A Computational Approach to Identifying Engineering Design Problems

Chijioke C. Obieke, first author
Systems Realization Laboratory, Division of Industrial Design, School of Engineering
University of Liverpool,
Brownlow Hill, Liverpool L69 3GH, United Kingdom
chijioke.obieke@liverpool.ac.uk

Jelena Milisavljevic-Syed, second author
Systems Realization Laboratory, Division of Industrial Design, School of Engineering
University of Liverpool,
Brownlow Hill, Liverpool L69 3GH, United Kingdom
j.milisavljevic-syed@liverpool.ac.uk

Sustainable Manufacturing Systems Centre
School of Aerospace, Transportation and Manufacturing (SATM)
Cranfield University, Cranfield, Bedfordshire MK43 0AL, United Kingdom
j.milisavljevicsyed@cranfield.ac.uk
ASME Membership No. 000102058626

Arlindo Silva, third author
Singapore University of Technology and Design
8 Somapah Rd, Singapore 487372
arlindo_silva@sutd.edu.sg
ASME Membership No. 102035333

Ji Han, fourth author[1]
INDEX, Business School,
University of Exeter
Tintagel House, 92 Albert Embankment, London SE1 7TY, United Kingdom
j.han2@exeter.ac.uk

[1] Corresponding author.
Ji Han, University of Exeter
j.han2@exeter.ac.uk

ABSTRACT

Identifying new problems and providing solutions are necessary tasks for design engineers at early-stage product design and development. A new problem fosters innovative and inventive solutions. Hence, it is expected that engineering design pedagogy and practice should equally focus on Engineering Design Problem-Exploring (EDPE) – a process of identifying or coming up with a new problem or need at the early stage of design, and Engineering Design Problem-Solving (EDPS) – a process of developing engineering design solutions to a given problem. However, studies suggest that EDPE is scarcely practiced or given attention to in academia and industry, unlike EDPS. The aim of this paper is to investigate the EDPE process for any information relating to its scarce practice in academia and industry. This is to explore how emerging technologies could support the process. Natural models and phenomena that explain the EDPE process are investigated, including the "rational" and "garbage can" models, and associated challenges identified. A computational framework that mimics the natural EDPE process is presented. The framework is based on Markovian model and computational technologies, including machine learning. A case study is conducted with a sample size of 43 participants drawn worldwide from the engineering design community in academia and industry. The case study result shows that the first-of-its-kind computational EDPE framework presented in this paper supports both novice and experienced design engineers in EDPE.

Keywords: Artificial intelligence; Computer-Aided Design; Conceptual Design; Creativity and Concept Generation; Data-Driven Design.

1 INTRODUCTION

One of the main tasks of a design engineer at early-stage product design and development is to provide an Engineering Design Solution (EDS) to a societal problem using personal knowledge, experience, and background [1, 2]. Another main task is identifying or conceptualising a new Engineering Design Problem (EDP) [3, 4]. The EDP would deliver societal values once it is solved. Many societal problems, such as “created” and “discovered” problems, are elusive and could remain elusive until discovered by a design engineer [5]. A “created” EDP is a problem that remains unknown or would not exist until conceptualized and translated by an engineer to make it an apparent EDP. A “created” EDP does not have a known formulation, method of solution, or solution. A “discovered” EDP is a type of problem that exists, identified by a design engineer, and may or may not have a known formulation, method of solution, or solution [6, 7]. An EDP triggers ideas for inventive or innovative solutions [8]. The value of an EDS depends on the type of EDP it solves. Hence, both coming up with an EDP and providing an EDS are equally important and are standard expectations in engineering design [9, 10].

Therefore, focus on the processes leading to an EDP and EDS is expected in engineering design academia and industry. However, this seems not to be the case. Studies suggest that over the years, the focus is mainly on Engineering Design Problem-Solving (EDPS) – the process of producing an EDS [11, 12]. The process of coming up with an EDP or a need at early stages in design [13], referred to as Engineering Design Problem-Exploring (EDPE) in this research, is rarely practiced or discussed in the literature [14-18]. This lack of attention on EDPE over the years has increasing consequences, including 1) limiting the comprehensive capability within engineering design, 2) delayed discovery, innovation, and invention, 3) less inventive solutions/products, 4) lack of development of a specific support tool for EDPE, 5) lack of creativity assessment standard and rewards for EDPE, and 6) decline in EDPE skills within the engineering design community [19-22]. Research interests in recent times suggest a need for a strong focus on EDPE. The solution-first approach in engineering design is attracting research interests and seems to be an approach that would necessitate computational facilitation in EDPE. In this approach, an EDS is created first in anticipation of a yet-to-be-identified EDP [27]. This is distinct from the widely known problem-first approach, where an EDS is sought for an EDP. The solution-first approach implies a greater need for a process to facilitate new EDPs that match EDS created beforehand. Albeit studies on the scarce attention and practice of EDPE are few, a study on its determinants, consequences, and mitigation lacks. A study that provides empirical evidence in support of reports on the scarce focus on EDPE also lacks.

In this paper, the aim is to investigate the EDPE process and practice for information relating to its scarce practice. This would facilitate possible interventions to encourage EDPE practice in academia and industry. Albeit behavioral interventions like formal teaching/training skills and reward systems for EDPE is possible, there are scholarly indications that EDPE involves heavy thinking and is a difficult, challenging, and time-consuming task [23-26]. Hence, special attention in this paper is on EDPE process activities with natural limitations that necessitate computational support. The models and phenomena related to the natural EDPE process are investigated including the “rational” and “garbage can” models, serendipity, and apophenia phenomena. This contributes to coming up with a first-of-its-kind computational approach and tool presented in this paper that mimic and support the natural EDPE process. To mimic the natural EDPE approach, a Markovian model (MM) is used in synergy with computational technologies, including data mining, machine learning (ML), natural language processing (NLP), duplication recognition, and python programming language. A case study is conducted in three parts with a sample size of 43 participants drawn worldwide from the engineering design community in academia and industry. In the first part of the case study, the aim is to obtain empirical evidence on the lack of focus on EDPE within the engineering design community. The aim of the second part of the case study is to test how closely a computationally framed EDP matches a naturally framed one by a human. In the last part of the case study, the value of the computational EDPE support tool – Pro-Explora V1 (Pro-Explora) presented for the first time in this paper is evaluated.

Presented in the following section are the natural EDPE process including possible determinants of its scarce practice. Section 3 is on the methodology used in this paper to come up with a data-driven computational EDPE framework. Section 4 is about the case study data collection methods. Qualitative and quantitative results of the case study are presented in Section 5 and discussed in Section 6. The paper is concluded in Section 7.

2 LITERATURE REVIEW

2.1 Models and phenomena related to the natural EDPE approach

The EDPE process is characterized by divergent thinking and decision-making for a new EDP. The “rational” model and “garbage can” model relate to the natural EDPE process. The “rational” model is a formal model of science which postulates that careful analyses of previous problems and theories underpin the discovery of a new problem [28].

It supports that a new problem is identified progressively or logically based on gaps in previous problems and theories. The "garbage can" model postulates that a new problem emerges stochastically rather than logically [29, 30]. It supports that a new problem comes up from a stochastic synthesis of previous problems that may not be related. A new problem based on the "garbage can" model is considered more creative than that based on the "rational" model [31]. The "garbage can" model relates to connectionism - a cognitive science concept that likens the connections in computer Artificial Neural Networks (ANN) to natural cognitive ability [32]. The computer ANN contains stochastic and complex interconnected nodes that distribute information for ML.

The "rational" and "garbage can" models describe the natural process through which new opportunities, ideas, or concepts are produced. Specifically, they are used to describe the process of coming up with research topics or titles. For example, Alter and Dennis [28] states that: "As faculty, we tend to teach our students a formal "rational model" of science in which research activity is driven by a solid understanding of prior work. Under this approach, research topics emerge from a careful analysis of prior research and theory." Also, project advice to students is to begin the "search for a suitable problem as soon as possible" [33]. In discussing where a new project is found and how a project is identified, selected, developed, and refined, Dennis and Valacich [29] state that the garbage can model is "a more useful model of how research projects are typically developed" where "the key elements of the project are thrown together into a garbage can, mixed together, and out comes the project." Also, Martin [30] states that an organization looking for a problem could be imagined as a garbage can where, as "members of the organization generate problems and solutions, they dump these into the garbage can" from which a new problem emerges.

Serendipity is a cognitive phenomenon related to EDPE or discovering something new and valuable by chance [34, 35]. It is described as one of the mechanisms of innovation [36], and "connotes the profound ability of finding out valuable things different from those who have been exploring by spending a lot of time or for years" [37]. Serendipity occurs when observation by a design engineer triggers unexplored possibilities. Usually, this is based on coincidence with the design engineer's interest, passion, experience, knowledge, cultural background, and so on. It is reported that the "ground effect" in aircraft is a serendipity discovery [38]. Apophenia is another cognitive phenomenon related to the discovery of a new EDP. It is a natural tendency to see or make meaningful, valuable, invisible connections between unrelated or random data [39]. Apophenia is related to the "garbage can" model and ANN. It could lead to an "invention: creating new, previously unimaginable meanings through accident" [40]. EDPE is considered challenging, and findings on some determinants are presented next.

2.2 Challenges with the natural EDPE

2.2.1 Memory limitation in the natural EDPE process

Studies show that the average amount of information the short-term memory can retain and process when exposed to a new concept is 7 ± 2 [41, 42]. The number approaches the minimum with an increased number of syllables in a word during processing of a sequence of words [43]. EDPE, as a cognitive activity, inherently involves processing word sequences, as the “rational” and “garbage can” models suggest. Hence, despite the complexity of the brain, its information-processing capacity is limited [44]. The cognitive demands of EDPE could push cognitive limits and present a level of difficulty, confusion, fixation, and demotivation, which could result in abandonment.

2.2.2 Cognitive fatigue in the natural EDPE process

Cognitive fatigue is mental exhaustion resulting from tasks requiring deep thinking and could occur within 30 minutes of commencing a cognitive task [45, 46]. An EDPE process, as explained by the “garbage can” model, involves creating unfamiliar concepts through stochastic information combinations, transformations, and/or explorations [47]. Cognitive fatigue could be induced during EDPE tasks and manifest as creative burnout, frustration, and/or tiredness leading to withdrawal from the task [48-52].

2.2.3 Insufficient knowledge and difficulty in prompt initiation of natural EDPE process

It could be inferred from the “garbage can” model that EDPE requires knowledge and information [53]. Sometimes it could be difficult to promptly recall previous knowledge or information. Hence, to “create” a new EDP by stochastically recalling and combining previous knowledge and information could be challenging. Also, knowledge is infinite and could hinder EDPE irrespective of the level of experience of a design engineer. Hence, novice design engineers who are not experienced, knowledgeable, or informed could find it more challenging to practice EDPE.

2.3 Potential computational technologies for EDPE

A computational system is not susceptible to the challenges associated with the natural EDPE process discussed in Section 2.2 [54]. Hence, it could be used to support EDPE. The "garbage can" model discussed in Section 2.1 could be computationally mimicked by using Artificial Intelligence (AI) [55]. A possible approach is to create a data (word) sequencing model using a Markovian model (MM), Bidirectional Encoder Representations from Transformers (BERT), or Long Short-Term Memory (LSTM). An MM is any model that exhibits the Markov chain (MC) property. BERT uses transformers - unique neural network architecture reported effective in modelling long-term dependencies in a sequence [56, 57]. BERT, like LSTM, uses long-term dependencies (depending on previous states) in its network to predict the next state in a sequence. For example, if the state represented with the ellipses is missing from the sequence - "Engineering design is a noble…", contextual word embeddings BERT can be used to predict (natural language inference, NLI) the next or missing state (masked word) as "profession". The next state is inferred relative to the previous states. BERT can also combine the context of both previous and next states (bidirectional) to predict a state in a sequence. This could be described as forward and backward determinism.

The persistence of previous states in BERT and LSTM networks could be disadvantageous because it can impact computational power, time, speed, and cost. Also, BERT requires significant data training. Although there are indications that BERT can be used to predict/generate a new sequence, it is used to predict/generate a sequence that already exists such as known answers to questions. It is reported that in sequence generation, BERT is formulated as a Markov random field language model without additional parameters or training [58]. When the generation of a new concept is intended such as in EDPE, the repetition of already known concepts or solved EDP is not desirable.

Unlike BERT and LSTM, MC is a model of a specific type of stochastic sequence [59, 60]. It is memoryless of past states, and the next state in the sequence only depends on the present state. The MC network uses only forward determinism and does not retain previous information. This results in lesser computing power, cost, time, and higher computing speed. In deterministic chaos theory, forward determinism is considered more important than backward determinism [61]. MC has prior stochastic and decision-making applications in engineering design including 1) modelling of transitions in communications, 2) modelling of sequential design decisions, and 3) analyses of behavioral patterns in engineering design [62-65]. An MM is used in this paper as the pioneering technology to provide the basis for comparison with BERT and LSTM in future research in EDPE applications.

Big data could be equivalent to the previous knowledge required in the “garbage can” model. Big data is a large volume of structured, semi-structured, and unstructured data from which new knowledge or information can be found [66]. Big data is associated with some computational technologies, including data extraction, NLP, ML, and duplication recognition [67]. Presented in Section 3 is how an MM is used in synthesis with big data and the associated technologies to come up with computational support in EDPE.

2.4 Research questions

Despite being an important and standard requirement, EDPE lacks practice within the engineering design community, as discussed in Section 1. There are scholarly opinions that EDPE is a challenging task with cognitive limitations (Section 2.2). Unlike EDPS, there is no specific support tool for EDPE. However, emergent computational technologies could provide support for EDPE. To this end, the research questions (RQs) that follow are addressed in this paper. The answers to the RQs would provide knowledge for interventions in the lack of EDPE practice in engineering design. In the first part of the case study presented in this paper, *RQ1* is addressed. In the second and third parts of the case study, *RQ2* is addressed:

RQ1: Do design engineers understand that EDPE is a standard requirement in engineering design, like EDPS?”

RQ2: Could emergent computational technologies support novice and experienced design engineers in EDPE?

3 METHODOLOGY

3.1 Theoretical framework for computational EDPE

Concerning *RQ2*, a theoretical data-driven computational framework, shown in Fig. 1, is presented to support EDPE as a challenging activity. The framework is based on the information presented in Table 1 on the natural EDPE approach and its computational equivalence. The natural EDPE approach in Table 1 is based on findings discussed in Section 2 and supported by most intellectual property offices [68].

