Computers in Industry 129 (2021) 103447

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Data science for engineering design: State of the art and future directions

Filippo Chiarello*, Paola Belingheri, Gualtiero Fantoni

University of Pisa, Italy

ARTICLE INFO

Article history:
Received 19 September 2020
Received in revised form 8 January 2021
Accepted 17 March 2021
Available online 25 March 2021

Keywords:
Engineering design
Data science
Literature review
Scoping review
State of the art

ABSTRACT

Engineering design (ED) is the process of solving technical problems within requirements and constraints to create new artifacts. Data science (DS) is the inter-disciplinary field that uses computational systems to extract knowledge from structured and unstructured data. The synergies between these two fields have a long story and throughout the past decades, ED has increasingly benefited from an integration with DS.

We present a literature review at the intersection between ED and DS, identifying the tools, algorithms and data sources that show the most potential in contributing to ED, and identifying a set of challenges that future data scientists and designers should tackle, to maximize the potential of DS in supporting effective and efficient designs. A rigorous scoping review approach has been supported by Natural Language Processing techniques, in order to offer a review of research across two fuzzy-confining disciplines.

The paper identifies challenges related to the two fields of research and to their interfaces. The main gaps in the literature revolve around the adaptation of computational techniques to be applied in the peculiar context of design, the identification of data sources to boost design research and a proper featurization of this data. The challenges have been classified considering their impacts on ED phases and applicability of DS methods, giving a map for future research across the fields. The scoping review shows that to fully take advantage of DS tools there must be an increase in the collaboration between design practitioners and researchers in order to open new data driven opportunities.

1. Introduction

Engineering Design (ED) can be defined as the process of solving technical problems within requirements and constraints to create new products (Pahl et al., 2007). It is an activity that combines both analytical and creative skills, which make it essentially a social endeavor, where participants from different backgrounds must exchange knowledge and information, negotiate, and reach viable compromises (Carleton and Leifer, 2009; De Toni and Nassimbeni, 2001; Hill et al., 2001). Recently, ED has been increasingly integrated with Data Science (DS) methodologies that are attempting to solve some of the challenges faced by modern designers. This article therefore analyses in depth the developments at the intersection between these two literature domains.

The discipline of ED has undergone several changes in the past decades, which have preceded or followed advances in manufacturing capabilities. In the 1960s, ED was mainly focused on combining systems design, product design, mechanical analysis and mechanisms design, and design components, while traditional manufacturing relied on manual machinery and the concept of specialization and mass production. In this period, important steps were taken by researchers in the fields of quality control and value analysis (Miles, 2015).

By the 1970s, new concepts such as learning-by-doing, prototyping, and testing rose to the fore. In the late 1970s two important journals related to both ED and Information and Communication Technologies emerged, i.e. *Design Studies* and *Computers in Industry*.

The 1980s saw the rise of mechatronics, combining notions of mechanical design with electrical engineering and computer programming, setting the first premise for the combination of design and Big Data (Carleton and Leifer, 2009). According to De Mauro et al. (2015: 103) "Big Data is the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" and the combination of engineering and computer programming was a necessary step in developing the appropriate analytical methods for this revolution. Among these analytical methods, a prominent role was carved out for Artificial Intelligence

* Corresponding author.
E-mail address: filippo.chiarello@unipi.it (F. Chiarello).

https://doi.org/10.1016/j.compind.2021.103447

(AI), that "applies advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions" (Gartner, 2018). Indeed, in the same period, the journals *Research in Engineering Design, Journal of Engineering* and *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* (AIEDAM) emerged. The latter is of particular interest for the present review, since it publishes works about significant AI theory and applications based on the most up-to-date research in all branches and phases of engineering (Pahl et al., 2007).

In the early 1990s, the instrumentation of the design process followed, which enabled more efficient benchmarking and redesign, coupled with the transition from traditional manufacturing to agile manufacturing that benefits from product modularity and the possibility to rapidly reconfigure production activities (Kusiak and Huang, 1996; Carleton and Leifer, 2009). Following the first steps in systematization taken in the 1980s, the 1990s started with the emergence of two important frameworks for ED: Function, Behaviour, Structure (FBS) theory (Umeda et al., 1990) and Axiomatic Design (Suh and Suh, 1990). In the second half of this decade also another important theory that attempts to systematize the ED process emerged, i.e. the Concept Knowledge-Theory (Hatchuel, 1996).

Through the beginning of this century, the inputs to the design process were expanded, with a stronger focus on customer requirements and the development of techniques such as Design Thinking (Goel and Shu, 2015). A stronger product and service personalization was advocated, reaching out to lead users to gain valuable inputs (Von Hippel, 1989), which led towards the trend of mass customization (Jiao and Zhang, 2005) and a larger focus on the management of complexity (Lindemann et al., 2008). In the same period, an interesting stream of publications started to study the use of external lexical sources (e.g. Wordnet, Wikipedia) in order to improve creativity and avoid fixation in the context of ED (Linsey et al., 2012).

Nowadays, design is recognized as a key element of the innovation process at-large, with a strong integration also with the domain of technology foresight (Anderson, 1997). Despite its long history, only in recent years has data-driven design become more prominent thanks to the recent developments in AI, focusing on the use of data coming from different design-related sources in order to make the design process more effective (Kusiak and Salustri, 2007). Furthermore, the influences of new societal challenges must be considered. As macro-trends such as climate change, population aging and pandemic phenomena become more prominent, new constraints are being placed upon design efforts such as sustainability and overall value (Iandoli et al., 2007; Isaksson et al., 2015; Umeda et al., 2012), positive social impact (Levine et al., 2016), product design's relationship to specific business models (Urbinati et al., 2018) or new design challenges in the post COVID-19 era (Fiorineschi et al., 2020).

With increased competition, and an increase in the digitalization of manufacturing, coupled with new methodologies to collect data on product characteristics, product performance and customer requirements, ED is often accompanied by big data (Wang and Alexander, 2015), the proper use of which can increase the efficiency of design processes, for example by detecting design defects, processing customer requirements, and increasing customer satisfaction (Gorgoglione et al., 2018; Wu et al., 2015). Design data is by nature multi-disciplinary, multi-dimensional and non-linear, (Romanowski et al., 2006) and it would be limiting to only consider its interaction with big data or machine learning. For this reason, we consider the broader field of DS, also considering that we are seeing an increase in the application of DS methods to ED and an increase in their sophistication.

DS bridges computer science, statistics, and domain knowledge to uncover the potential concealed in data. It is a growing field, and it is having an impact on several domains, including ED. In fact, DS is changing the way organizations are designing their products and is re-shaping the approaches adopted to handle data during the ED process. While using data-driven scientific methods to test new products in medicine, engineering, defense, and other safety-critical systems is common practice, using these methods in the design of consumer products is a relatively recent phenomenon.

In contrast to a large practical adoption of DS tools in ED, there is a lack of attention on the intersection between these two fields in the academic literature. In particular, no scientific framework exists to guide researchers interested in designing DS tools specifically for the ED process according to its peculiarities. Therefore, there is no theoretical foundation that allows for objective reasoning around the application of DS in the ED process. Furthermore, both DS and ED are complex fields by nature. They involve many sub-fields: engineering, computer sciences, social sciences, statistics, psychology and economics, to name a few. This makes it hard to understand their evolution and their intersections over time.

This article aims to highlight these gaps by providing a comprehensive Scoping Review of the literature on DS applications in the ED domain using both qualitative and quantitative approaches. It focuses on:

1 - the different ED phases and overarching processes that DS can serve: the ED process can be broken down into a series of phases that have different objectives. Indeed, it often starts with the collection of requirements from customers, followed by a period of creative thinking, and finally a further refinement and selection of ideas, prototypes, and solutions;
2 - the DS methodologies that have been applied in an ED setting: each ED phase, because of its different objectives, methodologies, and settings, can benefit from different DS methodologies that may be, to a varying degree, appropriate to tackle the issues or inefficiencies that ED professionals face in each phase.

The remainder of the paper is organized as follows. Section 2 illustrates the research methodology adopted for our Scoping Review. Section 3 presents our findings from the literature, relative to the main phases of ED and the main DS methodologies. Section 4 synthesizes the main themes and challenges, presenting the research gaps that should be addressed in further studies. Finally, Section 5 concludes.

2. Approach

In the context of multidisciplinary research, literature reviews are of particular interest, given their aim to provide a clear picture of existing knowledge and to support the research path for researchers from different backgrounds. Considering the novelty and the evolving rate of ED and DS, we decided to use a scoping review (Munn et al., 2018). A scoping review follows a structured approach in order to determine the coverage of a body of a field of study, indicating, among the other things, the volume of literature available in the field, its focus, how research is conducted, and which are the main gaps (Armstrong et al., 2011).

Following slightly modified version of a well-known framework in the context of scoping reviews (Arksey and O'Malley, 2005) the approach we used in order to map literature in the intersection between ED and DS is the following:

Stage 1 - identifying the research question: defining which aspects of the research question are particularly important.

Stage 2 - searching and extracting relevant scientific papers: retrieving primary studies suitable for answering the central research question.

Stage 3 - study selection: analyzing the articles retrieved in stage 2 in order to eliminate studies that did not address our central research question; we used our expertise in Natural Language Processing in order to make this phase more efficient, effective and reproducible.

Stage 4 - charting the data: synthesizing and interpreting the data filtered in stage 3 according to key issues and themes.

Stage 5 - summarizing and reporting: presenting our findings using both qualitative and quantitative evidence.

The five stages are described in further detail in the next sections.

2.1. Identifying the research question

The research question of the present review has been chosen to guide the strategies to build the knowledge base to analyze. Our research question is:

What does scientific literature tell us about the usage of Data Science tools in the context of Engineering Design?

We were aware that both ED and DS are ill-defined fields. We thus had to determine what phases of ED and what methodologies of DS were to be included in the review.

Different choices regarding the definition of the phases of ED and methodologies of DS are important for a proper scoping study. Vague definitions of ED and DS might make the review prone to introducing out-of-scope papers or to a too narrow focus. Following guidelines (Arksey & O'Malley, 2005), in stage 2 we maintained a wide approach in order to generate breadth of coverage. In stage 3, after having some sense of the volume and general scope of the field, we took decisions about how to properly filter bibliographic references. This has led to the identification of several meta-themes on which we can break down the main research questions. The meta themes are introduced in Subsection 2.3 and will be used to structure the discussion in Section 4.

2.2. Searching and extracting relevant scientific papers

The goal of this stage was to identify and collect articles that discuss Engineering Design. Two approaches were used, combining a recall-oriented and a precision-oriented one.

First, we collected all the scientific papers (journal and conference papers), published between 2000 and 2020 and containing the string "engineering design" in the title, abstract or keywords. We decided to use the Scopus database. This approach, executed in July 2020, led to the collection of 24,631 scientific papers. Considering the research question stated in Section 2.1, this set has a high recall, since it contains all the papers in the context of ED that use DS Tools, but also many papers in the same context that make no use of any of these tools.

Next, we collected all the scientific papers in the Scopus database, published between 2000 and 2020 and containing the string "design" in the name of the source (the name of the journal that published the study or the conference in which the study was presented) and containing in the title, abstract or keywords one of the following strings: "data science", "data analysis", "statistical analysis" or "artificial intelligence".

Finally, we merged the results and after removing the duplicated papers (scientific papers extracted with both approaches) we had a set of 25,620 articles. From this set, we selected only papers belonging to journals and conferences in the fields of Engineering, Computer Science, Business, Management and Accounting, Decision Sciences and Multidisciplinary (according to the ASJC classification), obtaining 20,766 documents.

The search was constrained to the last 20 years since even if DS has sub-fields, such as AI, that have roots in the early 1960s (Minsky, 1961), the specific field has emerged only in the early 2000s (Cleveland, 2001).

2.3. Study selection

The strategy pursued in stage 2 intentionally picked up a large number of papers, many of which potentially irrelevant for the scoping study. Design is a generic term used also in other fields with respect to engineering design (e.g., computer science, pharmaceutical design, drug design, lighting design). Even if some of these disciplines can bring interesting approaches to the discussion, the aim of stage 3 is to define the terminology for a correct scope, considering the research question stated in Section 2.1. In other terms, we needed an approach to eliminate studies that did not address our central research question.

The inclusion criteria used in our scoping study took into account:

- the ED phases addressed in the paper
- the DS tools used in the paper

After creating two lexicons of terms for the two criteria, the lexicons were used in order to filter the papers, i.e. only the papers containing at least one term for each lexicon (in the title, abstract or keywords) were considered.

Out of our original 20,766 references, 1024 were selected for stage 4. Sections 2.3.1 and 2.3.2 illustrate in greater detail the approach we followed in order to define the lexicons. It is important to stress here that the lexicons added a strong value to the present scoping review, since they have not only been used to filter the papers but also helped to segmented them into sub-topics and to guide the analysis that led to the results shown in Section 4.

2.3.1. Engineering design lexicon

A lexicon is a repertoire of words belonging to a specific semantic domain. It is an important linguistic resource needed for both shallow and deep language processing (Tjoa and Berger, 1993). Clearly defining a lexicon in the Engineering Design field is a known problem (Messac and Chen, 2000). Furthermore, for the present study we need to connect each word of the lexicon to a specific ED phase. In order to do so, we used a two-step (top-down, bottom-up) approach:

- Step 1: we revised the most important books in the field of ED, identifying the main phases in the ED process.
- Step 2: for each phase, we manually collected a set of words and multi-words, then used Natural Language Processing techniques to expand the lexicon.

