, Volume 1, Issue 3, September 2025, Article number 100030

Downloaded from a DSpace repository. https://dspace.org/