Table 1 Comparison of algorithms for natural and computational EDPE

Natural EDPE approach	Computational EDPE approach
1) Identify an EDP of societal relevance by accident (serendipity), stochastic synthesis (“garbage can” model), logical progression (“rational” model), and/or conceptualization (apophenia)	**1)** Frame an EDP of societal relevance by stochastic synthesis of big data, computational technologies (data extraction, ML, NLP), coding capabilities, connectionist theory, deterministic chaos, MM, BERT, and/or LSTM
2) Search manually for prior existence in relevant databases using search engines.	**2)** Make an automated search for prior existence in relevant databases using duplication recognitions.
3) Decide, subject to acceptance by society or a relevant authority	**3)** Decide, subject to a design engineer's acceptance.

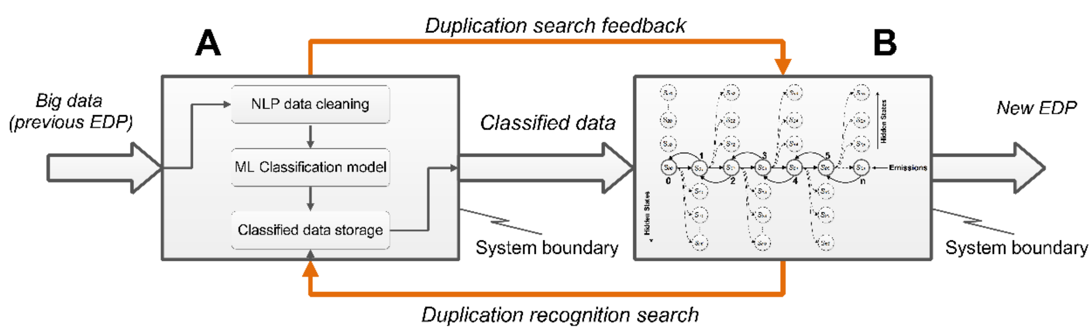

Fig. 1 Big data-driven computational EDPE framework

The computational EDPE framework shown in Fig. 1 comprises *System A* and *System B*. The input to *System A* is a collection or corpus of engineering design project titles extracted online from Compendex, Scopus, journals and conferences databases, and findaphd.com using a python data extraction tool “Scrapy”. The output from *System A* is processed data which feeds *System B* to produce a new EDP as output. The project titles used as input to *System A* are important lines of words that represent EDPs in engineering design projects [69]. A "project title should provide information about the topic being studied, and may consist of the actual problem statement" [70]. According to Martin [30], "the researcher should critically review the literature on a given topic in order to find an important issue which previous research has failed to resolve successfully". This 'important issue' is usually formulated as a title – the problem solved, being solved, or yet to be solved. Hence, the extracted titles in this paper are previous EDPs. As an example, in the project title: "Design of an Automatic Sprinkler Fire Fighting System", the EDP of fighting the fire using an automatic sprinkler system is described. The underlying EDP is: "What is a better way, among the existing alternatives, to fight fire?" The design project is conceived as a solution to fire fighting. It is believed that such a solution was not available when the project was conceived. This makes the project an EDP that needs to be addressed because there is a potential benefit in doing so. The EDP addressed in the project is described with the title. Albeit the title appears as an EDS, it is an EDP if that EDS is unavailable or yet to be realized. It remains an EDP until it is solved. Hence, computationally exploring and identifying a new EDP could lead to an invention or innovation.

The increasing volume of titles, continuously collected for EDPE, could be regarded as structured big data of previous EDPs. Hence, this is the concept of the 'Big data' indicated as input to the model in Fig. 1. To mimic the natural EDPE, the framework in Fig. 1 is created to use or learn from only the natural EDPs to come up with a computational equivalent. The output from *System B* (which contains the MM in Fig. 2) is a unique EDP distinct from the input. The preprocessing and processing of the corpus in *System A* for input in *System B* to produce a unique EDP are presented next.

3.1.1 Preprocessing of corpus for EDPE

The corpus extracted online is preprocessed in *System A* Fig. 1 using NLP and ML. The corpus is first prepared as a “tab-separated value” with each line in the corpus ending with a period. On inspection, some of the extracted titles in the corpus appear vague to describe an EDP. This necessitates an ML classification model to classify subsequent extracted titles that do not describe an EDP (Non-EDP). The extracted corpus is manually separated by inspection as a dataset of EDP and non-EDP. This is to enable the training of the algorithm for the classification model using supervised ML – an aspect of AI that provides computer systems with the ability to learn from data. The dataset size is 2133 (comprising 1833 EDP and 300 non-EDP), and a 20% test size is used for the ML. The training requires that the dataset is ‘cleaned’ and ‘tokenized’ as part of NLP [71, 72]. The ‘cleaning’ requires the removal of regular expressions or characters that specify a search pattern in extracted texts such as “?”, “@“, and “$”. It also requires the removal of stopwords from the dataset such as “a”, “for”, and “the” which are insignificant in NLP. Different algorithms are tried during the training, including RandomizedSearchCV, Naïve Bayes (Gaussian and Multinomial), and Random Forest. These algorithms are part of Scikit-learn – a library in Python that provides many unsupervised and supervised learning algorithms. Trying different algorithms to select the best based on performance is a common practice in ML. Two Scikit-learn performance evaluation libraries – classification report and confusion matrix [73], are used during the training to evaluate the performance of the algorithms in the classification model. For RandomizedSearchCV, the accuracy calculated using the confusion matrix is 93%. The classification report shows the precision, recall, and f1-score accuracy metrics as 94%, 93%, and 93%, respectively. These metrics suggest that only a few EDPs are wrongly classified as non-EDPs and vice versa. Hence, RandomizedSearchCV is used based on best performance values. The preprocessed corpus in *System A* Fig. 1 is stored and processed in *System B* Fig. 1 to produce a new EDP as presented next

3.1.2 Processing of corpus for EDPE

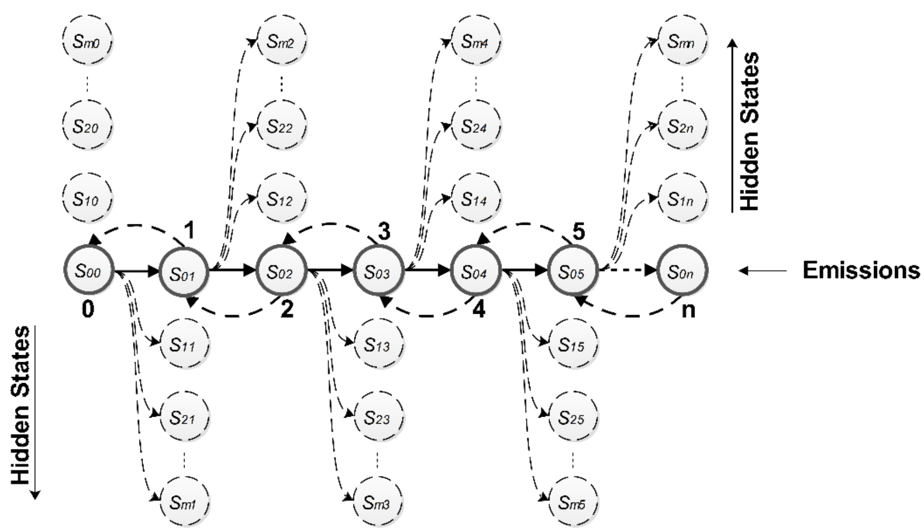

Fig. 2 A theoretical model for EDP sequencing [8]

The MM contained *System B* Fig. 1, discussed in Section 2.3, is enlarged in Fig. 2 for clarity. In Fig.2, the MM exhibits the properties of a two-stage MC known as a hidden

Markov model (HMM). It has hidden and physically observable states (emissions) [74, 75]. What constitutes the hidden states and emissions in this paper are explained in Section 3.3. MC is used as an HMM in many real-life problems, such as handwriting recognition, machine maintenance, and weather forecasting. This is because MC alone does not fully represent the intent in many real-life problems [76]. In this paper, as shown in Fig. 2, the hidden states (for example, $S_{10}S_{20}\ldots S_{m0}$) are discreet, while the emissions $S_{00}S_{01}S_{02}\ldots S_{0n}$ exhibit the Markov property (Section 2.4). The probability of the emissions depends on the probability of the hidden states. The emissions transit such that the next emission depends on the present emission and not on the past emission(s). The emission transitions are assumed to be observed at equal time intervals at the indices *0,1,2,3,… n*, known as epochs [77]. This means that time-homogeneity (discrete instead of real-valued time) applies. In this application, the actual time for transitions is computationally very small and justifies the time-homogeneity assumption. The stochastic process in Fig. 2 could be expressed as Eq. (1).

$$f: S_n \text{ x } \Lambda \rightarrow S_{n+1} \qquad n \in \mathbb{Z}^+;\ \mathbb{Z}^+ = \{0, 1, 2, 3, \ldots\} \tag{1}$$

In Eq. (1), as used in this paper, the output S_{n+1} is a function of two arguments, s_n and λ. This is such that, $s_n \in S_n$ and $\lambda \in \Lambda$. The function $f(s_n, .)$ is a random variable (S_{n+1}) for each $s_n \in S_n$, while for each $\lambda \in \Lambda$, $f(., \lambda)$ is a hidden function between S_n and S_{n+1}. This hidden function, $f(., \lambda)$, makes the output of *System B* Fig. 1 to be indeterministic which represents the unpredictability of the highest order [78]. Hence, the probability of any predictor's confidence in the output of *System B* Fig. 1, relative to the input, cannot be unity. Generally, the transition probability of the MM in Fig. 2 is given in Eq. (2).

$$p_{ij} = P(S_{0(n+1)} = T_j | \, S_{0n} = T_i) \tag{2}$$

Eq. (2) shows that, in Fig. 2, if the probability of an emission at epoch *n* is T_i then there is a probability that at epoch *n+1* the emission is T_j [79]. Note that the emission transitions occur if, and only if, $p_{ij} > 0$. As shown in Fig. 2, an emission is represented as S_{ij} where the row vector is the sequence index and the column vector is the epoch. For simplicity of analysis, let the emissions at epochs *0,1,2, … n* in Fig. 2 be *a,b,c,… z*. Specifically, applying Eq. (2) to Fig. 2 yields Eq. (3).

$$\left.\begin{aligned} &P(S_{01} = b, S_{02} = c, \dots S_{0n} = z | \, S_{00} = a) \\ &\quad = P(S_{00} = a) P(S_{01} = b | \, S_{00} = a) P(S_{02} = c | \, S_{01} = b) \dots \\ &\qquad\qquad \dots P(S_{0n} = z | \, S_{0(n-1)} = y) \\ &\quad = p_a^{(0)} p_{ab} p_{bc} \dots p_{yz} \end{aligned}\right\} \tag{3}$$

Relating Eq. (3) to Fig. 2, $p_a^{(0)}$ represents the initial emission probability for the stochastic process at epoch *0*. P_{ab} represents the transition probability from epoch *0* to epoch *1*. P_{bc} represents the transition probability from epoch *1* to epoch *2*. P_{yz} represents the transition probability from epoch *n - 1* to epoch *n*. The initial emission would be necessary if the observable state of the sequence after *n* transitions is of interest. If the initial emission probability is known, then $p_a^{(0)} = 1$, and Eq (3) results in Eq. (4).

$$= p_{ab} p_{bc} \dots p_{yz} \tag{4}$$

After an EDP is produced in *System B* Fig. 1 as a sequence of emissions, a duplication recognition search is performed. The sequence is only outputted from *System B* Fig. 1 as a new EDP if it does not have a duplicate (exact match) in the original corpus stored in *System A* Fig. 1. Otherwise, it is discarded and another sequence produced. The framework in Fig. 1 is deployed to produce a computational EDPE tool discussed next.

3.3 Pro-Explora – a computational support tool for EDPE

Pro-Explora is a computational support tool for EDPE. In Pro-Explora, the theoretical EDP sequencing model in *System B* Fig. 1 is realized by processing the input corpus in *System B* Fig. 1 as a python dictionary data structure. The corpus is split into single words with each word as the dictionary key. The value list of each key contains all words that come immediately after the key in all occurrences of the key in the corpus. To closely mimic a natural EDP, the initial word/emission at epoch *0* in Fig. 2 is randomly selected from the list of hidden states $S_{10}S_{20}… S_{m0}$. The hidden states $S_{10}S_{20}… S_{m0}$ comprise the first words of each EDP in the extracted corpus. For example, S_{10} will be “Design” and S_{20} will be “A” for a corpus that contains the two EDPs – [“Design of a mechanical intrusive force detection device.”, “A design of an automatic bottle opener.”]. The dictionary key “of” in the corpus will have the values “a” and “an”. After the initial emission, the rest of the emissions are constrained to be randomly chosen from the mutually exclusive hidden states at epochs *1,2,3,4,5,…n* based on the Markov property. These observable states in Fig. 2 represent the new EDP from *System B* Fig. 1. The framework in Fig. 1 mimics the natural EDPE (Section 2.1), especially the “garbage can” model. As presented next, a useful output from System B Fig. 1 requires specific adjustments to the MM in Fig. 2 [76].

Adjustment 1: Enhanced stochasticity

For a word ending with a period in a list of hidden states in Fig. 2, synonyms of the word are added to the list without replacing the word. For example, “extractor.”, “centrifuge.”, and “threshing machine.” are added to the list containing the word “separator.” The synonyms are obtained using the python wordhoard library and online thesaurus. This adjustment increases the stochasticity and number of Pro-Explora outputs with the same EDPs in the corpus and makes the tool create new EDPs that involve newer technologies.

Adjustment 2: Output word-count constraints

The number of words (*n*) in the new EDP from Pro-Explora is constrained to a minimum of 6 and a maximum of 12 ($6 \leq n \leq 12$). This is based on findings from studies and the result of the EDP word-count analysis on the extracted corpus as shown in Fig. 3.

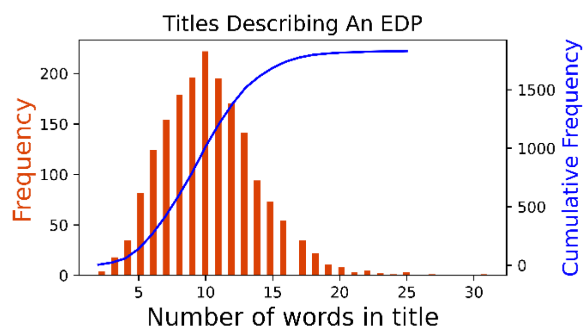

Fig. 3 Word count of titles describing EDP

Shown in Fig. 3 are the word counts mostly used in titles that describe an EDP (Section 3.1.1). It could be seen that the most used word counts range between 8 and 12. This range is significant in this paper and correlates with the scholarly suggestions that a maximum of 12 words should be used to describe an EDP to inspire thoughts and attract attention [80-83]. Also, as discussed in Section 2.3.1, the limit of words the brain can process at once is between 5 and 9. Hence, the word count limit of 6 - 12 for the Pro-Explora output is considered appropriate. Using python code, it is checked that the output from *System B* Fig. 1 ending with a period satisfies the word count limit.