During step 1 the following ED books were taken in consideration:

- Engineering design: a systematic approach (Pahl et al., 2007)
- Engineering design (Dieter and Schmidt, 2009)
- Product design and development (Ulrich, 2003)
- Engineering design process (Haik et al., 2018).

Since there is no common definition of the ED phases (Pahl et al., 2007), two authors independently identified 4 main ED phases, common to all the authors above. Furthermore, together with these vertical phases, we also identified horizontal phases (i.e. groups of activities that have a potential impact on every phase of the Engineering Design process). The results are summarized in Table 1,

Table 1
Number of papers in each ED phase that contains at least one entry in the DS Lexicon.

Design Phase	Number of Papers	% of Papers
1 - Problem Definition	157	15,33 %
2 - Conceptual Design	208	20,31 %
3 - Embodiment Design	382	37,30 %
4 - Detailed Design	125	12,21 %
Design Education	579	56,54 %
Design for X	24	2,34 %
Innovation & IP Management	325	31,74 %
Project Management	163	15,92 %
Prototyping	118	11,52 %
Sustainability	47	4,59 %

Table 2
Number and Percentage of papers citing Data Science methods.

Methods	N. of Papers	% of Papers
evolutionary algorithm	331	58,2 %
artificial neural networks	127	22,3 %
regression analysis	106	18,6 %
optimization algorithms	52	9,1 %
linear classifier	42	7,4 %
ensemble learning	22	3,9 %
bayesian methods	15	2,6 %
decision tree algorithms	14	2,5 %

Appendix A. Each identified phase will be described in greater detail in Section 4, considering each connected meta-theme.

During step 2, we revised the chapters of the selected books related to each phase and collected a preliminary lexicon of 84 words. We then used a word2vec approach based on Glove (Pennington et al., 2014) to find similar words and n-grams (a contiguous sequence of n words) contained in the articles under analysis. For each entry of the lexicon, we extracted the top-30 similar words in terms of cosine similarity, and manually selected the relevant ones. Thanks to this step we produced a lexicon of 334 words. For each word we kept the information regarding the relatedness to a given ED phase. The table that summarizes the selected ED phases with some exemplary entries of the network is shown in Table A1, Appendix A.

2.3.2. Data science lexicon

For this lexicon, our goal was to collect a list of DS techniques. Since DS has many broad sub-fields (e.g. machine learning, statistics, big data, data visualization) we used a similar approach to the one described in Section 2.3.1, mixing a top-down and a bottom up approach:

- Step 1: we collected a set of DS related lexicons available online,
- Step 2: we used Natural Language Processing techniques to expand the lexicons.

For step 1, we collected the following publicly available lexicons:

- The Machine Learning vocabulary developed by Google[1]
- The Glossary of common Machine Learning, Statistics and Data Science terms by Analytics Vidhya[2]
- Data Science Glossary by Kaggle[3]
- The Outline on Machine Learning by Wikipedia[4]

We merged these sources, manually revised them and after removing duplicates we obtained a list of 223 entries. We then used a word2vec approach based on Glove (Pennington et al., 2014) to find similar words and n-grams contained in the articles under analysis. For each entry of the lexicon, we extracted the top-30 similar words in terms of cosine similarity, and manually selected the relevant ones. Thanks to this step we finally have a lexicon of 514 words.

These words were manually classified by the authors into 22 classes as shown in Table B1 (Appendix B). The classes have been defined with the double aim of having a broad view of data science methods and algorithms and to cover the 10 specific meta-themes that will be discussed in greater detail in Section 4. Table 2 Appendix B shows, for each class, a sample of entries and the meta-theme to which the group of classes will answer.

2.4. Charting the data

Following this, we implemented a charting phase, where data were collected in a database using the statistical software RStudio (RStudio Team, 2020). The database is composed of 1024 scientific papers and has the following structure:

- Reference Number
- Scopus ID
- Authors
- Year of Publication
- Name of the Journal or of the Conference
- Title
- Paper type (Journal or conference paper)
- Number of Citations
- Lexicon Type
- Entry
- ED Lexicon Class
- DS Lexicon Class

The database is shared in a spreadsheet in the data set associated with this paper (see section Dataset).

2.5. Summarizing and reporting the results

In this final phase of the review, attention was given to the reading of a sub-set of the papers and to a numerical analysis of the extent, nature and distribution of the whole corpus. We focused both on single dimensions of the analysis (i.e. ED phases and DS methods) and on the intersections between the dimensions in order to identify research themes and open challenges in the bridge between ED and DS.

Concerning ED, we described in greater detail the phases identified in Section 2.3.1 through an in-depth analysis of the works belonging to each meta-theme connected to each of these phases. For the DS tools, we listed 10 meta-themes, using the classification developed in the DS lexicon (see Section 2.3.2). In order to do that, for each phase of ED and for each class of DS methods the reading list was made of the top-5 articles in terms of citations and the 5 most recent articles in terms of year of publication (around 200 papers identified). In addition, key papers that were not in the set (papers that are becoming central for a particular theme but not yet in the top-5) have also been identified and added to the reading list. The authors read the papers considering their expertise and background. The papers cited in Section 3 were chosen from the reading list in order to better explain each theme. Again, for a full list of the analyzed documents the reader can refer to the database shared in the dataset associated with this paper (see section Dataset).

To obtain a more quantitative view of the intersection between the dimensions, we used both lexicons to generate co-occurrence

[1] https://developers.google.com/machine-learning/glossary.
[2] https://www.analyticsvidhya.com/glossary-of-common-statistics-and-machine-learning-terms/.
[3] https://www.kaggle.com/shivamb/data-science-glossary-on-kaggle.
[4] https://en.wikipedia.org/wiki/Outline_of_machine_learning.

matrices. A co-occurrence matrix is used in Natural Language Processing to synthetize the strength of relations between two sets of words. The matrix has on the rows the first set of words (in our case DS methods) and on the columns the second set of words (ED phases). Intersections between the rows and columns contain the number of times that elements belonging to the two classes of words appear together in the same articles (abstract, title, keyword). These matrices were then discussed by the authors, in order to identify two phenomena concerning the literature in the intersection between ED and DS:

1 - the most important research themes, shown where the matrices had a high number of papers in the intersection between ED phases and DS tools,
2 - gaps and blind spots, identifying the emptier zones of the matrices. These zones have been explored, searching for evidence on the application of DS methods to specific ED phases.

These results are the input for the definition of the main themes and challenges, discussed in Section 4.

3. Key findings

The following subsections present key findings considering the two dimensions of the present scoping review: ED phases and DS methods. The specific meta-themes are listed in the same order as in Tables A1 and B1, respectively in Appendices A and B. The tables contain the meta themes with a sample of terms from the lexicon, to give a glimpse of their contents.

3.1. The engineering design process

Table 1 shows the number and the percentage of papers under analysis, containing at least one entry in the DS lexicon for each ED phase. Since every paper is associated with one (or more) ED phases and uses one (or more) DS methods, the table shows a distribution of the actual usage of DS methods in the ED phases (both vertical and horizontal).

3.1.1. Problem definition

The problem definition phase starts from customer needs and aims to frame the design problems to guide the ED process. In this phase, designers work closely with marketing specialists to identify explicit and latent customer needs. Identifying customers and eliciting needs is typically under the purview of the marketing staff, however it has a primary importance for the design process. One strategy is to identify lead users (Von Hippel, 1989) who make extensive and intensive use of the product and are therefore more likely to inspire new product ideas. Pajo et al. (2015) propose the FLUID methodology to automatically identify lead users through data mining of their Twitter feeds and their examination with a machine learning algorithm. Also, Chiarello et al. (2020) propose to mine twitter to identify technical problems and the advantages of new products.

The problem framing and scoping in the Problem Definition phase creates a multi-dimensional “design space” that contains possible designs that satisfy customer needs. Designers then need to evaluate the different possibilities by investigating the design space to come up with a series of concepts that can be further explored. This stage of the process may be characterized by a lack of relevant knowledge and difficulties in establishing requirements and design solutions (Quintana-Amate et al., 2015), which are exacerbated as the degree of innovation increases. Furthermore, it is hard to define the design space and to measure the value of each point of this area. After the design space has been defined, it can be searched and measured thanks to tools such as Quality Function Deployment, Axiomatic Design and Functional Analysis (Sridharan and Campbell, 2005).

3.1.2. Conceptual design

Although the definition of the design space can ideally start from scratch, most firms consider (their) previous products as a starting point for new designs (Tseng and Jiao, 1997). Reliance on past knowledge is summarized in the Knowledge-Based Engineering (KBE) approach where experts and software interact to identify the most suitable information for the ED process (Quintana-Amate et al., 2015). KBE has been impacted by the development of AI solutions, since AI tools can be applied to knowledge management in general (Liebowitz, 2001). A comparison of different AI tools to derive a model for manufacturing times based on previous designs in the aerospace sector shows that between linear regression, REPtree (tree-based) and M5R (rule-based) the latter was the most appropriate, providing a sound basis for expert learning and validation to further improve the model (Quintana-Amate et al., 2015). Once relevant information is available, new product designs can be derived from similar products through the recognition and modification of patterns in functional requirements (i.e. topology, classification, and templates). According to Tseng and Jiao (1997), Machine Learning can be used both to derive functional requirements from a pool of existing products, and to cluster these functional requirements to map those that describe the same functions, thereby simplifying the design experts' work in mapping customer requirements with product functionalities and enabling a more efficient discovery of design alternatives and their tradeoffs.

In this context, DS studies also focus on reducing the time and effort needed to explore a design space. Dasari et al. (2015) focus on Response Surface Modeling and evaluate the performance of Linear Regression, Support Vector Machines, M5P and Random Forest methodologies, finding that the latter two, being tree-based models, perform better also considering explainability.

Under the premise that quality designs come from experienced designers, Bryant et al. (2005) has implemented a system for automatically extracting experience in the form of design knowledge and stored for re-use in a web-based repository. The paper presented an automated concept generation tool that utilizes the repository of existing design knowledge to generate and evaluate conceptual design variants.

Usually, the problem definition phase ends with the definition of requirements (both functional and non-functional). Requirements can also be defined as the language with which designers define the problem.

Concerning functional requirements important work has been done for automatically collecting and mapping functions using different kinds of sources, such as patents (Fantoni et al., 2013) and to provide engineering designers with easy access to relevant functional knowledge (Wang et al., 2005a,b). New classes of user requirements, such as sustainability or total value, have also emerged in the past years as considerations that designers must contend with (Isaksson et al., 2015; Levine et al., 2016). Levine et al. (2016) propose a decision-support system including elements of sustainable product development and value-driven design, incorporating data mining techniques to support the analysis of different design concepts against these criteria. Another important non-functional requirement is modularity (Danese and Filippini, 2010). As the push towards modularity has grown over the years, DS has supported several studies in the creation or improvement of product families. These are groups of related products, based on an overlapping product architecture or platform, that support mass-customization through the addition of modules that provide specific functionalities. The first step in this process is the identification of the most suitable functionality grouping for the product family platform by examining possible variants. Agard and Kusiak (2004) base their approach on association rules between components and functional requirements, and through a neural networks approach, propose to identify candidate components, products, and processes, for standardization. Kusiak and Huang (1996) propose using a fuzzy neural network approach to evaluate the tradeoff between cost and performance in different modular designs. Moon et al. (2006) mine association rules and use fuzzy clustering to define the optimal organization of components into modules. A different approach addresses the level of generic bills of materials (GBOMs), used to represent variants within a product family. DS methodologies such as text mining, tree mining, tree union and constrained XML enable component family formation, similarity quantification and establishing design constraints to more effectively create product variants from a large collection of BOMs (Romanowski and Nagi, 2004).

3.1.3. Embodiment design

In this phase, considerations around the final product architecture are made. After the design has been defined and measured, designers start to search for the optimal design. Concepts can be thus embodied in products, following both functional and non-functional specifications.

This optimization phase can be considered as a constrained optimization problem. From our review it emerges that a body of research focuses on the use of genetic algorithms for finding the optimal solution (Mirjalili et al., 2017; Yu et al., 2007). Interesting results have been achieved by redesigning genetic algorithms for specific ED problems, in particular finding smarter ways for constraint handling (Deb, 2000). It is thus evident that genetic algorithms increase the performances in the context of ED when there is a customization of the algorithm itself in order to solve specific ED problems (Taboada et al., 2008). Among the other approaches used for optimal solution retrieval, recently Chaos Game Optimisation (Talatahari and Azizi, 2020), self organising maps with fuzzy clustering (Qiu et al., 2016) and the Monte Carlo method (Wang et al., 2005a,b) have been used.

3.1.4. Detailed design

In the final steps of the design process, the designer must detail the outputs, taking care to arrive at a precise final product specification (Giachetti et al., 1997). Many tools exist to support this step, mainly drawings, specifications, and models (i.e. prototypes, Gero, 1990) that can be physical or digital.

Concerning physical drawings and models, Computer-Aided Design (CAD) is one of the most used tools. In CAD, software applications connected to hardware systems are used to define new designs and to analyze and improve existing designs in a semi-automatic fashion (Groover and Zimmers, 1983). These software applications may constitute an interesting data source for DS applications in ED, for example as a performance measurement tool (Gao et al., 2008). However, to the author's knowledge, although CAD feature extraction for data analysis on CAD systems has been amply addressed, the problem that emerges from our review is defining what feature is. Since CAD is a native digital tool, it is creating a large amount of data and features both related to the designed product and of the design process. Indeed, from our review a gap emerges in the identification of latent features (e.g. temporal features, behavioral features) hidden in CAD data or even how to capture the design intent behind CAD applications (Rahman et al., 2019).