Pro-Explora GUI

Backend python codes enable the functionality of Pro-Explora. However, it is accessed through a simple web Graphical User Interface (GUI) shown in Fig 4 which requires a Login to access at https://www.explorefoss.com/. A logged-in user "Engineer" is shown in the Pro-Explora GUI in Fig. 4. The GUI has two settings that contain some options that could be selected based on preference before EDPE as explained next.

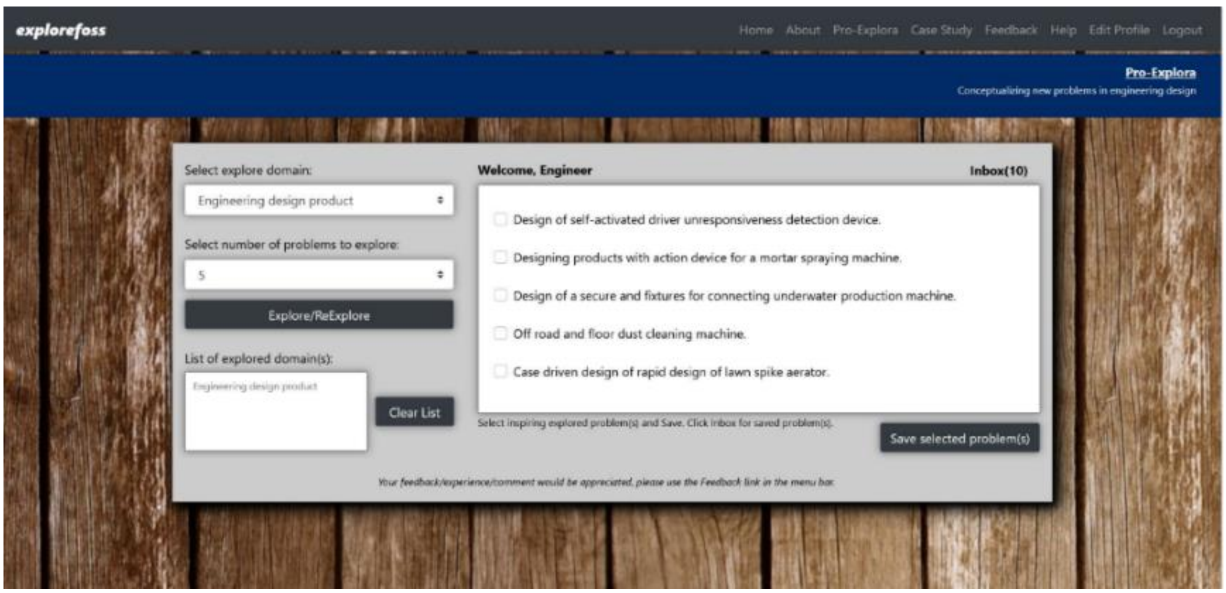

Fig. 4 Pro-Explora GUI with some framed design problems

The Pro-Explora GUI, as shown in Fig.4, has the "Select explore domain" and "Select number of problems to explore" settings. The "Select explore domain" setting has four domain options – "Engineering design product", "Engineering design research", "Engineering design machine intelligence", and "Engineering design cross-domain". The "Engineering design cross-domain" option has the largest database, a combination of the other three domains. The "Select number of problems to explore" setting has 1 and 5 as the minimum and maximum numbers to select with 1 as the default. The history of all explored domains is displayed in the "List of explored domain(s)" section of the GUI. With the preferential options set, clicking on the Explore/ReExplore button activates a new EDPE process. The GUI in Fig. 4 displays five framed EDPs and any or all could be selected and saved for later review. Pro-Explora can frame over 100 unique EDPs per minute. This would be impossible with the natural EDPE process (Section 2.1), considering the associated challenges (Section 2.2). The framework and tool for computational support in EDPE are presented in this section. Presented in the next section are the methods of data collection and analyses to answer the *RQs* in this paper.

4 DATA COLLECTION METHOD AND ANALYSES

4.1 Case study details

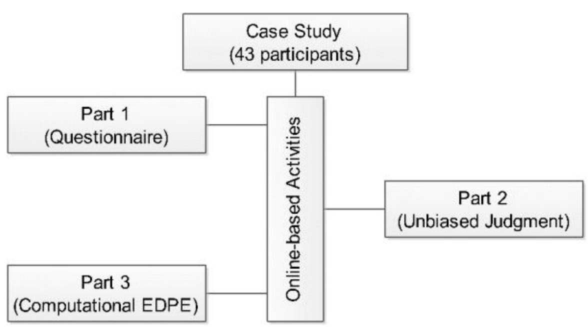

Fig. 5 Overview of data collection approach

As shown in Fig. 5, the data collection method is through a case study comprising three parts. Adverts for the case study are sent through channels targeting professionals and students (Year 2 upwards) within engineering design or design-related academic programmes. Participants for the case study include students from UK and Singapore

Universities and professionals from countries including Canada, France, India, Russia, Singapore, the United Kingdom, and the United States. The case study advert specifies eligibility for participation - healthiness, a basic understanding of creativity, and an engineering design-related background. All participants provide qualifications, years of experience, and other details (Table 2). Only one participant who showed interest is disqualified based on not meeting the criteria of educational background. The creativity “understanding” in the eligibility is important because creativity as a phenomenon for coming up with something new and useful is strongly correlated with EDPE [84-88].

Table 2 Case study participants' detail

Participants	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

B – Bachelors, MS – Masters

As shown in Table 2, the participants are categorized based on their years of experience as either “Novice” (Mean (*M)* = 1.5 years, Standard Deviation (*SD)* = 0.7 years) or “Experienced” (*M* = 8 years, *SD* = 3.8 years, Range (*R*) = 17 years). An inclusion criterion of 0 – 40 years of experience is applied to ensure that participants find it easy to perform computer-based activities. The three parts of the case study shown in Fig. 5 lasted about 30 minutes in total and are presented next.

Case Study: Part 1 - Questionnaire responses

This part of the case study addresses *RQ1*. The aim is to subtly test the consciousness of EDPE practice within the engineering design community. As previously mentioned, creativity correlates strongly with EDPE and EDPS. The participants are given the questions in Table 3 as an online questionnaire to respond to. In Table 3, Questions ‘a’, ‘c’, and ‘d’ are subtly designed to test if participants have a conscious understanding (personally or taught) of creativity relative to EDPE or identifying a new EDP. An effort is made to “avoid leading questions” by not mentioning EDPS or EDPE [89].

Table 3 Questionnaire for Part 1 of the case study

Questionnaire themes
(a) What does it mean to be creative?
(b) What are the major roles of creativity in engineering design?
(c) Do You Consider Yourself Creative?
(d) Why are you creative or uncreative?
(e) Were you taught creativity at University or at work?

Case Study: Part 2 - Unbiased judgment of natural and computational EDP

In this part of the case study, the participants are given the following instruction: “Below, you are presented with 20 unique engineering design-related problems. Professionals within the engineering design discipline conceptualized and produced some of these problems while some are computationally generated with a computational tool. You are required to go through each of the 20 problems and choose one option under each problem based on whether you think the problem is produced by a person or computationally generated.” Two options are provided for response.

Different sets of 20 EDPs similar to that in Table 4 are presented to each participant. However, each set contains a randomly arranged 5 EDPs framed by a design engineer (naturally framed) and 15 EDPs framed using Pro-Explora (computationally framed). The ratio (1:3) of the EDPs is intentionally not disclosed to the participants. Since the participants are unaware of this ratio, it helps to eliminate bias in their judgements. To make the reader guess, the categories of the EDPs – “naturally” or “computationally” framed, are not indicated here but in Section 5.4.

Table 4 A sample set of 20 EDPs for participants

Naturally and computationally framed EDPs
1. Design of a mechanical intrusive force detection device.
2. To design a portable water distillation device.
3. A sustainable packaging design for wine.
4. Designing an interactive interface for collaborative engineering design.
5. A design of an automatic bottle opener.
6. Towards intelligent emotion detection system for video traffic surveillance.
7. Ai-based learning models for video traffic surveillance.
8. Design and material properties to minimize biofilm deposits.
9. Design of human-powered hybrid electric-power shovel for the physically challenged.
10. Design of self-reconfigurable production equipment during operation.
11. Anti riot drone without traffic lights.
12. Investigation of anomaly detection in a critical materials.
13. Design of a self-timing solar seawater desalination machine.
14. Staging co-design for reverse modeling of product development.
15. Detecting aggressive driving behavior using scilab.
16. Design of remote intelligent home finance software.
17. Designing products by artificial intelligence design approach.
18. A computationally efficient real-time vehicle and speed detection using federated learning.
19. Automatic mechanical footstep power tiller machine.
20. Design of production information retrieval system.

Case Study: Part 3 - Evaluating the value of a computational EDPE support tool

This is the last part of the case study and contributes to answering *RQ2*. It is about evaluating a computational EDPE support tool – Pro-Explora, presented in Section 3.3. Participants use the tool to come up with at least 5 EDPs in about 10 minutes. On a Likert scale of 1 – 10, the participants rated the usefulness of the EDP framed by the tool. They also provide additional information on 1) the reason for the usefulness rating they provide, and 2) whether the EDP inspired or prompted them to think of a different EDP related or unrelated to the originally framed EDP.

4.2 Data analysis

Data from the case study is qualitatively and quantitatively analyzed. The qualitative analysis is performed with NVivo 12 - powerful software for qualitative data analysis, following the workflow in Fig. 6.

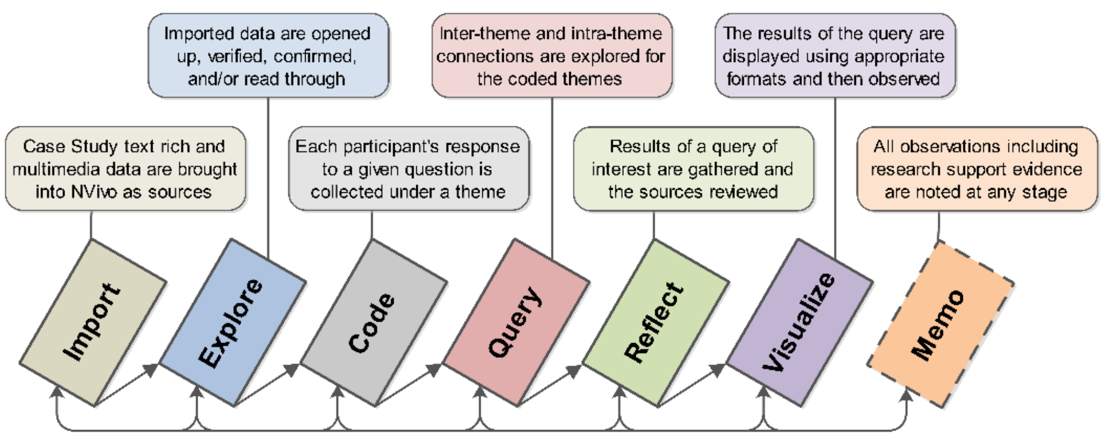

Fig. 6 Case study qualitative analysis workflow

In Fig. 6, the 7 stages involved in qualitative analysis with NVivo are shown. The various data collected are imported into NVivo and arranged. The data is coded - the process of gathering materials (participants' responses) by topics or themes. This is followed by querying the data for patterns and connections. The query results are reflected upon and visualized. Although the workflow in Fig. 6 is iterative, the first 6 stages should be sequentially completed before any iterative update can be made to any of them. However, the last stage (Memo) can be referenced from any stage at any time.

5 RESULTS

5.2 General understanding of creativity

In Part 1 of the case study (Section 4.1), Questions *'a', 'b', and 'd'* in Table 3 is to address *RQ1*. They are designed to subtly reveal the participants' general understanding of creativity relative to EDPE, which strongly correlates with creativity as mentioned in Section 4.1. Responses to the questions are qualitatively analyzed using the WFA in Fig. 7, Cluster Analysis (CA) in Fig. 8, and text query search (Fig. 9.)

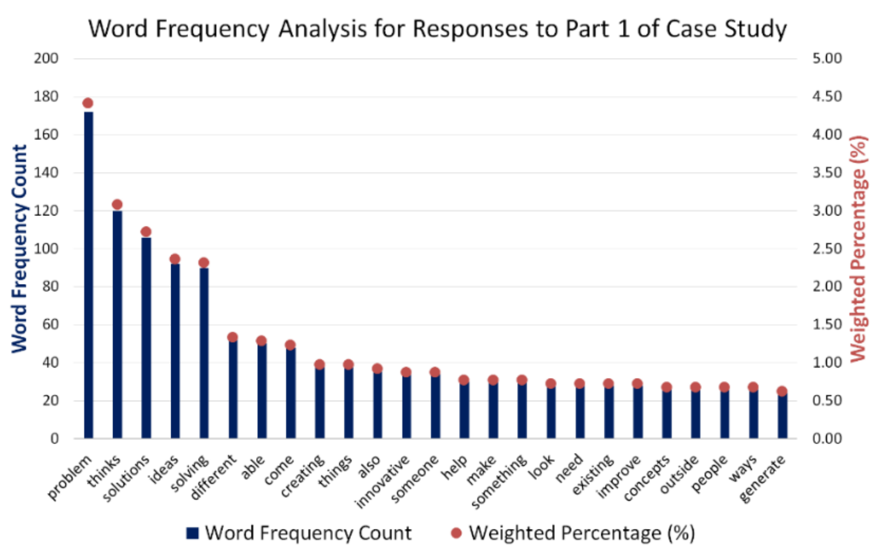

Fig. 7 Most frequent words in explaining creativity in engineering design

Presented in Fig. 7 are the 25 most frequent words used by the participants to explain what creativity means to them. It could be seen that the participants mainly associate the word ‘problem’ with creativity in engineering design. A cluster analysis (CA) is performed to observe the relationship between the 25 words in Fig. 7. The CA result is shown in Fig. 8.