DS, with its inclination to support optimization, is highly relevant when evaluating the performance of existing designs (Chiesa et al., 2007). Based on their performance characteristics, new designs' performance can be predicted through data mining and machine learning techniques (Bertoni et al., 2017). Another application is the reduction of existing product complexity, that can be achieved using a greedy modularity algorithm (Kusiak and Smith, 2007).

Alongside the technical performance, also cost models contribute to design evaluation. These are constructed starting from the estimation of manufacturing activity times given certain product features or functions and multiplying them by the relevant cost rates. DS techniques such as virtual manufacturing can be used in this context to generate simulated datasets that allow the best fitting of cost models for selected designs (Stockton et al., 2013a). From these, Cost Estimation Relationships (cost drivers) were mapped using product features, process features and process activity to estimate production costs of a design, comparing different DS models such as Stepwise Linear Regression, Symbolic Knowledge Acquisition Technology, and Artificial Neural Networks (Stockton et al., 2013b).

3.1.5. Project management

Since ED is a multidisciplinary endeavor, language and communication issues arise when discussing design inputs and decisions, that can be studied and overcome through various DS methodologies with the purpose of improving Project Management. In particular Natural Language Processing has been hypothesized as useful to act upon different aspects of the communication process, such as written and spoken language used to describe design artefacts and decisions or used to exchange information between design team members; image analysis can be used for the analysis of graphical and pictorial representations of the design objects; deep learning techniques, thanks to their ability to take as input low level features are interesting for the analysis of rules and relationships governing the combination of individual design elements or functions, the relationships between features and functions, the connection between physical principles and product functionality, and all the relevant numerical data that can be derived from the above (Dym et al., 2005).

In addition to communication between designers, Project Management has also intensive phases of search for useful information by designers as they navigate the ED process. This step has also been tackled by DS. In this case, several studies have focused on the creation of taxonomies to support designers in the search, indexing and retrieval of relevant knowledge. While the taxonomies themselves rely heavily on domain knowledge, DS methodologies such as Text Mining can support their final validation against a set of real design documents (Ahmed et al., 2007).

3.1.6. Innovation and IP management

The discipline of ED can be in itself a vehicle of innovation, when changes in existing design provide new meanings for the final customer that change their relationship with the product (Verganti, 2011). In addition, ED can support innovation efforts by accompanying technological breakthroughs. In recent decades, the pace of product innovation and the centrality of Intellectual Property (IP) management have gained increasing importance and firms are increasingly adopting automated data management processes to feed their innovation pipeline (Corso et al., 2001). There are two main objectives that DS methodologies are geared towards in the domain of knowledge management.

First, DS can be used to extract knowledge from written sources such as patents (Moehrle and Caferoglu, 2019), papers and requirement documents, which can be organized through DS, for example creating ontologies that are representative of a field of interest (Štorga et al., 2010; Trappey et al., 2009). Second, DS may support knowledge expansion, supporting creativity during the design process. In this vein, the literature stream on the Theory of Inventive Problem-Solving (TRIZ) is concerned with defining and solving problems, by supporting creativity through a structured process (Mann, 2001; Cascini, 2012). This has been successfully applied across ED efforts, especially for product innovation, including the identification of new product features and the evolution of product architectures (for a review see: Cascini, 2012). In this context, the examination of patents and textual data can support the evolution of product functions and the identification of contradictions in product design (Cascini and Russo, 2007; Cascini et al., 2009, 2011).

Recent literature has also demonstrated that Natural Language Processing techniques can be used to mine information from IP related documents, especially from patents (Chiarello et al., 2017, 2018a; Chiarello et al., 2019) and technical specifications (Fantoni et al., 2020). Recent Named Entity Recognition systems have been applied to patent documents in order to extract design relevant information (Lee and Hsiang, 2019; Chen et al., 2020; Kim and Yoon, 2021).

3.1.7. Design for X

Design for assembly and design for manufacturing led to important advancements in simplification, costs, quality, and time to market of products. During the early 2000's, environmental aspects have started to be considered in design, through efforts to reduce products' total life-cycle costs and environmental impact. Therefore, researchers began to focus their attention on design for environment, design for recyclability, design for life-cycle (DFLC), etc. These studies are sometimes referred to as Design for X (DFX) (Kuo et al., 2001).

From our scoping review, a lack of focus on the usage of DS tools for design for X has emerged. Even if this gap can be due to biases in our approach (as Design for X is not supported by a clear lexicon), also following a manual review of the identified articles we confirmed that many unexploited opportunities still exist in the usage of data in order to improve the development of products with the aim of maximizing non-functional requirements.

From the review, a specific focus on design for manufacturing in a concurrent design setting sees the implementation of fuzzy clustering to data, linking functions to design features and functions to manufacturing steps. These clusters are then used to optimize the production process for selected designs (Xue and Dong, 1997). Similarly, part-family formation for group-technology applications also benefited from fuzzy clustering to create part families that can be grouped by the required machining tasks (Liao, 2001).

3.1.8. Design education

There is ample literature dedicated to design education, focusing on methods and tools to be employed in design courses, as well as student performance in design courses. The research focuses on investigating effective ways to teach and assess the design process, and on the review of curricular structures and pedagogies that are commonly used in undergraduate engineering programs.

From the literature, a focus emerges on the use of data-driven tools in order to assess the design education phase (Davis et al., 2002). The tools show the possibility to efficiently measure student learning of engineering practices and student performance in a way that readily affords evaluation.

Concerning the most-used approaches for teaching design education using data tools, project-based learning (Lantada et al., 2013) and serious game (Wang et al., 2016) emerge. Both approaches show the need for more practical approaches for teaching engineering design in order to foster practice and not only theory. DS can be an interesting field of application since its products do not need physical environments for students to experiment different designs.

It is evident from our analysis, that there is a growing use not only of new methods for teaching but also of new hardware to enable ED education. Recent literature is growing in the use of robots and IoT solutions (De Luca et al., 2018). This is opening up many opportunities to measure teaching outcomes, since ED teaching is becoming a natively digital process.

In the next year, we expect also a growing interest in clearly identify and mapping the skills that designers will need to collaborate with Artificial Intelligence systems (Fareri et al., 2020).

3.1.9. Prototyping

Prototyping is of high importance in the ED space as it is the main tool through which designers interact and communicate with the final customer. Prototypes can be both physical or digital and represent the key knowledge that links functions, structure and behavior of the product (Gero, 1990).

Additive manufacturing in general, is having a strong impact on the way products are prototyped (Burckhard and Wampol, 2018). This is increasing the quantity of data that is produced during the prototyping phase, opening to new opportunities for prototyping performances assessment (Nguyen et al., 2020; Barclift et al., 2017). Interesting developments are emerging also on the use of virtual reality (Kuester et al., 2001; Gustavsen and Louka, 2012; Gulrez and Tognetti, 2014) and augmented reality (Liverani et al., 2004; Januszka and Moczulski, 2011; Ren et al., 2017; Gattullo et al., 2019).

3.1.10. Sustainability

The design of more sustainable products can be achieved by solving conflicts that often exist between, environmental impact, cost, and product performance. To support such a process, scholars are starting to use data driven approaches (Eddy et al., 2014), even if, among the design phases considered, sustainability seems to be the one least impacted by data tools. However, a growing interest in this context emerges from the literature review, especially on the topics of Life Cycle Assessment (Ostad-Ahmad-Ghorabi et al., 2011; Balochian and Baloochian, 2019) and Recycling (Dering and Tucker, 2015; Cuprak et al., 2008).

3.2. Data science tools

3.2.1. Data science problems

When using DS tools in the context of ED, the most discussed problem is Big Data (e.g. Hassannezhad and Clarkson, 2017; Soni and Mathai, 2015; Yin et al., 2018). Big Data is a well-studied problem in the context of DS and addresses the data systems that are too large or complex to be dealt with by traditional data-processing application software. In the context of ED, the focus is stronger on the second part of the problem of big data (complexity), since ED-related data contains a mix of structured and unstructured data coming from many different sources. The dimension of the data can be a problem in the context of ED especially when the data is coming from customers (Lin et al., 2017). In these cases, the company may need to deal with large amounts of data, not manageable with standard databases.

One interesting emerging problem in the context of ED, to which DS tools are proposing solutions, is Anomaly Detection (Tawhid and Savsani, 2018), the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data. The tools for anomaly detection can be used in the context of ED for both the identification of product anomalies (and thus failure identification) and for customer analysis, in order to identify weak signals coming from the market.

3.2.2. Data science applications

The most discussed application among the analyzed papers is Data Mining (e.g. Miguel and Miguel, 2012; Regli et al., 2010; Xie et al., 2018a,b). Data Mining is an application that is having an impact on all the phases of ED. It is in fact interesting for designers to mine knowledge from historical data, both for creative and structured tasks (e.g. concept generation, conceptual design, design space exploration) and for synthetic and structured tasks (e.g. optimization design, failure identification and analysis, CAD). It is interesting to notice that data mining is seeing a growing interest also in the context of engineering education.

The second most popular application concerns Expert Systems (e.g. Chakrabarti, 2001; Cheng et al., 2009; Rafiq et al., 2001; Tiwari et al., 2009). These are used both for knowledge extraction and as heuristic approaches to design optimization.

Another important application is represented by Time Series Analysis (e.g. Kasdirin et al., 2015; Regli et al., 2010; Zhang and Du, 2014). With these tools, engineers are exploring the dynamic perspective of the design process. In particular, these tools are used to assess performance over time both of ED teams and of the outputs of the ED processes (i.e. products, systems).

Among the weak signals emerging from our analysis, an interesting application is related to business analytics and intelligence (Loginov, 2010; Sun and Xiong, 2014; Zhang et al., 2012a,b). Since the ED process is a sociotechnical process, it is evident that typical business analytics approaches are emerging for the analysis of complex dynamics and the resulting trends that occur when social and technical systems intersect.

3.2.3. Data science methods

Among the 1161 papers analyzed, 569 (49,6 %) explicitly cite a DS method in the title, abstract or keywords. Table 2 shows a summary of the classes of methods and their distribution in the context of ED.

Among the evolutionary algorithms, the most used in the context of ED are genetic algorithms (Borgue et al., 2019; Brown, 2005; Deb et al., 2000; Peysakhov and Regli, 2003; Whitfield et al., 2003). Their main application is in the context of design optimization, but also as support design space exploration.

The second class of methods is artificial neural networks. Inside this class, a great importance is held by deep learning (Kale and Kulkarni, 2018; Neeley et al., 2013; Quintana-Amate et al., 2015).

Concerning regression analysis, both linear regression (Chen et al., 2006a; Matelli et al., 2009; Sadollah et al., 2015) and polynomial regression (Chen et al., 2006b; Miller, 2004; Talbi, 2009) are largely applied in the context of ED.

3.2.4. Data preprocessing approaches

Data preprocessing is one of the most important tasks in the DS process. In the context of ED, it has a particular importance since data is not usually designed for analytical purposes (e.g. CAD data, unstructured customer surveys, videos). In this sense, it is important to properly preprocess ED related data to avoid the well-known ğarbage in, garbage outëffect.

In our analysis, some articles explicitly focus on the structure of the data or on transformations to be performed on the data in order for it to be properly analyzed (e.g. Huang and Jin, 2010; Liu et al., 2010; Ong and Keane, 2002). Given the growing interest that has emerged in the last years around artificial intelligence and machine learning tools, the risk in this contest is that scholars do not properly focus on this important step of the analysis, assigning a higher importance to the modelling part. This can be a problem especially in the ED context, since it can lower the reproducibility of the presented works.

3.2.5. Dimensionality reduction approaches

Dimensionality reduction methods are used to transform data from a high-dimensional space into a lower dimensional one. The lower dimensional space is designed in such a way that when data is reprojected, it retaina some meaningful properties of the original space. It is not uncommon in the context of ED to work in high-dimensional spaces and it is usually undesirable for many reasons (i.e. sparsity, computational cost). Our analysis shows that dimensionality reduction in the context of ED is usually adopted to treat data with large numbers of variables. It is an evidence that the problem of having a high number of observations is rarer in this context. This is in line with the fact that data in the context of ED is usually more complex data than big-data.

The most used approach for dimensionality reduction is Principal Component Analysis (PCA) (Huque and Jahingir, 2002; Zheng et al., 2020) but also Latent Dirichlet Allocation (LDA) is used when researchers have to deal with textual data (Hamza et al., 2018; O'Halloran et al., 2018).

3.2.6. Data visualization

Data visualization is the process of graphically representing data, largely used to exploit the great skills that humans have in synthesizing large amounts of information when dealing with images. The choice of proper visualization tools has an important impact on the phases of data exploration and for synthesizing the output of the analytical process.

Surprisingly, very few of the analyzed papers explicitly talk about data visualization tools and methods (e.g. Becattini et al., 2013; Chiu and Kremer, 2011; Zhang et al., 2012a,b), leading to the conclusion that these tools are scarcely used in the context of ED. This evidence points to interesting future developments, considering also that many tools already used in the ED context involve the usage of images for handling complex data (e.g Computer Aided Design). In addition, it is interesting to note how the most adopted tools for visualizing (and exploring) data in this context are dashboards, indicating that data visualization for ED is nowadays more used in the phases of results synthesis with respect to the phases of data exploration, where other kind of tools (e.g. plots, bar charts, histograms) are used (e.g. Harwood and Revell, 2018; Li and Tate, 2013; Wang et al., 2014).

3.2.7. Machine learning hardware

Machine-Learning-specific hardware is a class of specialized hardware designed to make the process of learning more efficient. This class of hardware accelerates machine learning applications, especially when the model is based on deep artificial neural networks. From our analysis, it emerges that mostly graphics processing units (GPUs) (e.g. Kumar et al., 2020; Xie et al., 2018a,b; Wang et al., 2016) are used for applications in the context of ED.

3.2.8. Programming language

While there is a strong debate on which programming language best suits DS problems (Grover and Kar, 2017), it is becoming clear over the years that every programming language has its advantages and disadvantages. One of the most important advantages is the size of the community of practitioners (including students and academia) that works on or teaches about a specific class of problems using that programming language. This in fact, also thanks to open-source software, contributes to the creation and sharing of ready-to-use software packages that speed up the process of software development.

In ED, the most popular software is Matlab (e.g. Dudas et al., 2014; Hu et al., 2014; Zhou et al., 2017), followed by Python (e.g. Guzey et al., 2017; Kopena et al., 2005; Xie et al., 2014), while just a few applications use R and RStudio (Chong et al., 2018; Hu et al., 2017).

3.2.9. Unstructured data processing

Data is unstructured when it is not arranged with a pre-defined model or schematic. This kind of information is typically text- and image-heavy and is thus irregular and ambiguous. For this reason, to understand its latent content, specific pre-processing approaches are necessary that can structure the data. Considering that unstructured data comprises the vast majority (about 80 %) of data found in organizations (Shilakes and Tylman, 1998), it is clear how treating such data is important for the ED process.

In our set of papers, many works use Natural Language Processing (e.g. Field, 2007; Saha et al., 2011; Ye et al., 2009) and Text Mining techniques (e.g. Kudo et al., 2011; Ji and AbouRizk, 2018; Ye et al., 2009). More specifically, techniques such as language recognition (Zhou et al., 2015), machine translation (Zhang and Du, 2014), semantic retrieval (Guo et al., 2013), semantic web (Lin and Harding, 2007), question answering (Ferguson and Ohland, 2012; Savoie and Frey, 2012), speech recognition (Jiang et al., 2015), speech synthesis (McGoldrick et al., 2016) and topic modelling (e.g. Hamza et al., 2018; O'Halloran et al., 2018; Zhang and Althoefer, 2019) have already been used in different contexts of ED.

Concerning video, applications of computer vision (e.g. Costa and Sobek, 2004; Kamboj et al., 2020; Miao et al., 2017), object detection (e.g. Desrochers and Cherkaoui, 2002; Jiang et al., 2015; Lortal et al., 2006) and video processing (Radj and Senthilvelan, 2006) have been detected in our paper set.

3.2.10. Supervised vs unsupervised learning

This last theme is linked with others already discussed (see Section 3.2.3); here we want to understand if there are any differences in terms of numbers of papers that use supervised or unsupervised learning techniques in the context of ED. This difference is interesting since while supervised learning techniques map an input to an output based on example input-output pairs, unsupervised learning looks for previously undetected patterns in a dataset with no pre-existing labels. The main differences between the two approaches are thus the cost of human supervision and the need for labelled data.

We report that 80 % of papers use supervised approaches and 20 % use unsupervised approaches. This is in line with evidence that researchers tend to use DS tools when they have labelled data available. In the context of ED, given the difficulties to obtain data from companies, this can be a barrier to the adoption of DS tools, also unsupervised ones. At the state of the art, the most commonly adopted approaches for unsupervised learning are k-means clustering (e.g. Bailey et al., 2006; Singh, 2016) and self-organizing maps (Carraro et al., 2017).

4. Research themes and challenges

From the summary of our literature review presented in Section 3 and considering some seminal works on this topic (e.g. Reich et al., 1993), it is evident that DS and ED are already intertwined. This is due to the fact that DS is not a new field of research (although the concept has emerged in the last decade) but a convergence of many other existing fields (i.e. artificial intelligence, machine learning, statistics) some of which already overlapped with ED research. Despite this, DS is growing fast and ED researchers risk being left behind; furthermore, specific foci are emerging that may create a fixation effect, leaving many potential applications of DS tools in ED unexplored.

In this section we synthesize the results of Section 3 using graphs and charts to list the major research challenges emerging from our scoping review. We suggest ways in which our results may inform future research on the use of DS tools in the context of ED. Similarly to the structure of Section 3, this section is divided in ED and DS related challenges. Each challenge is connected to a specific meta-theme that has guided our scoping review. To conclude this section, we also suggest a list of possible additional challenges that did not emerge during our scoping review. Indeed, literature has highlighted a gap between design research and industrial practice (Chakrabarti and Lindemann, 2016). Considering the link between ED and DS only through the lens of research therefore introduces a bias that we have tried to mitigate in Section 4.3, where we list some additional challenges on how data science helps in dealing with iteration in ED practices across phases and sectors.

4.1. Engineering design challenges

To provide a synthetic view of the research themes, Fig. 1 shows a graph of the entries of the ED and DS lexicons. The graph is computed considering each entry as a node. An edge exists if the two entries appear together in at least 5 documents; the most these two entries will appear together, the closer they will be on the graph. The dimension of the nodes is proportional to the degree of the node (i.e. the sum of the weight of the edges hitting the edge). Finally, the colors of the node depend on the class they belong to.

We highlight ten main challenges from a DS viewpoint:

Challenge 1. Develop data-driven tools for problem definition: data can help in the exploratory phase of design, helping to find the right questions to answer and the right problems to solve. Scholars can use novel artificial intelligence approaches (especially NLP tools) to explore and synthesize large amounts of unstructured data, in order to support the phase of problem definition.

Challenge 2. Integrating top-down and bottom-up approaches in AI systems: Conceptual design is a transitionary phase. The customers' needs have to be translated in requirements, exploring the problem and the solution spaces. In this phase AI can help, but the design knowledge has to be injected into the system. This is why we need to develop explicit systems using formal declarative information models that follows an imperative procedure (top-down) in standard AI systems that do not need explicit data as input in their implementations.

Challenge 3. Move towards genetic algorithms for design optimization: as it emerges from Table 2, a strong focus is placed by literature on the use of genetic algorithms for optimization in design. ED researchers limit their access to state-of-the-art optimization systems (e.g. Neural-Networks-based solutions) coming from the DS field. This may also create a fixation effect. Scholars should explore other solutions for optimal design to advance the state of the art in ED and to incorporate more effective solutions coming from other fields of research.

Challenge 4. Design novel ways of using CAD as a data source: CAD is widely used by scholars and practitioners in ED, and Feature extraction on CAD has been addressed for a long time. The problem that emerges from our review is defining what a feature is. Since it is a native digital tool, CAD is creating a large amount of data and features both related to the designed products and of the design process. Scholars should study how to use DS tools in order to identify latent features (e.g. temporal features, behavioral features) hidden in CAD data (Rahman et al., 2019) or even how to capture the design intent behind them.

Challenge 5. Improve communication in design teams thanks to NLP: recently the state of the art of NLP tools has been advanced by contextual embedding systems. These systems are achieving great results in natural language understanding and generation (Brown et al., 2020). The vast amount of text generated for communication purposes during the design process can be analyzed by these systems to improve communication. The amount of digital data generated for communications is also being fostered by the advent of COVID-19, and the necessity to work remotely. Furthermore, generation systems could be used to automate many communication tasks.

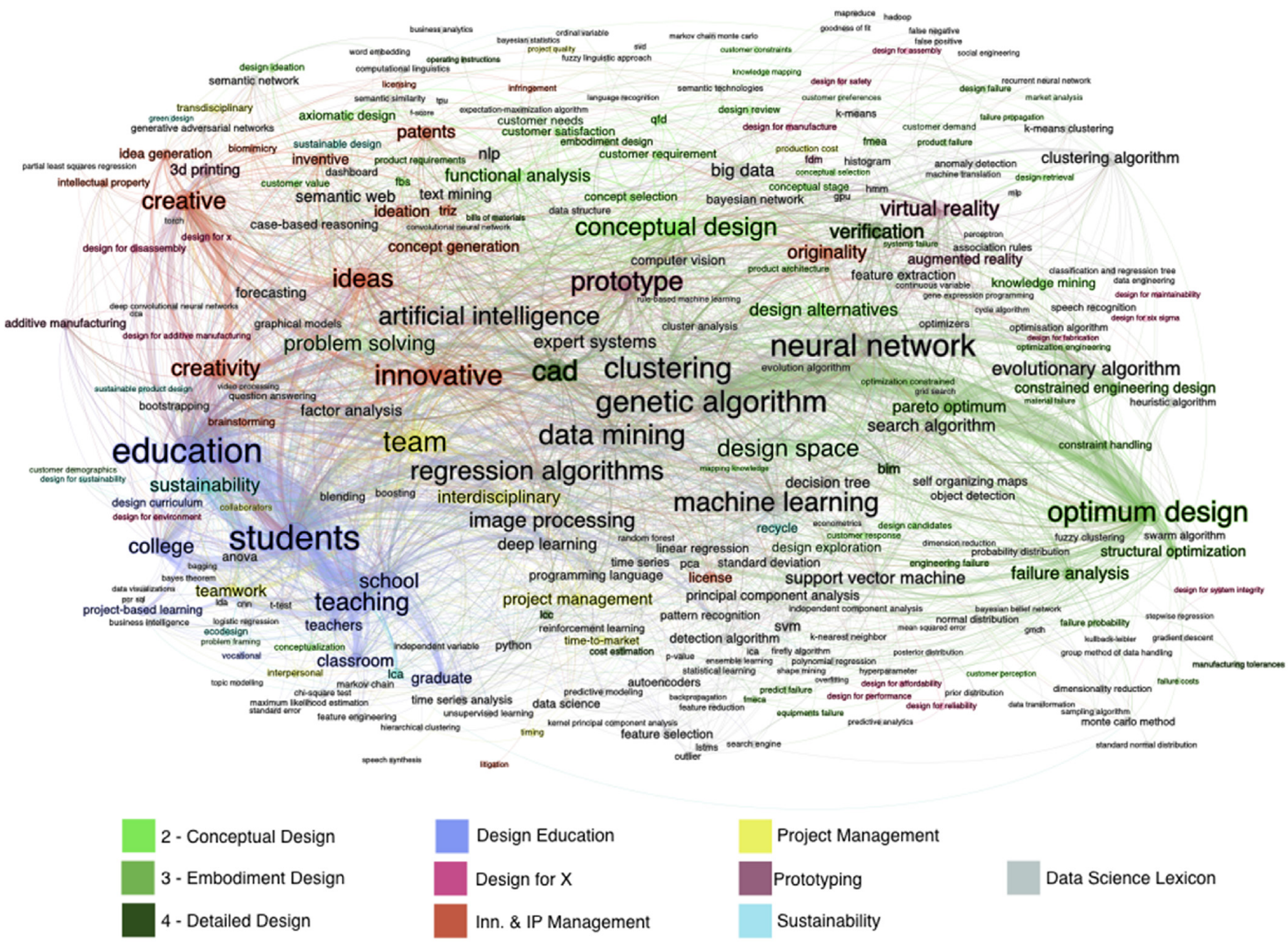

Fig. 1. Graph of ED and DS Lexicons and their Relationships in the Reviewed Articles.

Challenge 6. Use of AI to assist the inventive process: The state of the art on AI creativity is broad (Boden, 1998; Ritchie, 2007; Singh and Gu, 2012; Georgiev and Georgiev, 2018). Even if we can consider creativity as a new fronter in AI (Colton and Wiggins, 2012), it is reasonable to assume that there will be an increase in the usage of AI-aided systems for creativity support in ED. In particular, AI is interesting for ED for exploring the potential of conceptual spaces (Colton and Wiggins, 2012). We are aware that no breakthrough in the usage of AI in ED will be reached without a theory of discovery or invention. However, the usage of AI for exploring creativity tasks can help in better framing a theory of discovery or invention. ED Scholars may study systems to help designers in the generation phase and study the social, legal and ethical issues linked to this.

Challenge 7. Implement DS method in Design for X: from our review it is clear that the design for X process is lacking in the application of data science tools. Many opportunities exist for designers to improve this process thanks to the use of data, further developing standard knowledge-based inference methods and Ẅhat-If' analyses (Abramovici et al., 2009).

Challenge 8. Move beyond student-generated data: the analyzed literature shows an intense use of student-generated data in the context of ED studies related to education. This can create many limitations in terms of the phenomena that can be studied and in terms of reproducibility of the experiments. Scholars should strive to identify other data sources to study the teaching and learning of ED principles and tools. We acknowledge that this process requires trust-building and a deep connection to industrial settings that are traditionally hesitant to share design data.

Challenge 9. Speeding up the prototyping process: Data driven technologies are reshaping the way in which prototypes are developed, helping engineers in assessing the technical and financial feasibility of designs. The research on the bridge between ED and DS can focus on taking fast and objective decisions using *pretotypes*[5] (Savoia, 2011) and also rethinking the prototyping process, thanks to the opportunities offered by new technologies such as augmented reality (Gattullo et al., 2017).