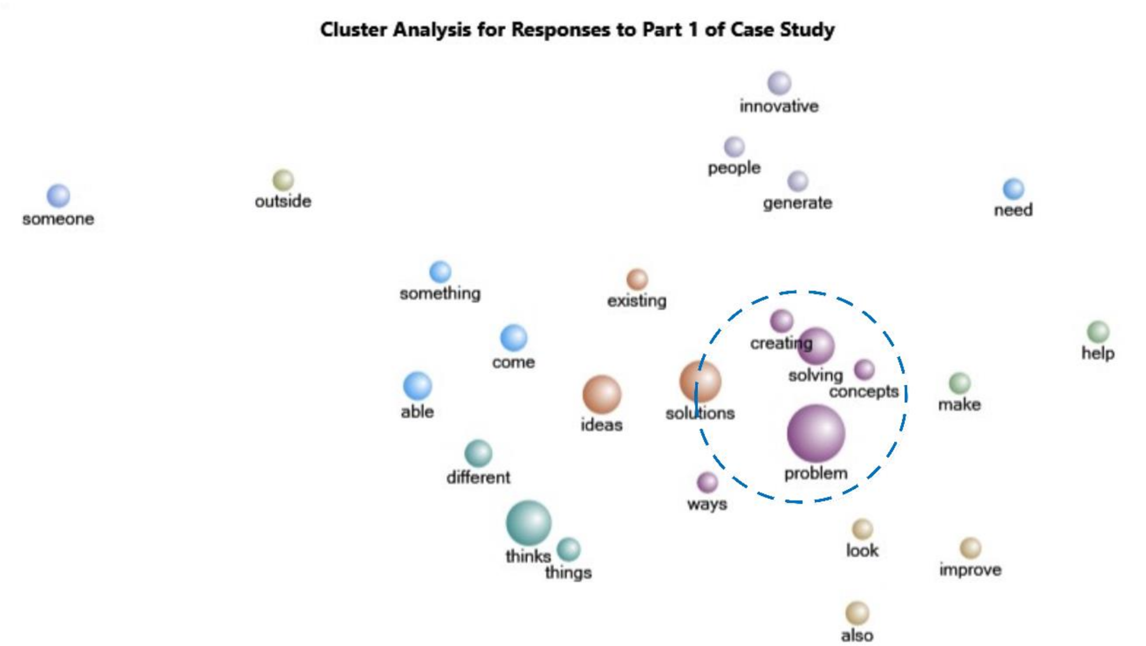

Fig. 8 Word cluster showing a relationship in words explaining creativity

Shown in Fig. 8, are clusters suggesting the relationships between the 25 most frequent words in Fig. 7. The highest single cluster of 5 words – ‘concepts’, ‘creating’, ‘solutions’, ‘solving’, and ‘problem’ could be seen encircled in Fig. 8. To understand the context of the encircled cluster, a text query search is run with the 5 cluster words, and the result is presented in Fig. 9.

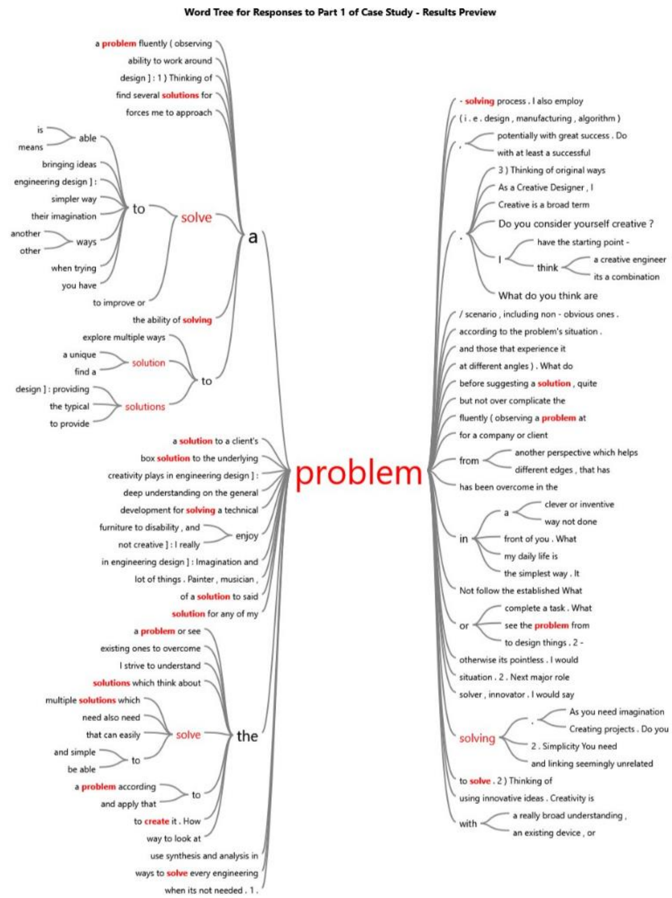

Fig. 9 Word Tree result for a text query search

The Word Tree (WT) in Fig. 9 shows the root term as ‘problem’ which is the most frequent word in the WFA in Fig. 7. For a clearer context of the relationships in the cluster words in Fig. 8, five words are allowed on either side of the root term in Fig. 9. As the WT shows, creativity is generally understood to be an EDPS phenomenon. This is likely to be the participants' understanding of creativity from academia and literature. There is no explicit association of creativity to EDPE by the participants. The next result presented is on the teaching of creativity in engineering design in academia.

5.3 Teaching creativity in engineering design

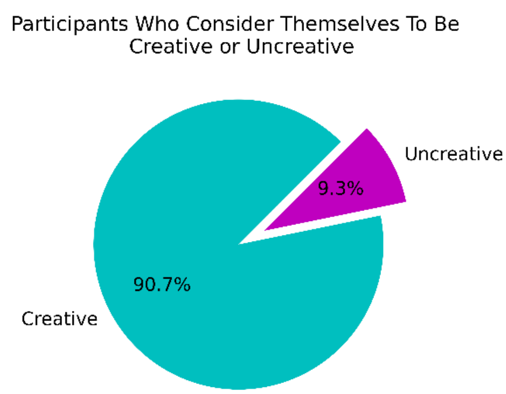

Fig. 10 Participants' responses to their creativity ability

The participants' responses to Questions 'c' and 'e' of Part 1 of the case study (Table 3) are presented in Fig. 10 and Fig. 11. Questions 'c' and 'e' are designed to reveal the adequacy and focus on creativity teaching in academia. As shown in Fig. 10, most of the participants consider themselves creative which shows that creativity is a popular skill in engineering design.

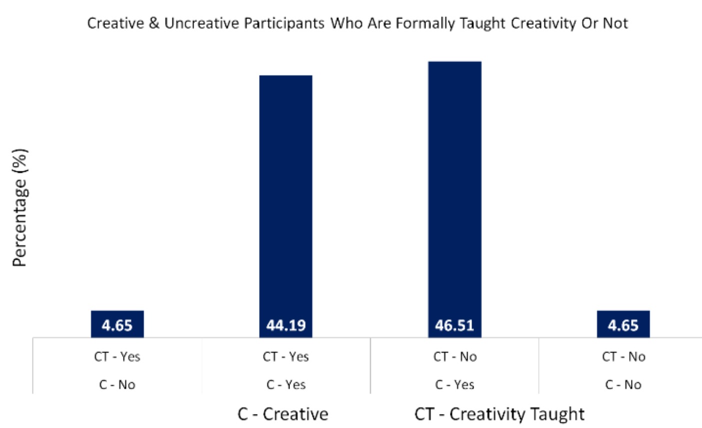

Fig. 11 An insight into creativity teaching in academia

Presented in Fig. 11, are the percentages of participants who are either taught creativity in academia or industry and those who are not. It could be seen that not all the participants who consider themselves creative (in Fig. 10) are formally taught creativity. Some design engineers could be naturally creative without being formally taught as shown in Fig. 11. However, this should not deter effort in teaching creativity techniques and skills formally in academia and industry. Every good natural ability needs formal support. For example, some people are naturally good at playing football but football academies exist. There is the possibility that those who are creative (C – Yes) but not taught creativity (CT – No) could have been more creative if formally taught creativity in academia. Also, the possibility exists that those who are not creative (C – No) and not taught creativity (CT – No) could have been creative if formally taught. The result presented in Fig. 11 suggests that creativity teaching in academia may be below average as over 50% of the participants are not formally taught creativity. For the lesser percentage that is formally taught creativity, the focus is on EDPS while EDPE is ignored as shown in Fig. 9. Following the completion of Part 1 of the case study, Part 2 is commenced and the results are presented next.

5.4 Differentiating a computationally and naturally framed EDP

In Part 2 of the case study, as part of answering *RQ2*, the intent is to test if the participants could differentiate between a computationally and naturally framed EDPs. In the set of 20 EDPs presented in Table 4, EDPs 1 – 5 are framed by a design engineer while EDPs 6 – 20 are framed by Pro-Explora. Participants are required to distinguish both categories of

EDPs, as explained in Section 4.1. The result of this activity for all the participants (Novice and Experienced) is presented in Fig. 12. The “Novice” and “Experienced” participants are code-named and presented in Table 5 for confidentiality.

Table 5 Novice and Experienced participants

Novice				Experienced			
1013	1019	1026	1058	1022	1035	1041	1048
1014	1020	1027		1028	1036	1042	1049
1015	1021	1029		1030	1037	1044	1057
1016	1023	1043		1031	1038	1045	1059
1017	1024	1050		1032	1039	1046	1060
1018	1025	1051		1033	1040	1047	1061

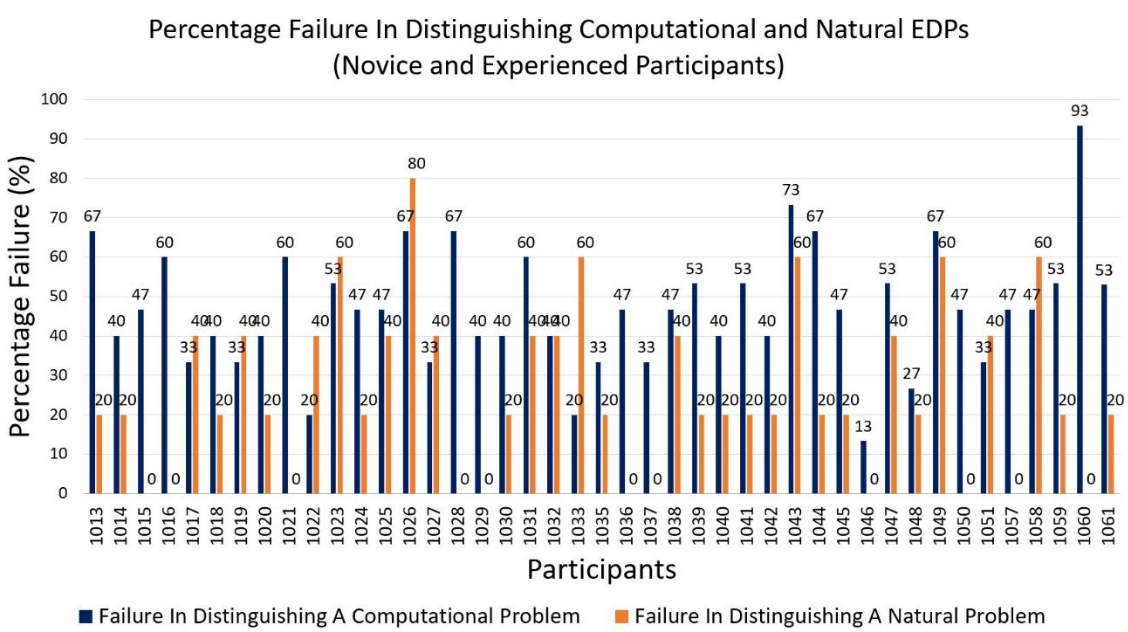

Fig. 12 Result of Part 2 (Unbiased Judgement) of the case study

As shown in Fig. 12, some of the participants have zero failures in distinguishing a naturally framed EDPs. These participants are “1015”, “1016”, “1021”, “1028”, “1029”, “1036”, “1037”, “1046”, “1050”, “1057”, and “1060”. It could be seen in Table 5 that some of these participants are “Novice” while some are “Experienced”. These zero failures suggest that the “computationally framed EDP” judged by the respective participants as a “naturally framed EDP” appears natural, useful, and meaningful. The correlation between the participants’ years of experience and failures in distinguishing a computationally and naturally framed EDP is tested for statistical significance. The results are presented in Table 6. Note that the participants are already categorized based on years of experience in Table 2. Hence, the failures in Table 6 are relative to the participants’ (Novice and Experienced) years of experience.

Table 6 Correlation between experience and distinguishing a computational EDP

(Null hypothesis (*Ho*): There is a significant relationship between the participants’ experience and their failures in distinguishing a computationally and naturally framed EDP)

	P-value	Pearson’s r
Failures in distinguishing a computationally framed EDP	0.78	0.04
Failures in distinguishing a naturally framed EDP	0.72	-0.06

P-value <0.05 defines statistical significance

Cosine similarity assessment

The result presented in Fig. 12 indicates a misjudgment of at least one naturally framed EDP or computationally framed EDP by all the participants. This suggests a similarity between the two categories of EDPs. As a further analysis, the naturally framed (EDP1 – EDP5) and computationally framed (EDP6 – EDP10) EDPs in Table 4 are assessed for differences or similarities. The first EDP in Table 4 is named correspondingly as EDP1, the second EDP2, and the tenth EDP10. A random quote (Q*) is added in Table 7 to see how its similarity compares with the EDPs. The result of the assessment is presented in Table 7. The assessment is performed using cosine similarity, which measures similarity between texts by "calculating the cosine of the angle between the two vectors" [90]. A web text trained Spacy pipeline, en_core_web_lg, is used to compute the cosine similarity. Spacy is an open-source python library for NLP. Cosine similarity ranges from 0 – 1 with 1 indicating 100% similarity. It could be seen in Table 7 that most similarities between the naturally and computationally framed EDPs are above 65%. This justifies the failures in judgments in Fig. 12. Since Q* in Table 7 is a quote, its similarity with the EDPs is the lowest across rows.

Table 7 Cosine similarity assessment result

	EDP6	**EDP7**	**EDP8**	**EDP9**	**EDP10**	**Q***
EDP1	0.8163	0.7000	0.7461	0.8465	0.8589	0.6163
EDP2	0.7164	0.6240	0.7838	0.6794	0.6686	0.5628
EDP3	0.6700	0.6656	0.6258	0.6776	0.6872	0.5249
EDP4	0.7842	0.7275	0.7191	0.7500	0.7710	0.5371
EDP5	0.6163	0.5767	0.5481	0.7164	0.7306	0.4750

Q* - "Anyone who has never made a mistake has never tried anything new." (Albert Einstein)

5.5 The value of computational support tool in EDPE

To address *RQ2*, in Part 3 of the case study, the participants used Pro-Explora as a support tool to come up with some EDPs. They rated 5 of the EDP on a Likert scale of 1 – 10 (with 10 being the highest). In Fig. 13, the mean of the ratings for all participants (Novice and Experienced) is shown with the standard error (SE).