Challenge 10. Exploit Smart Manufacturing Data to Design Sustainable Products: The analyzed literature shows a gap in the application of data driven tools for the assessment of the sustainability of designs. Therefore, the problem to be solved in this challenge is not only to develop data driven tools for measuring sustainability in design but obtaining meaningful sustainability measures that can drive the design process. In this context, Smart Manufacturing (i.e., a form of production integrating manufac-

[5] Pretotyping is a stage of design that should take place before building a prototype. Savoia (2011: 4) summarizes the key difference as: *"most prototypes are built to answer questions such as, "Can we build it?" or "Will it work as expected?" instead of focusing on questions such as "Should we build it at all?" or "If we build it, will people buy it and use it?" Unless you can answer the latter questions positively, the former questions are of little importance".*

turing assets with sensors, computing platforms, simulation, data modelling and predictive engineering) represent a huge opportunity for ED (Kusiak, 2018). Researchers can study how to use data coming from sensor-enabled processes in order to design more sustainable products. In addition, obtaining further data from the entire product lifecycle could also feed the design process with relevant measures able to orient the design actions.

4.2. Data science challenges

To switch to the DS related challenges, Figs. 2 and 3 break down respectively the DS methods (see Section 3.2.3) and the unstructured data processing techniques (see Section 3.2.9) used in each ED phase.

Drawn from these figures, the main challenges regarding DS are as follows.

Challenge 11. Mapping problems of ED in DS: In this review we highlighted some initial attempts. We expect additional works that support problem-mapping, to use novel data sources such as patents, papers and social media.

Challenge 12. Mixing Expert Systems and Machine Learning Approaches: from the review it emerges that DS tools used in the context of ED are polarized towards the use of expert systems (i.e. rule-based systems) and machine learning. ED can take great advantage of hybrid systems, exploiting both the knowledge of experts and the information coming from data sources. This process can be fostered by the new studies on the explainability of black-boxes models (Guidotti et al., 2018).

Challenge 13. Redesign ad-hoc DS methods: many DS systems have the limitation of being context-dependent. This is especially true for systems dealing with unstructured data (i.e. natural language, images, videos) where data are structured *ad hoc* before being analyzed. For this reason, many DS methods must be redesigned to work effectively in the ED domain. Scholars in the field of ED can join DS experts in order to redesign state of the art data-driven systems in order to properly perform in the specific context of ED and its subfields. Some examples are the fine tuning of popular contextual word embeddings (Devlin et al., 2018) for domain specific application (Beltagy et al., 2019; Lee and Hsiang, 2019).

Challenge 14. Structuring design related data: since the pre-processing phase is one of the most time-consuming ones of the analytical process, scholars may study common data structure for data related to ED. Future work may focus on the development of data structures and ontologies to define fuzzy concepts that are crucial in ED phases such as users, design objectives, requirements or product failures. This challenge is strongly linked to challenge 13 and must be faced by ED researchers working with Data Scientists.

Challenge 15. Search for meaningful feature representation: feature engineering is an important step for data analysis. A meaningful feature representation enables important steps forward for the performance of machine learning systems. This process is context dependent. Scholars can find meaningful feature representations for different ED tasks, in order to foster the accuracy of machine learning systems.

Challenge 16. Define more effective visualization tools for ED: data visualization is used both for communicating within the research team (i.e. data exploration) and for communicating the results of the analytical process. Identifying which are the most effective ways of synthesizing data in the context of ED can bring important improvements in the field and enable a better communication of data-driven analysis.

Challenge 17. Scale experiments using advanced hardware: ED related data is usually more complex (many features, unstructured input data) than big (number of observations), although some applications need big data. In this context, the use of the appropriate hardware can make it possible to scale up the experiments. Research may test and describe experimental set-up using state-of-the-art hardware systems (i.e. GPU, TPU) in order to make it easier for other scholars to re-use these experimental set-ups.

Challenge 18. Develop ad-hoc packages for ED: packages are the fundamental units of reproducible code and include functions, documentation, and sample data. To foster the use of data in the context of ED, scholars and practitioners may develop packages especially designed for the ED context, as well as examples and datasets associated to particular methods. This may lead to the development of contextually indexed generic methods.

Challenge 19. Combine textual and visual data: scholars in ED are using text and images separately. ED processes produce both types of data and almost any ED documents contain drawings, pictures, sketches, digital images, and text. This opens to the use of Artificial Intelligence systems able to combine text and images since from a recent stream of research (Qiao et al., 2019) it is evident that there exists great value in using them together in Artificial Intelligence systems. The combination can have an impact on two ED Related activities. The first one is image generation, that has recently reached remarkable results[6] thanks to the implementation of the language systems GPT-3 (Brown et al., 2020). This task can be of great interest for supporting the creative phases. The second is increasing the accuracy of AI systems in ED, since using the text and images together can open to the uncover of unexplored latent knowledge produced during the ED process (e.g. mixing CAD images and project reports over time).

Challenge 20. Automatic data labelling: one of the main problems to be solved in the context of DS is the need for machine learning systems to work properly, in particular receiving as input labelled data from which to learn. It is hard to find open labelled data in the context of ED and this is limiting the development of machine learning systems. There are two possible approaches to address this issue. The first one is the development of curated labelled data that may be extracted from multiple sources. This is an optimal solution but is expensive and labor intensive. Another possible solution is to test and describe approaches to automatically label data, using small samples of already labelled databases or using expert systems in order to accomplish this task. For both these solutions, a greater interaction between ED researchers and practitioners has to be reached.

4.3. Interface challenges

There are aspects and challenges of ED that can be mitigated by DS that have been deliberately left out from our scoping review. These challenges are related to the fact that DS has blurry boundaries between its phases (see Section 3.1), with other disciplines and with other sectors, while the approach of ED researchers when dealing with DS techniques, is a vertical one (Reich et al., 1993). However, during our research some of these challenges related to the interfaces of ED have emerged; in this section we briefly classify and discuss these challenges with the goal of guiding future research on the topic.

Blurry boundaries of DS create gaps that can be summarized as "interface challenges", that can be classified as follows:

Challenge 21. Interfaces between ED research and industrial practice: A specific stream of studies investigates the relationship with and the impact of ED research on Industrial Practice (Chakrabarti and Lindemann, 2016). DS tools have already been used in ED by many companies, but scant evidence exists of a suc-

[6] https://openai.com/blog/dall-e/#fn1.

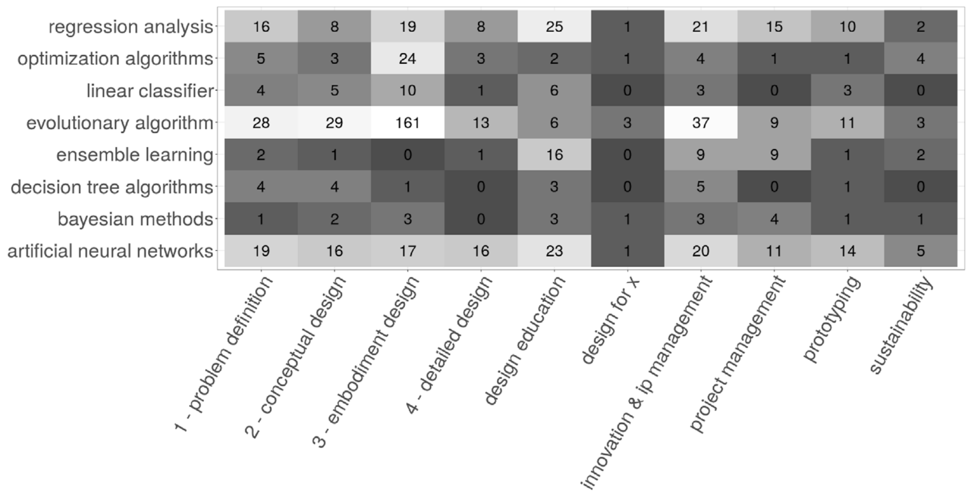

Fig. 2. Number of papers in the dataset mentioning DS Methods per ED Phase.

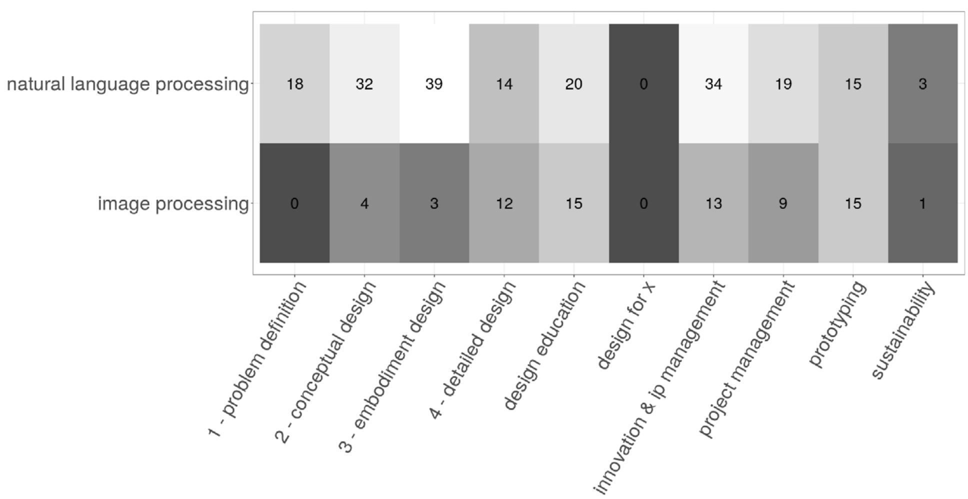

Fig. 3. Number of papers in the dataset mentioning Data Processing Techniques per ED Phase.

cessful transfer of the outcomes of data-driven design research into practice, as well as methods and guidelines for the successful transfer of research into practice. The risk is that in the coming years, the same problems highlighted in more general ED research will be replicated considering the interaction between ED and DS. To mitigate this risk, better communicating design research results, using tools offered by DS such as interactive dashboards (Chang et al., 2020), reproducible reports (Xie et al., 2020), open-source repositories (e.g., Github) and the democratization of the usage of these tools, can ensure that design researchers can widely disseminate their results in a convincing and useable format (Hales, 2005). Also, rigorous approaches to carry out industrial case studies can improve the knowledge of the phenomena we are studying.

Challenge 22. Interfaces between ED research and other disciplines: ED embraces competences of several scientific sectors and several disciplines of study. Solving design problems properly and exploring the problem-solutions dynamical spaces, requires the integration of skills from different domains. This integration is often indispensable to face certain problems in a complete and effective manner, especially concerning modern digitalized products (Kügler et al., 2018), the industry 4.0 revolution (Chiarello et al., 2018b) and the rise of Product Service Systems (Rapaccini et al., 2013). This creates a skill gap, to which many scholars have proposed educational solutions (Yan and Sawada, 2006; Telenko et al., 2016). ED practice still lacks approaches and guidelines for managing this multidisciplinary approach in a scientific way and to overcome the fragmentation of knowledge.

In this context, DS may present a bridge between ED and other sectors thanks to availability of new data, new methods and new skills to manage and exploit different knowledge sources. Some examples are research on cognitive psychology (Howard et al., 2008), marketing and operations management (Krishnan and Ulrich, 2001), sociology and anthropology (Sonalkar et al., 2011). ED scholars can work with data scientist on cross-domain datasets in order to find hard evidence that can guide design practice in the near future.

Challenge 23. Interfaces between non-contiguous ED phases: DS works within and across phases of ED (Hales, 1987) and this aspect is usually not visible in the literature. The reason why most academic papers do not address this issue, is that the localized application of DS methods is easily experimented with from research perspective. Linking to challenge 21, a better understanding of what happens in practice is needed in order to understand the relationships between phases of ED. Since data flows among different phases and processes, data can be the right media to study interfaces between non-contiguous DS phases. The advantages of data driven horizontal and vertical integration are already visible thanks to Industry 4.0 (Pérez-Lara et al., 2018). ED researchers can study these practices in order to design better products.

5. Conclusions

DS and ED are two fields of study that are becoming increasingly intertwined. On the one hand, firms require rapid innovation in products that are becoming more complex, therefore pressuring designers into streamlining the design process through the reuse of past knowledge and through a more context-aware approach that considers customer needs and product lifecycles. On the other hand, design tools are producing increasing amounts of data that can be exploited to support all the different phases of ED, including its most creative aspects that are traditionally not supported by classical optimization-based methodologies. This paper reviews the intersection between the DS and ED literature streams, by searching for papers whose terminology matches both ED and DS lexicons. Our review has been heavily based on Natural Language Processing systems and has identified a series of challenges, related both to the integration of DS methodologies into the DS process, to fully capitalize on their benefits, and to the advancement of DS methodologies to better serve the ED process.

This paper therefore represents a call to ED and DS but also to researchers and practitioners to work towards new applications of DS with new data sources and to solve challenges in methodologies, to support designers in better serving their customers in an efficient and sustainable manner.

Author statement

Chiarello Filippo: Conceptualization, Resources, Investigation, Methodology, Software, Visualization, Writing - Original Draft.

Belingheri Paola: Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing.

Fantoni Gualtiero: Conceptualization, Writing - Original Draft, Supervision, Writing - Review & Editing.

Dataset

Chiarello, Filippo; Belingheri, Paola; Fantoni, Gualtiero (2020), “Data Science for Engineering Design: State of the Art and Future Directions”, Mendeley Data, V1, doi: 10.17632/w59xtmrpk4.1.