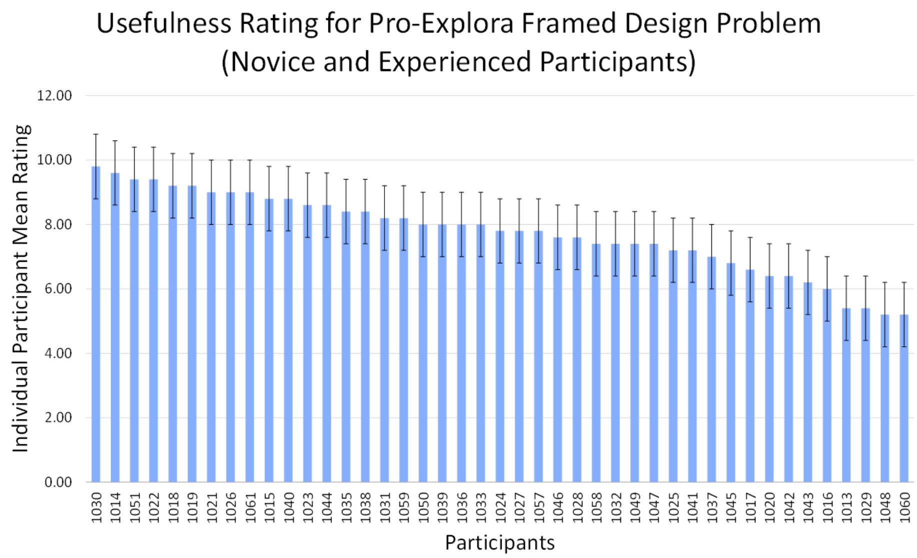

Fig. 13 Participants' usefulness rating for Pro-Explora framed EDP (with ±1 SE bar)

The overall mean usefulness rating of the participants (Novice and Experienced) shown in Fig.13 is 7.74. Coincidentally, the separate mean usefulness ratings of the "Novice" and "Experienced" participants is 7.74 and 7.74, respectively. The mean for a Likert scale of 1 – 10 is 5.5. Hence, the overall usefulness rating of the participants (7.74) for Pro-Explora generated EDPs is above the mean value (5.5) of the Likert scale. This usefulness rating of 7.74 out of a maximum of 10 on the Likert scale could be considered high. In Fig. 13, the error bar overlaps give a visual insight on the variabilities in the individual rating values which aggregate to the means used for the plot. For example, in Fig. 13, the error bar for participant "1030" overlaps with others to the right of participant "1030" up to participant "1033". This indicates that some of the separate ratings of the participants to the right of “1030” are higher than that of participant “1030”. As seen in Table 5, these participants belong to either the “Novice” or “Experienced” category.

Some of the participants mention that they are inspired or prompted to think of a different or related EDP based on the EDP framed by Pro-Explora. The correlation between the participants' experience and their usefulness ratings is statistically analyzed and presented in Table 8. The analyses in Table 8 are relative to the participants’ years of experience, as indicated in Table 2. As shown in Table 8, the analysis is performed for the overall participants (Novice and Experienced). As a further confirmation, the analysis is also performed separately for only the “Novice” and only the “Experienced” participants. Coincidentally, as shown in Table 8, the p-value for the overall rating is the same as that of the “Experienced” participants.

Table 8 Correlation between experience and rating of a computational EDP

(Null hypothesis (*Ho*): There is a significant relationship between the participants’ level of experience and their usefulness ratings of Pro-Explora framed EDP)

	P-value	Pearson’s r
Usefulness Rating (Overall)	0.69	-0.06
Usefulness Rating (Novice participants only)	0.16	-0.34
Usefulness Rating (Experienced participants only)	0.69	-0.08

P-value <0.05 defines statistical significance

6 DISCUSSION

6.1 Academic implications

The findings and results contribute to knowledge by providing empirical evidence on the 1) lack of focus on EDPE within the engineering design community and 2) value of computational support in the EDPE process for the first time. The lack of attention on EDPE contrasts with the standard expectation of design engineers in identifying societal EDP using their experience, knowledge, and background [4, 10, 13, 91]. The natural EDPE process investigated in this paper requires creativity. Over 50% of the participants in the case study indicate that they were not formally being taught creativity. This suggests a lack of creativity teaching in academia within engineering design disciplines [92]. Also, the effort in teaching creativity in academia is focused on EDPS while EDPE is ignored. The case study result indicates that the general understanding of creativity is about EDPS within the engineering design community. This understanding is likely from the teachings provided in academia. Hence, effort in teaching creativity in engineering design disciplines should equally focus on both EDPS and EDPE.

6.2 Industry implications

There are scholarly opinions, as mentioned previously, that EDPE is a challenging activity. However, a paper on why EDPE is challenging lacks. This paper highlights the possible determinants of the challenges associated with EDPE. Hence, this makes it possible to extensively investigate some computational technologies that could support the natural EDPE process, while it was previously indicated that computational EDPE would be impossible [93]. This paper would provide opportunities for further research in the area of computational EDPE in engineering design and other fields. For example, it could be applied in the medical field to identify new possibilities. The results presented in this paper show that both novice and experienced design engineers can come up with at least 5 EDPs in about 10 minutes. This feat would be difficult or impossible within natural limits due to cognitive limitations and fatigue (Section 2.2). In the natural EDPE process, it is considered “entirely reasonable to spend several months or longer thinking about potential problems” to solve [94]. The EDP framed by Pro-Explora is given an average “usefulness” rating of 7.74 out of 10 by both novice and experienced design engineers (Section 5.5). This indicates that design engineers could be computationally and intentionally inspired, prompted, or supported in using their knowledge in EDPE. The inspiration occurs when the Pro-Explora framed EDP coincides with the design engineer’s knowledge, experience, and/or background. This is similar to serendipity discovery (Section 2.1), and some participants agree that the EDP framed by Pro-Explora inspired them to think of a different EDP. Knowledge is infinite, and design engineers cannot measure their knowledge or intentionally recall all they know [95, 96]. Hence, computationally prompting the design engineer of an EDP that may be within the domain of their knowledge to solve is advantageous and a rapid way of discovering a new EDP, an invention, or innovation.

6.3 Limitations and opportunities

The results are based on the direct responses provided by the participants. No further verification of the information is carried out. For example, the Universities attended by the participants who reported that they are not taught creativity are not contacted for verification. Also, being an online activity, it is not certain whether the participants spent longer or lesser than 10 minutes during EDPE with Pro-Explora. However, an instruction to spend 10 minutes on the task is provided. During the case study, the participants used Pro-Explora once for EDPE, rated its outputs above average, and requested access for continued use which is granted. Further trials would be necessary to monitor the subsequent rating for Pro-Explora and ensure an increased rating. The uniqueness of Pro-Explora framed EDP is based on a duplication recognition search in the original corpus used in generating the EDP. This search is not extended to the google and patent databases which are popular for verifying uniqueness. However, during a pilot test, a manual search on google returned no duplicate for any Pro-Explora framed EDP.

Although BERT and LSTM technologies are potential computational technologies for EDPE (Section 2.3), they have not been used to compare with the MM used in this paper. Being in its infancy (Version 1), Pro-Explora will be improved further based on the feedback received from the participants. This will include optimizing its outputs and exploring other related NLP technologies including BERT and LSTM. Data collection for Pro-Explora database will continue, and its model will be updated continuously.

7 CONCLUSIONS

In this paper, case study-based evidence is provided to highlight the lack of attention on EDPE - an important aspect of engineering design at early-stage product design and development. Albeit there are few studies on the lack of attention on EDPE, a study providing empirical evidence and determinants for it lacks. The natural approaches related to EDPE are investigated including the “garbage can” model and serendipity phenomenon. Some challenges and natural limitations associated with the natural EDPE approach are identified including cognitive fatigue. This suggests that computational support could be advantageous in the process. In response, a data-driven computational EDPE framework and support tool – Pro-Explora are presented. The tool is the first-of-its-kind computational technology that mimics the natural EDPE process. It is based on a synergy of the MM and some big data technologies including ML and NLP.

A case study is conducted with 43 participants including novice and experienced design engineers. During the case study, the participants could not distinguish EDP framed by Pro-Explora when presented alongside naturally framed ones. Using Pro-Explora as support, novice and experienced participants come up with at least 5 new EDPs in about 10 minutes. This would be difficult or impossible with the natural EDPE approach. The overall average rating provided by the participants on the usefulness of Pro-Explora framed EDP is 7.74 out of 10. This is promising for accelerated innovations and inventions in the industry. Further, the result shows that over 50% of the participants in the case study did not receive any formal teaching on creativity in academia. This highlights the importance of focusing on teaching creativity in engineering design-related disciplines which is fundamental in EDPE.

ACKNOWLEDGMENT

All the students and professionals including academics and inventors who contributed to this study by participating in the case study are hereby acknowledged.

REFERENCES

[1] Olewnik, A., Yerrick, R., Simmons, A., Lee, Y., and Stuhlmiller, B., 2020, "Defining open-ended problem solving through problem typology framework," International Journal of Engineering Pedagogy (IJEP), 10(1).

[2] Stevens, R., Johri, A., and O’connor, K., 2014, "Professional engineering work," Cambridge handbook of engineering education research, pp. 119-137.

[3] Vale, J., Gordon, K., Kirkscey, R., and Hill, J., 2020, "Student and Faculty Perceptions of Capstone Purposes: What can Engineering Learn from Other Disciplines?," Proceedings of the Canadian Engineering Education Association (CEEA).

[4] Taura, T., 2015, Principia Designae-Pre-Design, Design, and Post-Design: Social Motive for the Highly Advanced Technological Society, Springer.

[5] Walter, C., and Richards, E., 1996, "Engineering and the law [custodians of information]," IEEE Engineering in Medicine and Biology Magazine, 15(5), pp. 138-139.

[6] Jaussi, K. S., and Topaloglu, E., 2020, "Intentionality," Encyclopedia of Creativity (Third Edition), S. Pritzker, and M. Runco, eds., Academic Press, Oxford, pp. 672-677.

[7] Getzels, J. W., 1979, "Problem finding: A theoretical note," Cognitive Science - A Multidisciplinary Journal, 3(2), pp. 167-172.

[8] Obieke, C. C., Milisavljevic-Syed, J., and Han, J., 2021, "Data-driven creativity: computational problem-exploring in engineering design," Proceedings of the Design Society, 1, pp. 831-840.

[9] NRC - National Research Council, 2012, A framework for K-12 science education: Practices, crosscutting concepts, and core ideas, National Academies Press.

[10] NAE - National Academy of Engineering, U. S., 2004, The engineer of 2020: Visions of engineering in the new century, National Academies Press Washington, DC.

[11] Obieke, C., Milisavljevic-Syed, J., and Han, J., 2020, "Supporting Design Problem-exploring with Emergent Technologies," Procedia CIRP, 91, pp. 373-381.

[12] Abdulla, A. M., Paek, S. H., Cramond, B., and Runco, M. A., 2018, "Problem Finding and Creativity: A Meta-Analytic Review," Psychology of Aesthetics, Creativity, and the Arts, 14(1), pp. 3-14.

[13] Heimlich, J., Wasserman, D., and Hayde, D., 2014, "Human Plus: Real Lives+ Real Engineering," Center for Research and Evaluation.

[14] Karwowski, M., Jankowska, D. M., Brzeski, A., Czerwonka, M., Gajda, A., Lebuda, I., and Beghetto, R. A., 2020, "Delving into creativity and learning," Creativity Research Journal, 32(1), pp. 4-16.

[15] Design Council, 2018, "The Design Economy 2018: The state of design in the UK," Design Council.

[16] Mullen, C. A., 2018, Creativity Under Duress in Education?: Resistive Theories, Practices, and Actions, Springer.

[17] AlMaian, R. Y., "Analysis of the stakeholders of engineering education system to improve the creativity of engineering education," Proc. 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, pp. 110-114.

[18] Chulvi, V., Sonseca, Á., Mulet, E., and Chakrabarti, A., 2012, "Assessment of the relationships among design methods, design activities, and creativity," Journal of Mechanical Design, 134(11), p. 11.

[19] Miller, S. R., Hunter, S. T., Starkey, E., Ramachandran, S., Ahmed, F., and Fuge, M., 2021, "How should we measure creativity in engineering design? a comparison between social science and engineering approaches," Journal of Mechanical Design, 143(3).

[20] Charyton, C., 2015, Creativity and innovation among science and art: A discussion of the two cultures, Springer.

[21] Gross, D. P., "Hiding in Plain Sight: The Identification and (Humble) Origins of General Purpose Technologies."," Proc. One Hudred Flowes" Conference. All-UC Group in Economic History. Cosponsored by the Berkeley Economic History Lab (BEHL) and the Dept. of Economics, UC Berkeley, Berkeley, CA, USA.

[22] Shneiderman, B., 2007, "Creativity support tools: Accelerating discovery and innovation," Communications of the ACM, 50(12), pp. 20-32.

[23] Nicholl, B., and McLellan, R., "The Contribution of Product Analysis to Fixation in Students' Design and Technology Work," Proc. The Design and Technology Association International Research Conference 2007, p. 71.

[24] Harris, S. D., and Zeisler, S., 2002, "Weak signals: Detecting the next big thing," The Futurist, 36(6), p. 21.

[25] Fischer, G., 1994, "Turning breakdowns into opportunities for creativity," Knowledge-Based Systems, 7(4), pp. 221-232.

[26] Einstein, A., and Infeld, L., 1938, The evolution of physics, Simon & Schuster, New York.

[27] Lee, J. W., Daly, S. R., Huang-Saad, A., Rodriguez, G., and Seifert, C. M., 2020, "Cognitive strategies in solution mapping: How engineering designers identify problems for technological solutions," Design studies, 71, p. 100967.

[28] Alter, S., and Dennis, A. R., 2002, "Selecting research topics: Personal experiences and speculations for the future," Communications of the Association for Information Systems, 8(1), p. 21.

[29] Dennis, A. R., and Valacich, J. S., 2001, "Conducting experimental research in information systems," Communications of the association for information systems, 7(1), p. 5.

[30] Martin, J., 1982, "A garbage can model of the research process," Judgment calls in research, pp. 17-40.

[31] Russ, S. W., and Hoffmann, J. D., 2020, "Associative Theory☆," Encyclopedia of Creativity (Third Edition), S. Pritzker, and M. Runco, eds., Academic Press, Oxford, pp. 76-82.

[32] Banan, S., Ridwan, M., and Adisaputera, A., 2020, "A Study of Connectionism Theory," Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, 3(3), pp. 2335-2342.

[33] Gall, M., Gall, J., and Borg, W., 2007, "edition 8," Educational research: An introduction. Boston. Pearson/Allyn & Bacon.

[34] Darbellay, F., 2020, "Serendipity," Encyclopedia of Creativity (Third Edition), S. Pritzker, and M. Runco, eds., Academic Press, Oxford, pp. 470-474.

[35] Mednick, S., 1962, "The associative basis of the creative process," Psychological review, 69(3), p. 220.
[36] Dawes, H., Dawes, C., Martin, G., and Macfarlane, A., 2006, "Making Things from New Ideas," History of Technology Volume 26, 2005: Including Special Issue: Engineering Disasters, 26, p. 1.
[37] Ishikawa, A., 2010, "Discovery, invention and serendipity," Chinese Business Review, 9(11), p. 61.
[38] Hubbell, M., Hard, S., Boots, M., Clarke, M. A., and Smith, J. E., "Pitch Stability of an Unpowered Ground Effect Vehicle," Proc. ASME International Mechanical Engineering Congress and Exposition, pp. 191-199.
[39] Thaler, S., 2020, "The creativity machine paradigm," Encyclopedia of creativity, invention, innovation, and entrepreneurship, E. G. Carayannis, ed., Springer Nature Switzerland AG, pp. 650 - 658.
[40] Revell, T., and Andersen, K., 2021, "THINGS: IMAGINING WITH," Designing Smart Objects in Everyday Life: Intelligences, Agencies, Ecologies, p. 57.
[41] Heath, S., and Shine, B., 2021, "Teaching Techniques to Facilitate Time Management in Remote and Online Teaching," Journal of Teaching and Learning with Technology, 10(1).
[42] Miller, G. A., 1956, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychological review, 63(2), p. 81.
[43] Jones, D. M., "The 7±2 urban legend," Proc. MISRA C 2002 conference www. knosof. co. uk/cbook/misart. pdf.
[44] Marois, R., and Ivanoff, J., 2005, "Capacity limits of information processing in the brain," Trends in cognitive sciences, 9(6), pp. 296-305.
[45] Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., and Roelands, B., 2017, "The effects of mental fatigue on physical performance: a systematic review," Sports medicine, 47(8), pp. 1569-1588.
[46] Smith, M. R., Coutts, A. J., Merlini, M., Deprez, D., Lenoir, M., and Marcora, S. M., 2016, "Mental fatigue impairs soccer-specific physical and technical performance," Med Sci Sports Exerc, 48(2), pp. 267-276.
[47] Boden, M. A., 2007, "Creativity in a nutshell," Think, 5(15), pp. 83-96.
[48] Hurley, P. J., 2021, "Reconceptualizing Ego Depletion as Transient Cognitive Fatigue," Available at SSRN 3797263.
[49] Stanton, P., 2018, Conscious Creativity: Look, Connect, Create, Leaping Hare Press.
[50] Ariza-Montes, A., Arjona-Fuentes, J. M., Han, H., and Law, R., 2017, "Employee responsibility and basic human values in the hospitality sector," International Journal of Hospitality Management, 62, pp. 78-87.
[51] Marcora, S. M., Staiano, W., and Manning, V., 2009, "Mental fatigue impairs physical performance in humans," Journal of applied physiology, 106(3), pp. 857-864.
[52] Southern, S., and Domzalski, S., "Developing Intuition: The Key to Creative Futures Research," Proc. Annual Meeting of the American Educational Research Association, p. 105.
[53] Chakrabarti, A., "Towards a measure for assessing creative influences of a creativity technique," Proc. DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm.

[54] Sommerville, I., 2011, "Software engineering 9th Edition," ISBN-10, 137035152, p. 18.
[55] Bridle, J., 2018, New dark age: Technology and the end of the future, Verso Books.
[56] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., 2018, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805.
[57] Arras, L., Arjona-Medina, J., Widrich, M., Montavon, G., Gillhofer, M., Müller, K.-R., Hochreiter, S., and Samek, W., 2019, "Explaining and interpreting LSTMs," Explainable ai: Interpreting, explaining and visualizing deep learning, Springer, pp. 211-238.
[58] Wang, A., and Cho, K., 2019, "Bert has a mouth, and it must speak: Bert as a markov random field language model," arXiv preprint arXiv:1902.04094.
[59] Banerjee, A., Fuchsbauer, G., Peikert, C., Pietrzak, K., and Stevens, S., "Key-homomorphic constrained pseudorandom functions," Proc. Theory of Cryptography Conference, Springer, pp. 31-60.
[60] Meyn, S., and Tweedie, R. L., 2009, Markov Chains and Stochastic Stability, Cambridge University Press, Cambridge.
[61] Schurz, G., 1996, "Kinds of unpredictability in deterministic systems," Law and prediction in the light of chaos research, Springer, pp. 123-141.
[62] Krishnakumar, S., Berdanier, C., Lauff, C., McComb, C., and Menold, J., 2022, "The story novice designers tell: How rhetorical structures and prototyping shape communication with external audiences," Design Studies, 82, p. 101133.
[63] Mehta, P., Malviya, M., McComb, C., Manogharan, G., and Berdanier, C. G., 2020, "Mining design heuristics for additive manufacturing via eye-tracking methods and hidden Markov modeling," Journal of Mechanical Design, 142(12).
[64] McComb, C., Cagan, J., and Kotovsky, K., 2017, "Capturing human sequence-learning abilities in configuration design tasks through Markov chains," Journal of Mechanical Design, 139(9).
[65] Rahman, M. H., Xie, C., and Sha, Z., 2021, "Predicting sequential design decisions using the function-behavior-structure design process model and recurrent neural networks," Journal of Mechanical Design, 143(8).
[66] Anandakumar, H., Arulmurugan, R., and Onn, C. C., 2019, "Big Data Analytics for Sustainable Computing," Mobile Networks and Applications, 24(6), pp. 1751-1754.
[67] Thanaki, J., 2017, Python natural language processing, Packt Publishing Ltd.
[68] IPO - Intellectual Property Office UK, 2017, IP Basics, Crown Copyright, Newport.
[69] Langford, C. A., and Pearce, P. F., 2019, "Increasing visibility for your work: The importance of a well-written title," American Association of Nurse Practitioners, 31(4), pp. 217-218.
[70] Greenspan, Y. F., 2016, A guide to teaching elementary science: Ten easy steps, Brill Sense.
[71] Sri, M., 2021, "NLP in Customer Service," Practical Natural Language Processing with Python, Springer, pp. 13-63.
[72] Hajba, G., 2018, "Website Scraping with Python: Using BeautifulSoup."
[73] Mikes, A., Edmonds, K., Stone, R. B., and DuPont, B., "Optimizing an algorithm for data mining a design repository to automate functional modeling," Proc. International

Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, p. V11AT11A014.

[74] Plötz, T., and Fink, G. A., 2011, Markov Models for Handwriting Recognition, Springer.

[75] Yadav, P., Ladha, S., Deshpande, S., and Curry, E., "Computational model for urban growth using socioeconomic latent parameters," Proc. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, pp. 65-78.

[76] Rabiner, L. R., 1989, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, 77(2), pp. 257-286.

[77] Privault, N., 2013, Understanding Markov Chains, Springer.

[78] Eagle, A., 2005, "Randomness is unpredictability," The British Journal for the Philosophy of Science, 56(4), pp. 749-790.

[79] Sheskin, T. J., 2011, Markov chains and decision processes for engineers and managers, CRC Press, New York.

[80] Koronis, G., Casakin, H., Silva, A., and Kang, J. K. S., 2021, "The influence of design brief information on creative outcomes by novice and advanced students," Proceedings of the Design Society, 1, pp. 3041-3050.

[81] Bahadoran, Z., Mirmiran, P., Kashfi, K., and Ghasemi, A., 2019, "The principles of biomedical scientific writing," International Journal of Endocrinology and Metabolism, 17(4).

[82] Awasthi, R., and Kulkarni, G. T., 2014, "It's All In The Title: Writing a Proper Title to the Paper," Journal of Chronotherapy and Drug Delivery, 5(3), pp. S13 - S16.

[83] Hays, J. C., 2010, "Eight recommendations for writing titles of scientific manuscripts," Public Health Nursing, 27(2), pp. 101-103.

[84] Wohl, H., 2022, "Innovation and creativity in creative industries," Sociology Compass, 16(2), p. e12956.

[85] Godart, F., Seong, S., and Phillips, D. J., 2020, "The sociology of creativity: Elements, structures, and audiences," Annual Review of Sociology, 46, pp. 489-510.

[86] Han, J., Forbes, H., and Schaefer, D., "An exploration of the relations between functionality, aesthetics and creativity in design," Proc. Proceedings of the Design Society: International Conference on Engineering Design, Cambridge University Press, pp. 259-268.

[87] Polanyi, M., 1958, Personal Knowledge: Towards a post critical philosophy, Routledge & Kegan Paul, London.

[88] Jørgensen, U., "Engineering design competences–controversial relations between techno-science discipline and engineering practice domains pointing to new foundations for engineering knowledge," Proc. INES Workshop, Virginia Tech, Citeseer, pp. 1-16.

[89] Malhotra, N. K., 2006, "Questionnaire design and scale development," The handbook of marketing research: Uses, misuses, and future advances, pp. 83-94.

[90] Kulkarni, A., and Shivananda, A., 2021, Natural language processing recipes : unlocking text data with machine learning and deep learning using Python / Akshay Kulkarni, Adarsha Shivananda, Apress, Berkeley, CA.

[91] ABET - Accreditation Board for Engineering and Technology, 2019, "Criteria for Accrediting Engineering Programs 2020-2021," Engineering Accreditation Commission, ABET, Baltimore.

[92] Kazerounian, K., and Foley, S., 2007, "Barriers to creativity in engineering education: A study of instructors and students perceptions," Journal of Mechanical Design, 129(7), pp. 761–768.
[93] Celik, Y., 2019, "Rules for Doing and Managing Research Project," https://dergi.biruni.edu.tr/wp-content/uploads/2019/12/2-Yusuf-Celik.pdf.
[94] Gall, M. D., Borg, W. R., and Gall, J. P., 1996, Educational research: An introduction, Longman Publishing.
[95] Wood, D., 2013, "Fichtes Conception of Infinity in the Bestimmung des Menshcen," State University of New York Press, pp. 155-171.
[96] Saad, I., and Chakhar, S., "Aggregation procedure based on majority principle for collective identification of firm’s crucial knowledge," Proc. World Summit on Knowledge Society, Springer, pp. 346-352.