Appendix A

Table A1
ED Manuals and Characterisation of ED Phases with sample of entries of the lexicon.

Engineering Design Books' Authors					
Meta Theme		(Pahl et al., 2007)	(Dieter and Schmidt, 2009)	(Ulrich, 2003).	(Haik et al., 2018
ED1	Problem Definition	Clarify the task; Setting up requirements list	Identify the customer needs; Gathering information	Opportunity identification; Identifying customer needs	Identify customer needs; Market analysis; Defining goals
ED2	Conceptual Design	Abstracting to identify the essential problems; Establish function structure; Developing working structure; Developing concepts	Conceptualization; Concept selection; Refinement of the product design specification; Design review	Product specification; Concept generation; Concept selection; Concept testing	Establishing functions; Task specification; Conceptualization; Evaluating alternatives
ED3	Embodiment Design	Determining requirements; Produce scale drawings; Develop and select preliminary layouts and form; Evaluate against technical and economic criteria; Optimise and complete form design; Check for errors and disturbing factors; Prepare preliminary list and production documents	Determining product architecture; Configuration design of parts and components; Parametric design of parts		
ED4	Detailed Design	Definitive Layout; Finalise details; Complete detailed drawings; Integrate into overall layout drawings, assembly drawings and parts lists; Complete production documents with production, assembly, transport and operating instructions; Check all documents for standards, completeness and correctness	Drawings; Verification testing of prototypes; Assembly drawings and assembly instructions; Detailed production specifications; Make or Buy; Detailed cost estimation; Design review	Prototyping	Analysis and optimization; Prototyping
ED5	Project Management	x		x	x
ED6	Innovation and IP Management			x	
ED7	Design for X	x	x	x	
ED8	Design Education				x
ED9	Prototyping				
ED10	Sustainability				

Appendix B

Table B1
DS topics and sample of entries of the lexicon.

Meta Theme	Class	Sample of entries
DS1	data science problems	big data; bias-variance dilemma; multi-class classification; underfitting; learning to rank
DS2	data science applications	expert systems; hybrid recommender systems; predictive analytics; pattern recognition; shape mining
DS3	artificial neural networks	Neural network Gaussian process; deep convolutional neural networks; neural-network; lstm; RNN
	bayesian methods	bayesian networks; k naive bayes; gaussian naive bayes; bayesian knowledge base; bayesian statistics
	decision tree algorithms	chaid; id3; id3 algorithm; tree based models; c5.0 algorithm
	ensemble learning	gradient boosting; blending; bootstrap aggregating; adaboost; ensemble learning
	evolutionary algorithms	genetic algorithms; evolutionary algorithm; Evolutionary programming; genetic programming; evolution algorithm
	linear classifier	logistic-regression; support vector machine; perceptron; fisher's linear discriminant; logistic regression
	optimisation algorithm	optimisation algorithm; evolutionary multimodal optimization; optimizers; swarm algorithm; firefly algorithm
	regression analysis	regression spline; lasso regression; linear regression; multivariate regression; regression algorithms
DS4	data preprocessing	data structure; nosql; data-augmentation; data transformation; dplyr
DS5	dimensionality reduction	feature hashing; minimum redundancy feature selection; directional-derivatives; lda; pcr
DS6	data visualisation	histogram; box plot; plotly; ggplot2; bar chart
DS7	machine learning hardware	vpu; tensor processing unit; tpu; graphics processing unit; gpu
DS8	programming language	scala; weka; python; stata; r studio
DS9	image processing	image recognition; video processing; computer vision; image processing; optical character recognition
	natural language processing	speech recognition; autoencoder; automatic summarization; textblob; nltk
DS10	supervised learning	instance-based learning; gmdh; group method of data handling; inductive logic programming; anova
	unsupervised learning	Action model learning; generative topographic map; k means clustering; structured knn; growing self-organizing map
MISC	machine learning algorithms	expectation–maximization algorithm; rprop; forward–backward algorithm; stochastic gradient descent; linde–buzo–gray algorithm
	machine learning methods	Adversarial machine learning; online machine learning; gradient-accumulation; occam learning; lvq
	statistical methods	frequentist statistics; skewness; categorical variable; standard deviation; maximum likelihood estimation

Declaration of Competing Interest

The authors report no declarations of interest.

References

Abramovici, M., Neubach, M., Fathi, M., Holland, A., 2009. Knowledge-based feedback of product use information into product development. Proceedings of ICED 09, the 17th International Conference on Engineering Design, 227–238, 8.

Agard, B., Kusiak, A., 2004. Standardization of components, products and processes with data mining. International Conference on Production Research Americas, 1–9.

Ahmed, S., Kim, S., Wallace, K.M., 2007]. A methodology for creating ontologies for engineering design. J. Comput. Inf. Sci. Eng. 7 (2), 132–140.

Anderson, J., 1997]. Technology foresight for competitive advantage. Long Range Plann. 30 (5), 665–677.

Arksey, H., O'Malley, L., 2005]. Scoping studies: towards a methodological framework. Int. J. Soc. Res. Methodol. 8 (1), 19–32.

Armstrong, R., Hall, B.J., Doyle, J., Waters, E., 2011]. Scoping the scope of a Cochrane review. J. Public Health 33 (1), 147–150.

Bailey, M., Clothier, M., Gebbie, N., 2006. Realtime dome imaging and interaction: towards immersive design environments. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, January (Vol. 42578, pp. 813–821).

Balochian, S., Baloochian, H., 2019]. Social mimic optimization algorithm and engineering applications. Expert Syst. Appl. 134, 178–191.

Barclift, M., Simpson, T.W., Alessandra Nusiner, M., Miller, S., January 2017]. An investigation into the driving factors of creativity in design for additive manufacturing. In: ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

Becattini, N., Borgianni, Y., Cascini, G., Rotini, F., 2013]. Question/answer techniques within CAD environments: an investigation about the most effective interfaces. Comput. Des. Appl. 10 (6), 905–917.

Beltagy, I., Lo, K., Cohan, A., arXiv preprint, arXiv:1903.10676 2019]. SciBERT: A Pretrained Language Model for Scientific Text.

Bertoni, A., Larsson, T., Larsson, J., Elfsberg, J., 2017]. Mining data to design value: a demonstrator in early design. In: DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology, Vancouver, Canada, 21–25.08. 2017, pp. 021–029.

Boden, M.A., 1998]. Creativity and artificial intelligence. Artif. Intell. 103 (1–2), 347–356.

Borgue, O., Panarotto, M., Isaksson, O., 2019]. Modular product design for additive manufacturing of satellite components: maximising product value using genetic algorithms. Concurr. Eng. 27 (4), 331–346.

Brown, D.C., 2005]. Artificial intelligence for design process improvement. In: Design Process Improvement. Springer, London, pp. 158–173.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Agarwal, S., arXiv preprint arXiv:2005.14165 2020]. Language Models are Few-Shot Learners.

Bryant, C.R., McAdams, D.A., Stone, R.B., Kurtoglu, T., Campbell, M.I., 2005. A computational technique for concept generation. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, January (Vol. 4742, pp. 267-276).

Burckhard, S.R., Wampol, C., June 2018]. Board 41: using 3-D printing in a laboratory setting to teach design principles. 2018 ASEE Annual Conference & Exposition.

Carleton, T., Leifer, L., 2009]. Stanford's ME310 course as an evolution of engineering design. In: Proceedings of the 19th CIRP Design Conference–Competitive Design, Cranfield University Press.

Carraro, F., Lopez, R.H., Miguel, L.F.F., 2017]. Optimum design of planar steel frames using the search Group Algorithm. J. Braz. Soc. Mech. Sci. Eng. 39 (4), 1405–1418.

Cascini, G., 2012]. TRIZ-based anticipatory design of future products and processes. J. Integr. Des. Process. Sci. 16 (3), 29–63.

Cascini, G., Russo, D., 2007]. Computer-aided analysis of patents and search for TRIZ contradictions. Int. J. Prod. Dev. 4 (1–2), 52–67.

Cascini, G., Rotini, F., Russo, D., 2009]. Functional modeling for TRIZ-based evolutionary analyses. In: DS 58-5: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 5, Design Methods and Tools (Pt. 1), Palo Alto, CA, USA, 24.-27.08. 2009.

Cascini, G., Rotini, F., Russo, D., 2011. Networks of trends: systematic development of system evolution scenarios. 8th ETRIA TRIZ Future Conference, 355–367.

Chakrabarti, A., 2001]. Improving efficiency of procedures for compositional synthesis by using bidirectional search. Ai Edam 15 (1), 67–80.

Chakrabarti, A., Lindemann, U., 2016. Impact of Design Research on Industrial Practice. Springer.

Chang, W., Cheng, J., Allaire, J.J., Xie, Y., McPherson, J., 2020. Shiny: Web Application Framework for R. R Package Version 1.5.0. http://shiny.rstudio.com.

Chen, L., Shadbolt, N.R., Goble, C.A., 2006a]. A semantic web-based approach to knowledge management for grid applications. IEEE Trans. Knowl. Data Eng. 19 (2), 283–296.

Chen, L., Shadbolt, N.R., Goble, C., Tao, F., 2006b]. Managing semantic metadata for web/grid services. Int. J. Web Serv. Res. (IJWSR) 3 (4), 73–94.

Chen, L., Xu, S., Zhu, L., Zhang, J., Lei, X., Yang, G., 2020. A deep learning based method for extracting semantic information from patent documents. Scientometrics 125 (1), 289–312.

Cheng, M.Y., Tsai, H.C., Hsieh, W.S., 2009]. Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. Autom. Constr. 18 (2), 164–172.

Chiarello, F., Fantoni, G., Bonaccorsi, A., 2017]. Product description in terms of advantages and drawbacks: exploiting patent information in novel ways. In: DS 87-6 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 6: Design Information and Knowledge, Vancouver, Canada, 21–25.08. 2017, pp. 101–110.

Chiarello, F., Cimino, A., Fantoni, G., Dell'Orletta, F., 2018a]. Automatic users extraction from patents. World Pat. Inf. 54, 28–38.

Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G., 2018b]. Extracting and mapping industry 4.0 technologies using wikipedia. Comput. Ind. 100, 244–257.

Chiarello, F., Cirri, I., Melluso, N., Fantoni, G., Bonaccorsi, A., Pavanello, T., 2019. Approaches to automatically extract affordances from patents. In: Proceedings of the Design Society: International Conference on Engineering Design, Cambridge University Press, July (Vol. 1, No. 1, pp. 2487–2496).
Chiarello, F., Bonaccorsi, A., Fantoni, G., 2020]. Technical sentiment analysis. Measuring advantages and drawbacks of new products using social media. Comput. Ind. 123, 103299.
Chiesa, V., Frattini, F., Lazzarotti, V., Manzini, R., 2007]. Measuring performance in new product development projects: a case study in the aerospace industry. Proj. Manag. J. 38 (4), 45–59.
Chiu, M.C., Kremer, G.E.O., 2011]. Investigation of the applicability of Design for X tools during design concept evolution: a literature review. Int. J. Prod. Dev. 13 (2), 132–167.
Chong, W.K., Naganathan, H., Liu, H., Ariaratnam, S., Kim, J., 2018. Understanding infrastructure resiliency in Chennai, India using Twitter's Geotags and texts: a preliminary study. Engineering 4 (2), 218–223.
Cleveland, W.S., 2001]. Data science: an action plan for expanding the technical areas of the field of statistics. Int. Stat. Rev. 69 (1), 21–26.
Colton, S., Wiggins, G.A., 2012]. Computational creativity: the final frontier? Ecai, August (Vol. 12, pp. 21–26).
Corso, M., Martini, A., Paolucci, E., Pellegrini, L., 2001]. Knowledge management in product innovation: an interpretative review. Int. J. Manag. Rev. 3 (4), 341–352.
Costa, R., Sobek, D.K., 2004. How process affects performance: an analysis of student design productivity. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, January (Vol. 46962, pp. 137–143).
Cuprak, R., Rajadas, J., Georgeou, T., 2008. Experimental study of waste egress from collection vehicle. ASEE Annual Conference and Exposition, Conference Proceedings.
Danese, P., Filippini, R., 2010]. Modularity and the impact on new product development time performance. Int. J. Oper. Prod. Manage. 30 (11), 1191–1209.
Dasari, S.K., Lavesson, N., Andersson, P., Persson, M., 2015. Tree-based response surface analysis. In: International Workshop on Machine Learning, Optimization and Big Data, Springer, Cham, pp. 118–129.
Davis, D., Gentili, K., Trevisan, M., Calkins, D., 2002]. Engineering design assessment processes and scoring scales for program improvement and accountability. J. Eng. Educ. 91 (2), 211–221.
De Luca, G., Li, Z., Mian, S., Chen, Y., et al., 2018. Visual programming language environment for different IoT and robotics platforms in computer science education. CAAI Trans. Intell. Technol. 3 (2), 119–130.
De Mauro, A., Greco, M., Grimaldi, M., 2015]. What is big data? A consensual definition and a review of key research topics. In: AIP Conference Proceedings, American Institute of Physics (Vol. 1644, No. 1, pp. 97-104).
De Toni, A., Nassimbeni, G., 2001]. A method for the evaluation of suppliers' co-design effort. Int. J. Prod. Econ. 72 (2), 169–180.
Deb, K., 2000]. An efficient constraint handling method for genetic algorithms. Comput. Methods Appl. Mech. Eng. 186 (2–4), 311–338.
Deb, K., Pratap, A., Moitra, S., September 2000. Mechanical component design for multiple objectives using elitist non-dominated sorting ga. In: International Conference on Parallel Problem Solving from Nature. Springer, Berlin, Heidelberg, pp. 859–868.
Dering, M.L., Tucker, C.S., 2015. A computer vision approach for automatically mining and classifying end of life products and components. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, August (Vol. 57113, p. V004T05A007).
Desrochers, A., Cherkaoui, S., September 2002. A multi-agent architecture for automated product technical specifications verification in CAD environments. In: The 7th International Conference on Computer Supported Cooperative Work in Design, IEEE, pp. 236–240.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., arXiv preprint, arXiv:1810.04805 2018]. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.
Dieter, G.E., Schmidt, L.C., 2009]. Engineering Design. McGraw-Hill Higher Education., Boston.
Dudas, C., Ng, A.H., Pehrsson, L., Boström, H., 2014]. Integration of data mining and multi-objective optimisation for decision support in production systems development. Int. J. Comput. Integr. Manuf. 27 (9), 824–839.
Dym, C.L., Agogino, A.M., Eris, O., Frey, D.D., Leifer, L.J., 2005]. Engineering design thinking, teaching, and learning. J. Eng. Educ. 94 (1), 103–120.
Eddy, D., Krishnamurty, S., Grosse, I., Witherell, P., Wileden, J., Lewis, K., 2014]. An integrated approach to information modeling for the sustainable design of products. J. Comput. Inf. Sci. Eng. 14 (2).
Fantoni, G., Apreda, R., Dell'Orletta, F., Monge, M., 2013]. Automatic extraction of function–behaviour–state information from patents. Adv. Eng. Inform. 27 (3), 317–334.
Fantoni, G., Coli, E., Chiarello, F., Apreda, R., Dell'Orletta, F., Pratelli, G., 2020]. Text mining tool for translating terms of contract into technical specifications: development and application in the railway sector. Comput. Ind. 124, 103357.
Fareri, S., Fantoni, G., Chiarello, F., Coli, E., Binda, A., 2020]. Estimating Industry 4.0 impact on job profiles and skills using text mining. Comput. Ind. 118, 103222.
Ferguson, D.M., Ohland, M.W., 2012]. What is engineering innovativeness? Int. J. Eng. Educ. 28 (2), 253–262.
Field, B.W., 2007]. Visualization, intuition, and mathematics metrics as predictors of undergraduate engineering design performance. J. Mech. Des.-Trans. ASME 129 (7), 735–743.
Fiorineschi, L., Frillici, F.S., Rotini, F., 2020]. Challenging COVID-19 with creativity: supporting design space exploration for emergency ventilators. Appl. Sci. 10 (14), 4955.
Gao, L., Shao, B., Zhu, L., Lu, T., Gu, N., 2008]. Maintaining time and space consistencies in hybrid CAD environments: framework and algorithms. Comput. Ind. 59 (9), 894–904.
Gartner, Inc. IT glossary, 2018. Referred 03.12.2020, Available at https://www.gartner.com/it-glossary/.
Gattullo, M., Uva, A.E., Fiorentino, M., Scurati, G.W., Ferrise, F., 2017]. From paper manual to AR manual: do we still need text? Procedia Manuf. 11, 1303–1310.
Gattullo, M., Scurati, G.W., Fiorentino, M., Uva, A.E., Ferrise, F., Bordegoni, M., 2019]. Towards augmented reality manuals for industry 4.0: a methodology. Robot. Comput. Manuf. 56, 276–286.
Georgiev, G.V., Georgiev, D.D., 2018]. Enhancing user creativity: semantic measures for idea generation. Knowledge Based Syst. 151, 1–15.
Gero, J.S., 1990]. Design prototypes: a knowledge representation schema for design. AI Mag. 11 (4), 26.
Giachetti, R.E., Young, R.E., Roggatz, A., Eversheim, W., Perrone, G., 1997]. A methodology for the reduction of imprecision in the engineering process. Eur. J. Oper. Res. 100 (2), 277–292.
Goel, A.K., Shu, L.H., 2015]. Analogical thinking: an introduction in the context of design. Ai Edam 29 (2), 133–134.
Gorgoglione, M., Petruzzelli, A.M., Panniello, U., 2018]. Innovation through tradition in the Italian coffee industry: an analysis of customers' perceptions. Rev. Manag. Sci. 12 (3), 661–682.
Groover, M., Zimmers, E.W.J.R., 1983]. CAD/CAM: Computer-Aided Design and Manufacturing. Pearson Education.
Grover, P., Kar, A.K., 2017]. Big data analytics: a review on theoretical contributions and tools used in literature. Glob. J. Flex. Syst. Manag. 18 (3), 203–229.
Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018]. A survey of methods for explaining black box models. ACM Comput. Surv. (CSUR) 51 (5), 1–42.
Gulrez, T., Tognetti, A., 2014]. A sensorized garment controlled virtual robotic wheelchair. J. Intell. Robot. Syst. 74 (3–4), 847–868.
Guo, Y., Peng, Y., Hu, J., 2013]. Research on high creative application of case-based reasoning system on engineering design. Comput. Ind. 64 (1), 90–103.
Gustavsen, M.A., Louka, M.N., 2012]. Supporting human factors engineering design review activities using virtual control room mockups. Topical on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT 2012), 22–26.
Guzey, S.S., Harwell, M., Moreno, M., Peralta, Y., Moore, T.J., 2017]. The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics. J. Sci. Educ. Technol. 26 (2), 207–222.
Haik, Y., Sivaloganathan, S., Shahin, T.M., 2018]. Engineering Design Process. Nelson Education.
Hales, C., Doctoral dissertation 1987]. Analysis of the Engineering Design Process in an Industrial Context. University of Cambridge.
Hales, C., 2005]. Adding value to design research. Ai Edam 86.
Hamza, F., Abderazek, H., Lakhdar, S., Ferhat, D., Yıldız, A.R., 2018]. Optimum design of cam-roller follower mechanism using a new evolutionary algorithm. Int. J. Adv. Manuf. Technol. 99 (5–8), 1267–1282.
Harwood, A.R., Revell, A.J., 2018]. Interactive flow simulation using Tegra-powered mobile devices. Adv. Eng. Softw. 115, 363–373.
Hassannezhad, M., Clarkson, P.J., 2017]. Internal and external involvements in integrated product development: a two-step clustering approach. In: Complex Systems Engineering and Development Proceedings of the 27th CIRP Design Conference, Elsevier BV, May (Vol. 60, pp. 253–260).
Hatchuel, Armand, 1996. Théories et modèles de la conception. Cours d'ingénierie de la conception.
Hill, A., Song, S., Dong, A., Agogino, A., September 2001. Identifying shared understanding in design using document analysis. In: Proceedings of the 13th International Conference on Design Theory and Methodology, American Society of Mechanical Engineers, pp. 9–12.
Howard, T.J., Culley, S.J., Dekoninck, E., 2008]. Describing the creative design process by the integration of engineering design and cognitive psychology literature. Des. Stud. 29 (2), 160–180.
Hu, Q., Li, F., Chen, C.F., 2014]. A smart home test bed for undergraduate education to bridge the curriculum gap from traditional power systems to modernized smart grids. IEEE Trans. Educ. 58 (1), 32–38.
Hu, J., Ma, J., Qi, J., Peng, Y., 2017]. Knowledge modelling and innovative analogy methodology of biologically inspired design. J. Mech. Eng. 53 (15), 21–31.
Huang, Z., Jin, Y., 2010]. A prior and data validation and adjustment scheme for Bayesian reliability analysis in engineering design. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, January (Vol. 44137, pp. 483–496).
Huque, Z., Jahingir, N., 2002. Application of collaborative optimization on a RBCC inlet/ejector system. 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, July, p. 3604.
Iandoli, L., Klein, M., Zollo, G., 2007]. Can We Exploit Collective Intelligence for Collaborative Deliberation? The Case of the Climate Change Collaboratorium. MIT Sloan School of Management, Working Paper 4675-08.
Isaksson, O., Bertoni, M., Hallstedt, S., Lavesson, N., 2015. Model based decision support for value and sustainability in product development. 20th International Conference on Engineering Design (ICED), Milan. The Design Society.
Januszka, M., Moczulski, W., 2011]. Augmented reality system for aiding engineering design process of machinery systems. J. Syst. Sci. Syst. Eng. 20 (3), 294.