Figure Captions List

Table Caption List

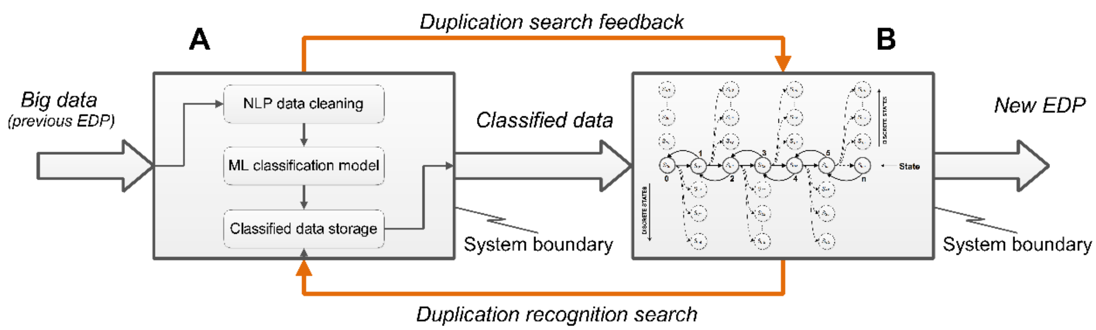

Fig. 1 Big data-driven computational EDPE framework

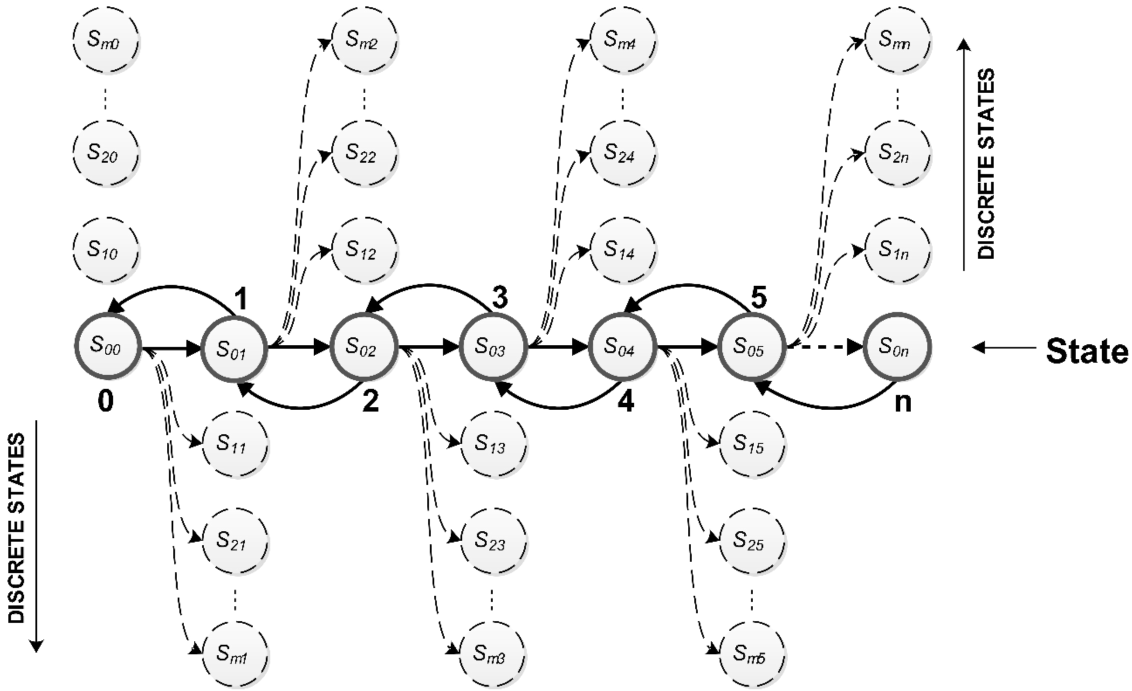

Fig. 2 A theoretical model for EDP sequencing

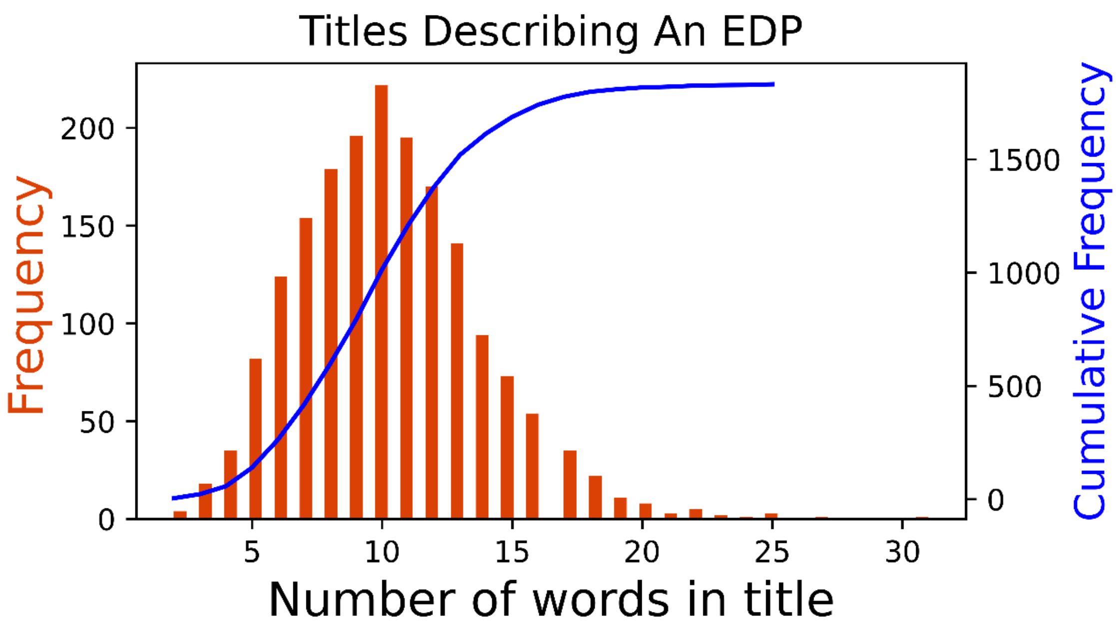

Fig. 3 Word count of titles describing EDP and Non-EDP

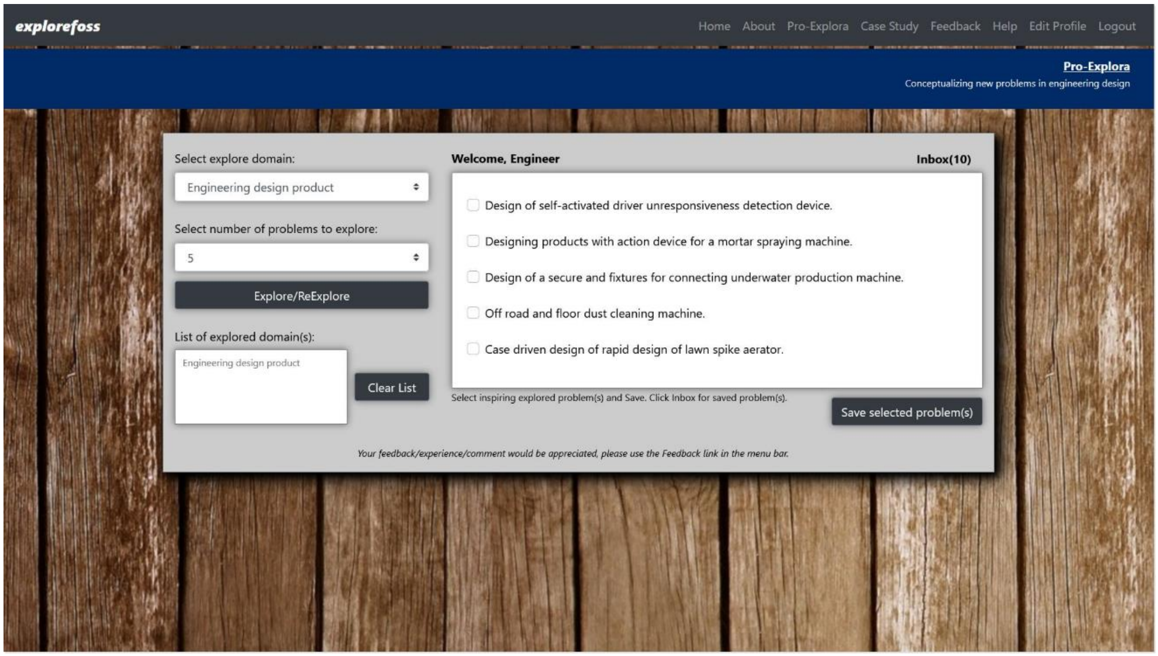

Fig. 4 Pro-Explora GUI with some framed design problems

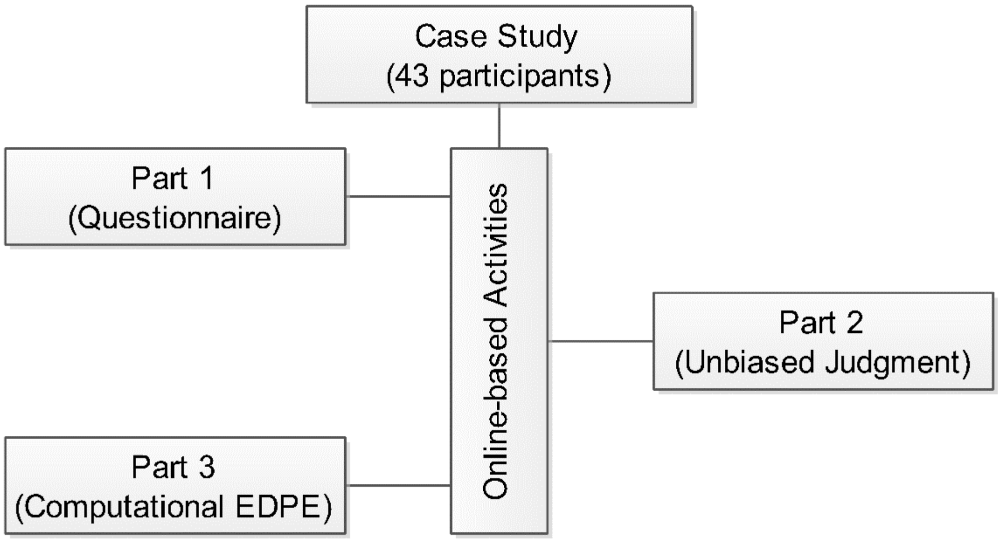

Fig. 5 Overview of data collection approach

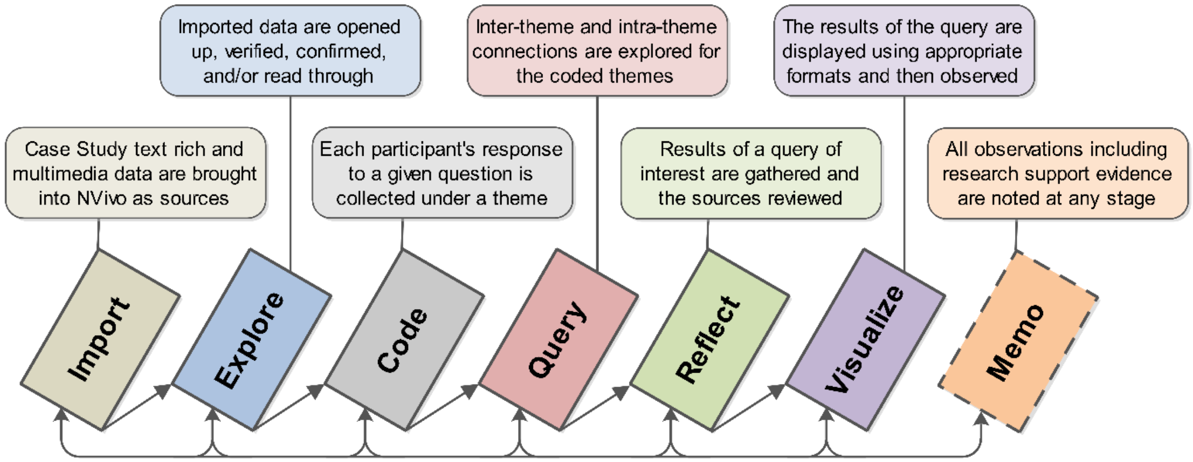

Fig. 6 Case study qualitative analysis workflow

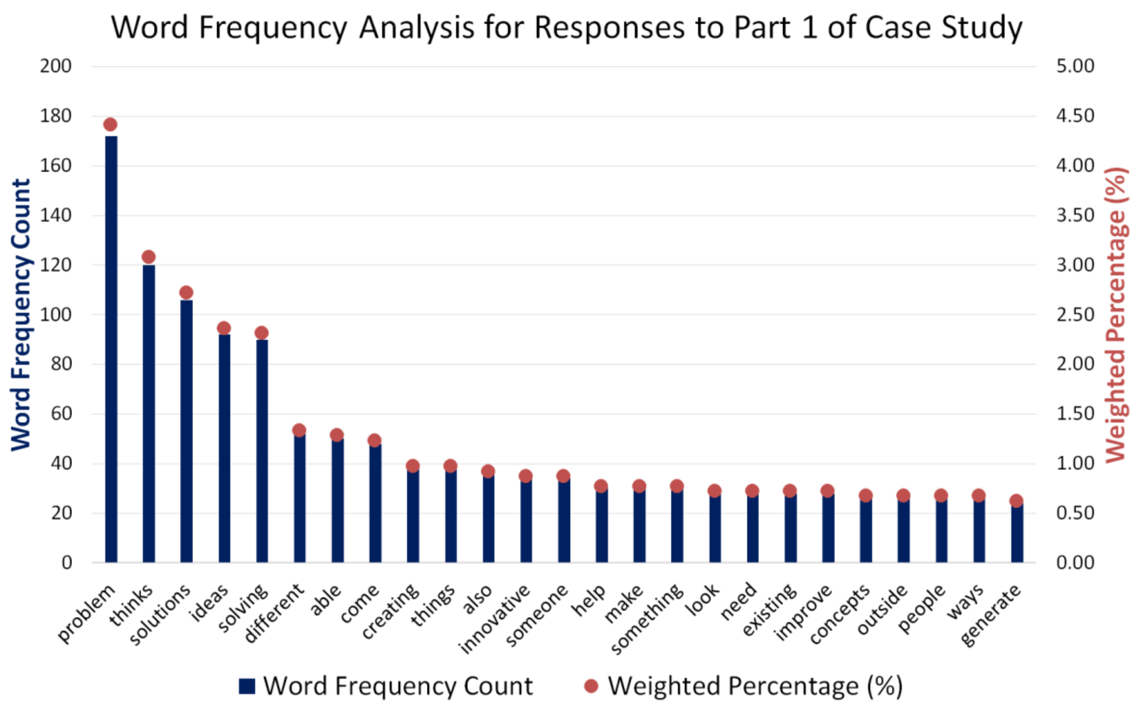

Fig. 7 Most frequent words in explaining creativity in engineering design

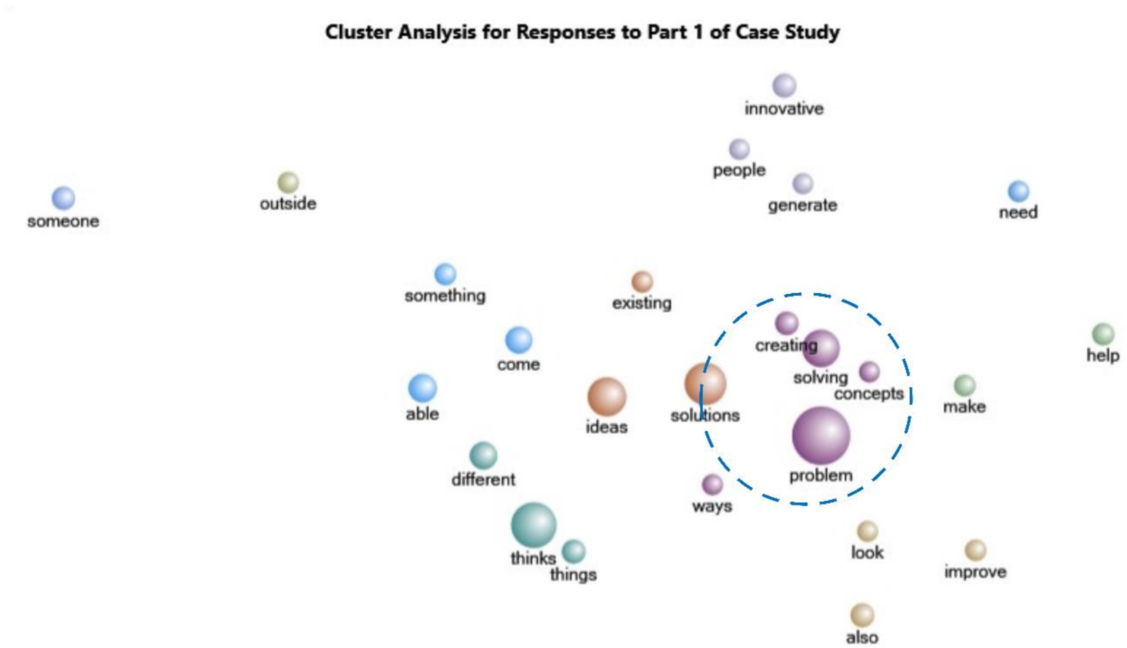

Fig. 8 Word cluster showing a relationship in words explaining creativity

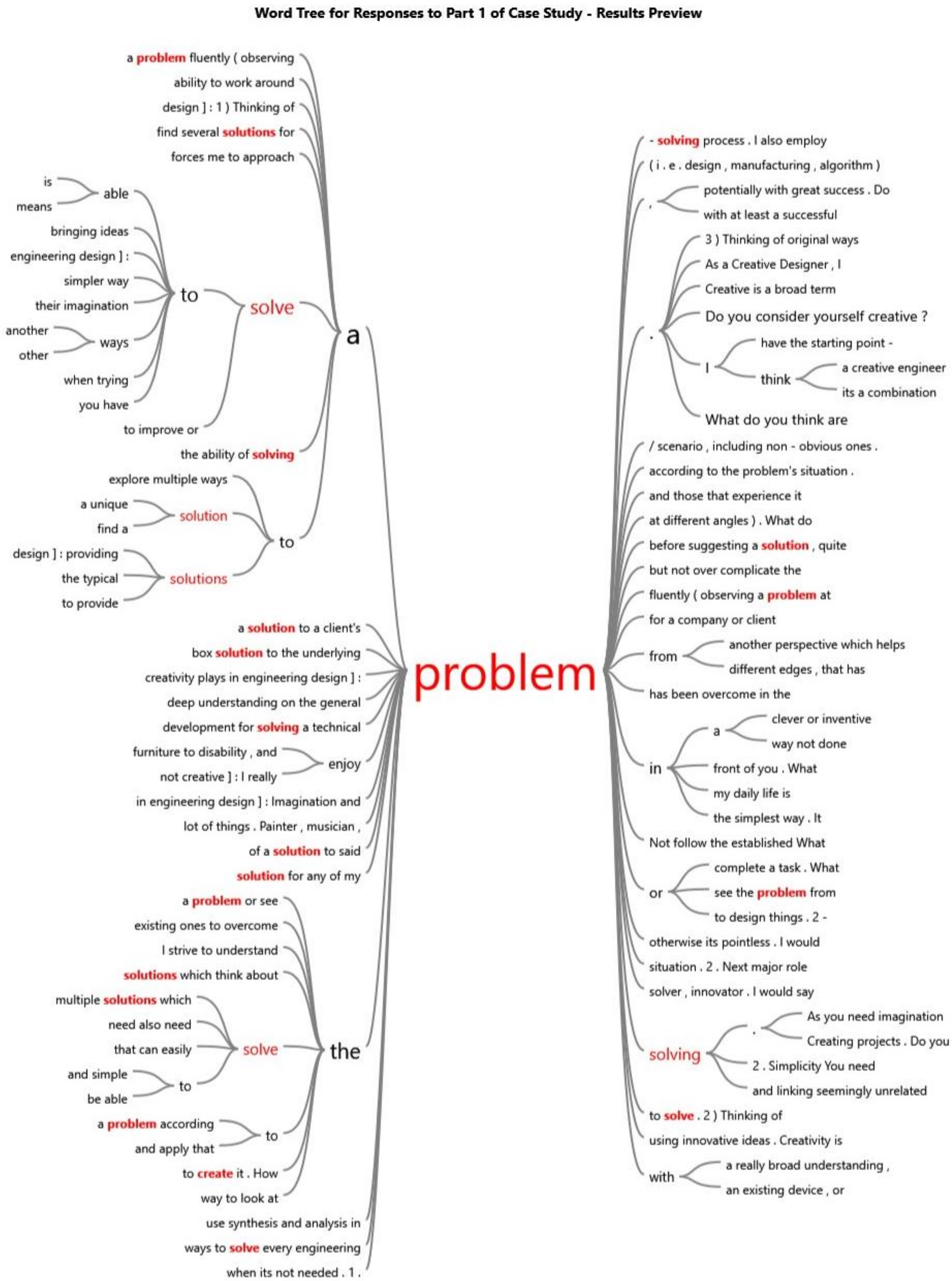

Fig. 9 Word Tree result for a text query search

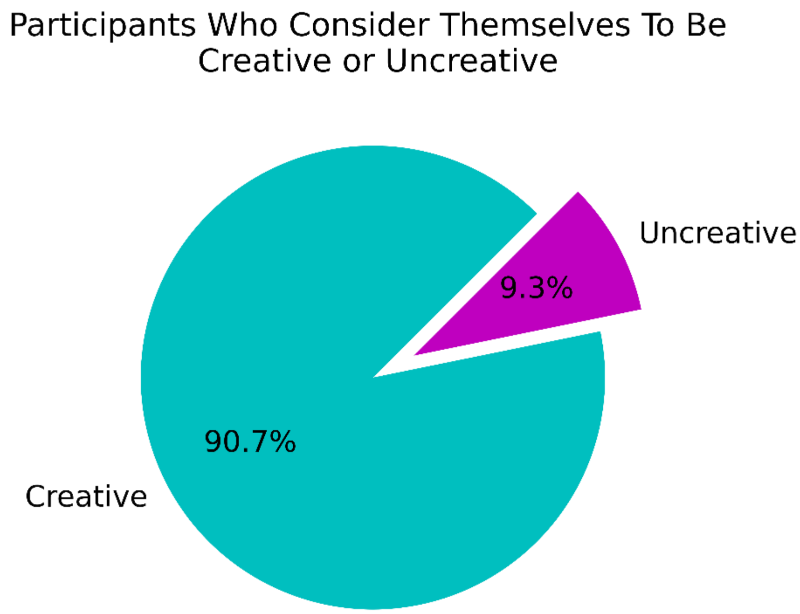

Fig. 10 Participants' responses to their creativity ability

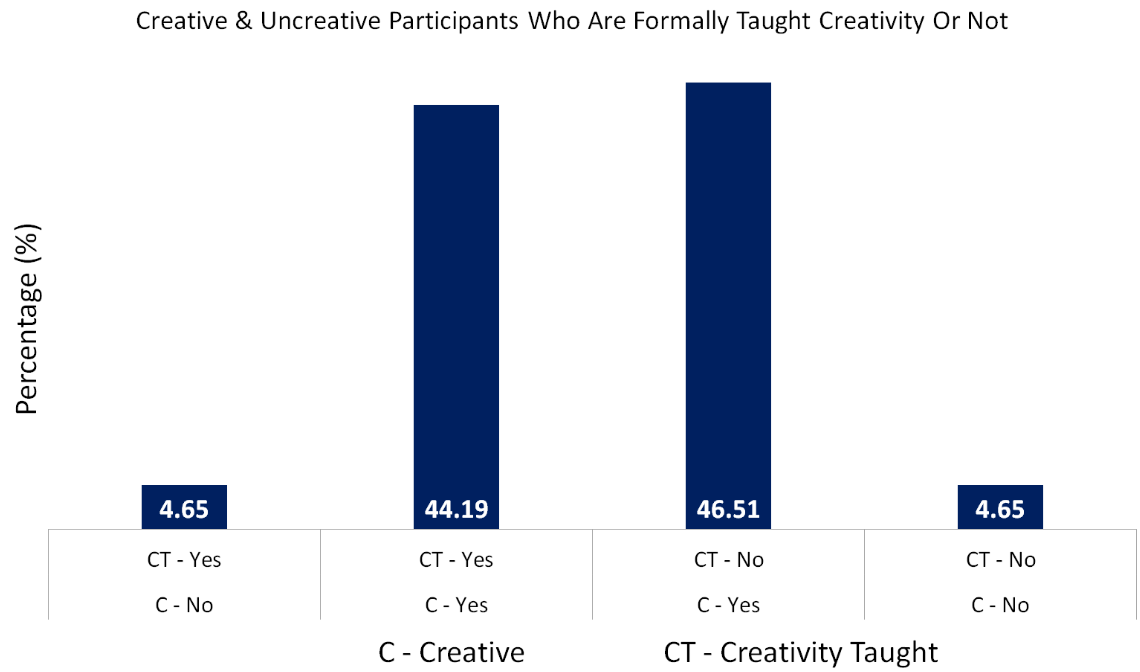

Fig. 11 An insight into creativity teaching in academia

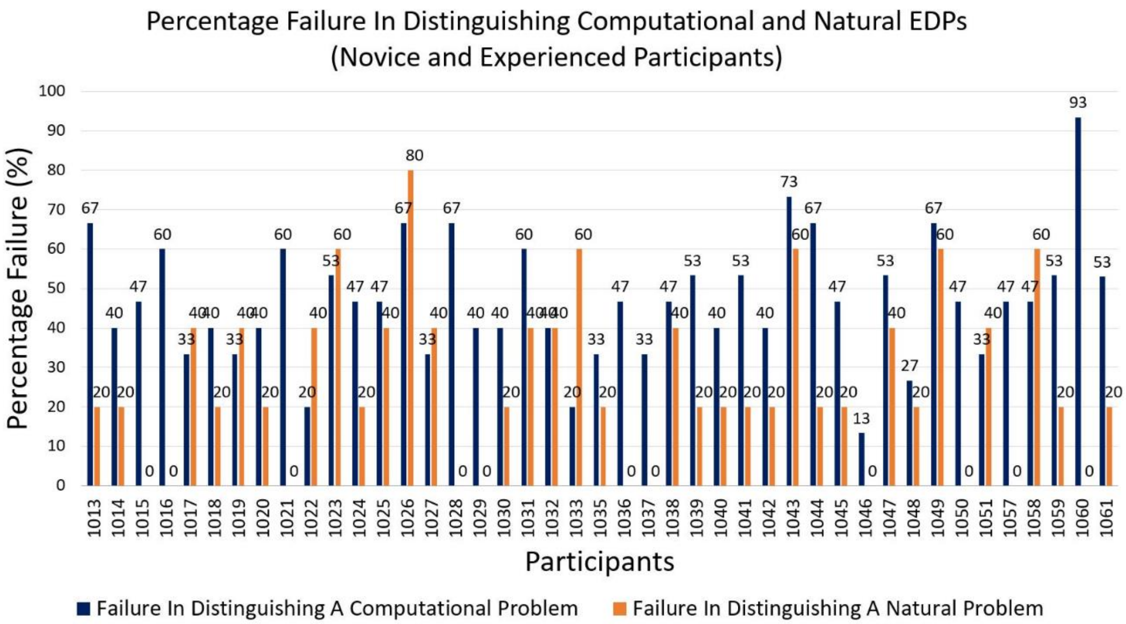

Fig. 12 Result of Part 2 (Unbiased Judgement) of the case study

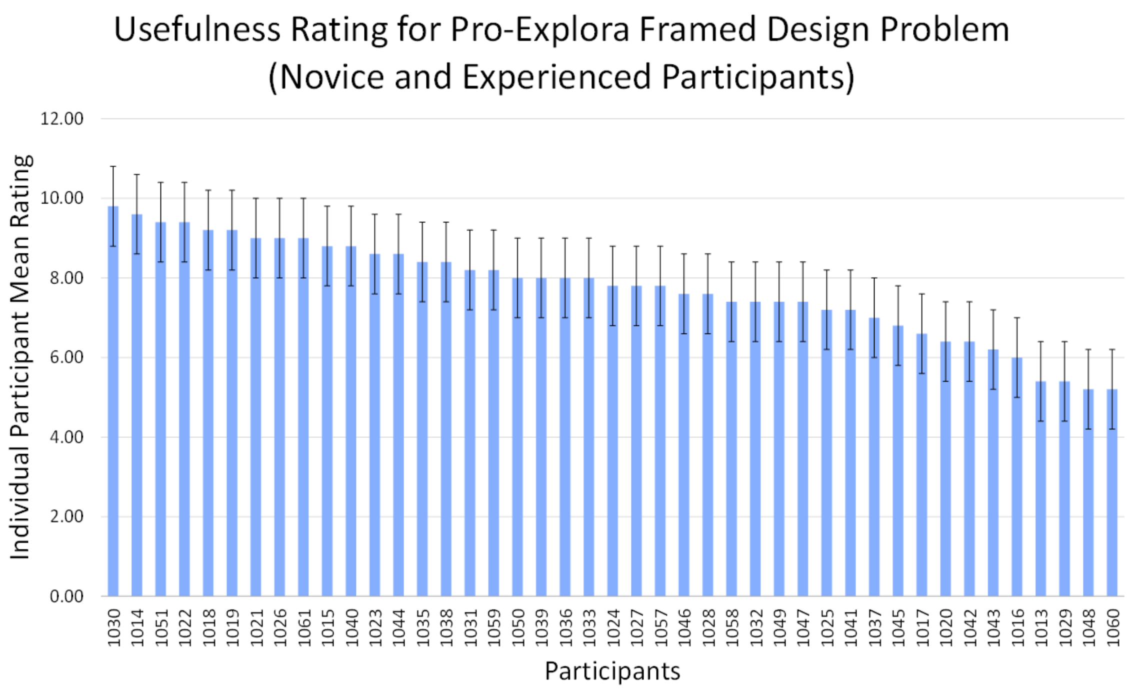

Fig. 13 Participants' usefulness rating for Pro-Explora framed EDP (with ±1 SE bar)

Table 1 Comparison of algorithms for natural and computational EDPE

Natural EDPE approach	Computational EDPE approach
Identify an EDP of societal relevance by accident (serendipity), stochastic synthesis (“garbage can” model), logical progression (“rational” model), and/or conceptualization (apophenia)	Frame an EDP of societal relevance by stochastic synthesis of big data, computational technologies (data extraction, ML, NLP), coding capabilities, connectionist theory, deterministic chaos, MM, BERT, and/or LSTM
Search manually for prior existence in relevant databases using search engines.	Make an automated search for prior existence in relevant databases using duplication recognitions.
Decide, subject to acceptance by the society or a relevant authority	Decide, subject to a design engineer's acceptance.

Table 2 Case study participants' detail

Participants	Gender			Academic Qualification (Obtained/in view)			Experience (Years)	
	Male	Female	Total	B	M	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

B – Bachelors, M – Masters

Table 3 Questionnaire for Part 1 of the case study

Questionnaire themes
(a) What does it mean to be creative? (b) Major are the roles of creativity in engineering design? (c) Do You Consider Yourself Creative? (d) Why are you creative or uncreative? (e) Were you taught creativity at University or at work?

Table 4 A sample set of 20 EDP for participants

Naturally and computationally framed EDP
1. Design of a mechanical intrusive force detection device.
2. To design a portable water distillation device.
3. A sustainable packaging design for wine.
4. Designing an interactive interface for collaborative engineering design.
5. A design of an automatic bottle opener.
6. Towards intelligent emotion detection system for video traffic surveillance.
7. Ai-based learning models for video traffic surveillance.
8. Design and material properties to minimize biofilm deposits.
9. Design of human-powered hybrid electric-power shovel for the physically challenged.
10. Design of self-reconfigurable production equipment during operation.
11. Anti riot drone without traffic lights.
12. Investigation of anomaly detection in a critical materials.
13. Design of a self-timing solar seawater desalination machine.
14. Staging co-design for reverse modeling of product development.
15. Detecting aggressive driving behavior using scilab.
16. Design of remote intelligent home finance software.
17. Designing products by artificial intelligence design approach.
18. A computationally efficient real-time vehicle and speed detection using federated learning.
19. Automatic mechanical footstep power tiller machine.
20. Design of production information retrieval system.

Table 5 Novice and Experienced participants

Novice	Experienced

1013	1019	1026	1058	1022	1035	1041	1048
1014	1020	1027		1028	1036	1042	1049
1015	1021	1029		1030	1037	1044	1057
1016	1023	1043		1031	1038	1045	1059
1017	1024	1050		1032	1039	1046	1060
1018	1025	1051		1033	1040	1047	1061

Table 6 Correlation between experience and distinguishing a computational EDP

	P-value		Pearson's r	
	Novice	Experienced	Novice	Experienced
Experience vs Failed C differentiation	0.19	0.62	0.32	0.11
Experience vs Failed N differentiation	0.38	0.56	0.21	0.13

C – computationally framed EDP N – naturally framed EDP

Table 7 Cosine similarity assessment result

	EDP6	EDP7	EDP8	EDP9	EDP10	Q*
EDP1	0.8163	0.7000	0.7461	0.8465	0.8589	0.6163
EDP2	0.7164	0.6240	0.7838	0.6794	0.6686	0.5628
EDP3	0.6700	0.6656	0.6258	0.6776	0.6872	0.5249
EDP4	0.7842	0.7275	0.7191	0.7500	0.7710	0.5371
EDP5	0.6163	0.5767	0.5481	0.7164	0.7306	0.4750

Table 8 Correlation between experience and rating of a computational EDP

	P-value	Pearson's r
Experience vs Usefulness Rating (Novice)	0.16	-0.34
Experience vs Usefulness Rating (Experienced)	0.69	-0.08