Ji, W., AbouRizk, S.M., 2018]. Simulation-based analytics for quality control decision support: pipe welding case study. J. Comput. Civ. Eng. 32 (3), 05018002.
Jiang, H., Kwong, C.K., Liu, Y., Ip, W.H., 2015]. A methodology of integrating affective design with defining engineering specifications for product design. Int. J. Prod. Res. 53 (8), 2472–2488.
Jiao, J., Zhang, Y., 2005]. Product portfolio identification based on association rule mining. Comput. Des. 37 (2), 149–172.
Kale, I.R., Kulkarni, A.J., 2018]. Cohort intelligence algorithm for discrete and mixed variable engineering problems. Int. J. Parallel Emergent Distrib. Syst. 33 (6), 627–662.
Kamboj, V.K., Nandi, A., Bhadoria, A., Sehgal, S., 2020]. An intensify Harris Hawks optimizer for numerical and engineering optimization problems. Appl. Soft Comput. 89, 106018.
Kasdirin, H.A., Yahya, N.M., Tokhi, M.O., December 2015. Hybridizing firefly algorithm with invasive weed optimization for engineering design problems. In: 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), IEEE, pp. 1–6.
Kim, S., Yoon, B., 2021]. Patent infringement analysis using a text mining technique based on SAO structure. Comput. Ind. 125, 103379.
Kopena, J.B., Cera, C.D., Regli, W.C., 2005. Conceptual design knowledge management and the semantic web. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, January (Vol. 47403, pp. 1005-1013).
Krishnan, V., Ulrich, K.T., 2001]. Product development decisions: a review of the literature. Manage. Sci. 47 (1), 1–21.
Kudo, F., Yoshikawa, T., Furuhashi, T., June 2011]. A study on analysis of design variables in Pareto solutions for conceptual design optimization problem of hybrid rocket engine. In: 2011 IEEE Congress of Evolutionary Computation (CEC), IEEE, pp. 2558–2562.
Kuester, F., Hamann, B., Joy, K.I., 2001]. Virtualexplorer: a plugin-based virtual reality framework. In: Stereoscopic Displays and Virtual Reality Systems VIII, International Society for Optics and Photonics, June (Vol. 4297, pp. 436–442).
Kügler, P., Schleich, B., Wartzack, S., 2018. Consistent digitalization of engineering design–an ontology-based approach. In: DS 91: Proceedings of NordDesign 2018, Linköping, Sweden, 14th-17th August 2018.
Kumar, A., Verma, S., Jeng, J.Y., 2020]. Supportless lattice structures for energy absorption fabricated by fused deposition modeling. 3D Print. Addit. Manuf. 7 (2), 85–96.
Kuo, T.C., Huang, S.H., Zhang, H.C., 2001]. Design for manufacture and design for 'X': concepts, applications, and perspectives. Comput. Ind. Eng. 41 (3), 241–260.
Kusiak, A., 2018. Smart manufacturing. Int. J. Prod. Res. 56 (1-2), 508–517.
Kusiak, A., Huang, C.C., 1996]. Development of modular products. IEEE Trans. Compon. Packag. Manuf. Technol. Part A 19 (4), 523–538.
Kusiak, A., Salustri, F.A., 2007]. Computational intelligence in product design engineering: review and trends. IEEE Trans. Syst. Man Cybern. Part C 37 (5), 766–778.
Kusiak, A., Smith, M., 2007]. Data mining in design of products and production systems. Annu. Rev. Control 31 (1), 147–156.
Lantada, A., et al., 2013]. Towards successful project-based teaching-learning experiences in engineering education. Int. J. Eng. Educ. 29 (2), 476–490.
Lee, J.S., Hsiang, J., arXiv preprint, arXiv:1906.02124 2019]. Patentbert: Patent Classification with Fine-Tuning a Pre-Trained Bert Model.
Levine, D.I., Agogino, A.M., Lesniewski, M.A., 2016]. Design thinking in development engineering. Int. J. Contin. Eng. Educ. Life-long Learn. 32 (3), 1396–1406.
Li, Z., Tate, D., 2013. Interpreting design structure in patents using an ontology library. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, August (Vol. 55928, p. V005T06A004).
Liao, T.W., 2001]. Classification and coding approaches to part family formation under a fuzzy environment. Fuzzy Sets Syst. 122 (3), 425–441.
Liebowitz, J., 2001]. Knowledge management and its link to artificial intelligence. Expert Syst. Appl. 20 (1), 1–6.
Lin, H.K., Harding, J.A., 2007]. A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration. Comput. Ind. 58 (5), 428–437.
Lin, N., Kaipin, W., Xu, W., 2017]. An online matching customized method facing customers ' needs. J. Chongqing Univ. 40 (3), 25–33.
Lindemann, U., Maurer, M., Braun, T., 2008. Structural complexity management: an approach for the field of product design. Springer Science & Business Media.
Linsey, J.S., Markman, A.B., Wood, K.L., 2012]. Design by analogy: a study of the WordTree method for problem re-representation. J. Mech. Des. 134 (4).
Liu, Y., Liang, Y., Kwong, C.K., Lee, W.B., 2010]. A new design rationale representation model for rationale mining. J. Comput. Inf. Sci. Eng. 10 (3).
Liverani, A., Amati, G., Caligiana, G., 2004]. A CAD-augmented reality integrated environment for assembly sequence check and interactive validation. Concurr. Eng. 12 (1), 67–77.
Loginov, D., October 2010. Synergetic modelling: application possibilities in engineering design. In: Proc. 10th WSEAS International Conference on System Science and Simulation in Engineering (ICOSSSE'10), Iwate Prefectural University, Japan, pp. 111–116.
Lortal, G., Lewkowicz, M., Todirascu-Courtier, A., December 2006]. AnT&CoW: share, classify and elaborate documents by means of annotation. In: 2006 1st International Conference on Digital Information Management, IEEE, pp. 332–337.
Mann, D., 2001]. An introduction to TRIZ: the theory of inventive problem solving. Creat. Innov. Manag. 10 (2), 123–125.
Matelli, J.A., Bazzo, E., da Silva, J.C., 2009]. An expert system prototype for designing natural gas cogeneration plants. Expert Syst. Appl. 36 (4), 8375–8384.
McGoldrick, C., Shivaram, S., Huggard, M., June 2016]. Experiences of integrating UAVs into the curriculum through multidisci-plinary engineering projects. Proc. ASEE 123rd Annual Conf. and Exhibition, 26–29.
Messac, A., Chen, W., 2000. The engineering design discipline: is its confounding lexicon hindering its evolution? J. Eng. Eval. Cost Anal. Decis.-Based Des.: Status Prom. 3, 67–83.
Miao, C., Du, G., Jiao, R.J., Zhang, T., 2017]. Coordinated optimisation of platform-driven product line planning by bilevel programming. Int. J. Prod. Res. 55 (13), 3808–3831.
Miguel, L.F.F., Miguel, L.F.F., 2012]. Shape and size optimization of truss structures considering dynamic constraints through modern metaheuristic algorithms. Expert Syst. Appl. 39 (10), 9458–9467.
Miles, L.D., 2015]. Techniques of Value Analysis and Engineering. Miles Value Foundation.
Miller, D.P., 2004. Using robotics to teach computer programming & AI concepts to engineering students. Proceedings of the AAAI Spring Symposium on Accessible Hands-on Artificial Intelligence and Robotics Education.
Minsky, M., 1961]. Steps toward artificial intelligence. Proc. IRE 49 (1), 8–30.
Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M., 2017]. Salp Swarm Algorithm: a bio-inspired optimizer for engineering design problems. Adv. Eng. Softw. 114, 163–191.
Moehrle, M.G., Caferoglu, H., 2019. Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. Technol. Forecast. Soc. Change 146, 776–784.
Moon, S.K., Kumara, S.R., Simpson, T.W., 2006. Data mining and fuzzy clustering to support product family design. In: ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, pp. 317–325.
Munn, Z., Peters, M.D., Stern, C., Tufanaru, C., McArthur, A., Aromataris, E., 2018]. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med. Res. Methodol. 18 (1), 143.
Neeley Jr., W.L., Lim, K., Zhu, A., Yang, M.C., 2013. Building fast to think faster: exploiting rapid prototyping to accelerate ideation during early stage design. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, August (Vol. 55928, p. V005T06A022).
Nguyen, V.H., Huynh, T.N., Nguyen, T.P., Tran, T.T., 2020]. Single and multi-objective optimization of processing parameters for fused deposition modeling in 3D printing technology. Int. J. Automot. Mech. Eng. 17 (1), 7542–7551.
O'Halloran, B.M., Papakonstantinou, N., Van Bossuyt, D.L., July 2018. Assessing the consequence of cyber and physical malicious attacks in complex, cyber-physical systems during early system design. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), IEEE, pp. 733–740.
Ong, Y.S., Keane, A.J., 2002]. A domain knowledge based search advisor for design problem solving environments. Eng. Appl. Artif. Intell. 15 (1), 105–116.
Ostad-Ahmad-Ghorabi, H., Rahmani, T., Gerhard, D., 2011]. Integrating LCA into PDM for ecodesign. World Acad. Sci. Eng. Technol. 7 (81), 223–228.
Pahl, G., et al., 2007. In: Wallace, K., Blessing, L. (Eds.), Engineering Design: A Systematic Approach. Springer-Verlag, London.
Pajo, S., Vandevenne, D., Duflou, J., 2015]. Systematic online lead user identification-case study for electrical installations. Proceedings of the 20th International Conference on Engineering Design (ICED15) (Vol. 10, No. DS 80-10, pp. 1-8). DESIGN SOC.
Pennington, J., Socher, R., Manning, C.D., 2014]. Glove: global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543.
Pérez-Lara, M., Saucedo-Martínez, J.A., Marmolejo-Saucedo, J.A., Salais-Fierro, T.E., Vasant, P., 2018]. Vertical and Horizontal Integration Systems in Industry 4.0. Wireless Networks, pp. 1–9.
Peysakhov, M., Regli, W.C., 2003]. Using assembly representations to enable evolutionary design of Lego structures. Ai Edam 17 (2), 155–168.
Qiao, T., Zhang, J., Xu, D., Tao, D., 2019]. Mirrorgan: learning text-to-image generation by redescription. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1505–1514.
Qiu, H., Xu, Y., Gao, L., Li, X., Chi, L., 2016]. Multi-stage design space reduction and metamodeling optimization method based on self-organizing maps and fuzzy clustering. Expert Syst. Appl. 46, 180–195.
Quintana-Amate, S., Bermell-Garcia, P., Balcazar, L., Tiwari, A., A new knowledge sourcing framework to support KBE development 2015. DS 80-8 Proceedings of the 20th International Conference on Engineering Design (ICED 15) , Innovation and Creativity, Milan, Italy, 27–30.07. 15, pp. 111–120, Vol. 8.
Radj, B.M., Senthilvelan, T., 2006]. Statistical analysis of friction stir welded AA 5052-H34 weldments by applying taguchi technique. ARPN J. Eng. Appl. Sci. 11 (18), 11062–11067.
Rafiq, M.Y., Bugmann, G., Easterbrook, D.J., 2001]. Neural network design for engineering applications. Comput. Struct. 79 (17), 1541–1552.
Rahman, M.H., Xie, C., Sha, Z., 2019]. A deep learning based approach to predict sequential design decisions. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 59179.
Rapaccini, M., Saccani, N., Pezzotta, G., Burger, T., Ganz, W., 2013]. Service development in product-service systems: a maturity model. Serv. Ind. J. 33 (3–4), 300–319.

Regli, W.C., Kopena, J.B., Grauer, M., Simpson, T.W., Stone, R.B., Lewis, K., Bohm, M.R., Wilkie, D., Piecyk, M., Osecki, J., 2010]. Semantics for digital engineering archives supporting engineering design education. AI Mag. 31 (1), 37–50.
Reich, Y., Konda, S.L., Levy, S.N., Monarch, I.A., Subrahmanian, E., 1993]. New roles for machine learning in design. Artif. Intell. Eng. 8 (3), 165–181.
Ren, D., Lee, B., Höllerer, T., 2017]. Stardust: Accessible and Transparent Gpu Support for Information Visualization Rendering, Vol. 36. Computer Graphics Forum, pp. 179–188, June, No. 3.
Ritchie, G., 2007]. Some empirical criteria for attributing creativity to a computer program. Minds Mach. 17 (1), 67–99.
Romanowski, C.J., Nagi, R., 2004]. A data mining approach to forming generic bills of materials in support of variant design activities. J. Comput. Inf. Sci. Eng. 4 (4), 316–328.
Romanowski, C.J., Nagi, R., Sudit, M., 2006]. Data mining in an engineering design environment: OR applications from graph matching. Comput. Oper. Res. 33 (11), 3150–3160.
RStudio Team, 2020. RStudio: Integrated Development for R. *RStudio, PBC*, Boston, MA http://www.rstudio.com/.
Sadollah, A., Eskandar, H., Kim, J.H., 2015]. Water cycle algorithm for solving constrained multi-objective optimization problems. Appl. Soft Comput. 27, 279–298.
Saha, A., Ray, T., Smith, W., June 2011. Towards practical evolutionary robust multi-objective optimization. In: 2011 IEEE Congress of Evolutionary Computation (CEC), IEEE, pp. 2123–2130.
Savoia, A., 2011]. Pretotype It. Pobrane.
Savoie, T.B., Frey, D.D., 2012]. Detecting mistakes in engineering models: the effects of experimental design. Res. Eng. Des. 23 (2), 155–175.
Shilakes, C., Tylman, J., 1998. Merrill-Lynch Report on Enterprise Information Portals. Merrill-Lynch ltd.
Singh, H.K., 2016]. Development of optimization methods to deal with current challenges in engineering design optimization. AI Commun. 29 (1), 219–221.
Singh, V., Gu, N., 2012]. Towards an integrated generative design framework. Des. Stud. 33 (2), 185–207.
Sonalkar, N., Jung, M., Mabogunje, A., 2011]. Emotion in engineering design teams. In: Emotional Engineering. Springer, London, pp. 311–326.
Soni, D., Mathai, K.J., 2015. An efficient content based image retrieval system based on color space approach using color histogram and color correlogram. In: 2015 Fifth International Conference on Communication Systems and Network Technologies, IEEE, pp. 488–492.
Sridharan, P., Campbell, M.I., 2005]. A study on the grammatical construction of function structures. Ai Edam 19 (3), 139–160.
Stockton, D.J., Khalil, R.A., Mukhongo, M.L., 2013a]. Cost model development using virtual manufacturing and data mining: part I—methodology development. Int. J. Adv. Manuf. Technol. 66 (5–8), 741–749.
Stockton, D.J., Khalil, R.A., Mukhongo, L.M., 2013b]. Cost model development using virtual manufacturing and data mining: part II—comparison of data mining algorithms. Int. J. Adv. Manuf. Technol. 66 (9–12), 1389–1396.
Štorga, M., Andreasen, M.M., Marjanović, D., 2010]. The design ontology: foundation for the design knowledge exchange and management. J. Eng. Des. 21 (4), 427–454.
Suh, N.P., Suh, P.N., 1990]. The Principles of Design (No. 6). Oxford University Press.
Sun, Y., Xiong, H.G., 2014]. Function optimization based on quantum genetic algorithm. Res. J. Appl. Sci. Eng. Technol. 7 (1), 144–149.
Taboada, H.A., Espiritu, J.F., Coit, D.W., 2008]. MOMS-GA: a multi-objective multi-state genetic algorithm for system reliability optimization design problems. IEEE Trans. Reliab. 57 (1), 182–191.
Talatahari, S., Azizi, M., 2020]. Optimization of constrained mathematical and engineering design problems using chaos game optimization. Comput. Ind. Eng. 145 (1), 106560.
Talbi, E.G., 2009]. Metaheuristics: from Design to Implementation, vol. 74. John Wiley & Sons.
Tawhid, M.A., Savsani, V., 2018]. A novel multi-objective optimization algorithm based on artificial algae for multi-objective engineering design problems. Appl. Intell. 48 (10), 3762–3781.
Telenko, C., Wood, K., Otto, K., Rajesh Elara, M., Foong, S., Leong Pey, K., et al., 2016]. Designettes: an approach to multidisciplinary engineering design education. J. Mech. Des. 138 (2).
Tiwari, S., Teegavarapu, S., Summers, J.D., Fadel, G.M., 2009]. Automating morphological chart exploration: a multi-objective genetic algorithm to address compatibility and uncertainty. Int. J. Prod. Dev. 9 (1–3), 111–139.
Tjoa, A.M., Berger, L., 1993. Transformation of requirement specifications expressed in natural language into an EER model. In: International Conference on Conceptual Modeling. Springer, Berlin, Heidelberg, pp. 206–217.
Trappey, A.J., Trappey, C.V., Hsu, F.C., Hsiao, D.W., 2009]. A fuzzy ontological knowledge document clustering methodology. IEEE Trans. Syst. Man Cybern. Part B 39 (3), 806–814.
Tseng, M.M., Jiao, J., 1997]. A variant approach to product definition by recognizing functional requirement patterns. J. Eng. Des. 8 (4), 329–340.
Ulrich, K.T., 2003]. Product Design and Development. Tata McGraw-Hill Education.
Umeda, Y., Takeda, H., Tomiyama, T., Yoshikawa, H., 1990]. Function, behaviour, and structure. Appl. Artif. Intell. Eng. 1, 177–193.
Umeda, Y., Takata, S., Kimura, F., Tomiyama, T., Sutherland, J.W., Kara, S., Herrmann, C., Duflou, J.R., 2012]. Toward integrated product and process life cycle planning—an environmental perspective. CIRP Ann. 61 (2), 681–702.
Urbinati, A., Latilla, V.M., Chiaroni, D., 2018]. The role of product design in circular economy business model. In: ISPIM Conference Proceedings, The International Society for Professional Innovation Management (ISPIM), pp. 1–19.
Verganti, R., 2011]. Radical design and technology epiphanies: a new focus for research on design management. J. Prod. Innov. Manage. 28 (3), 384–388.
Von Hippel, E., 1989]. New product ideas from "lead users". Res.-Technol. Manage. 32 (3), 24–27.
Wang, et al., 2016]. Let them play: the impact of mechanics and dynamics of a serious game on student perceptions of learning engagement. IEEE Trans. Learn. Technol. 10 (4), 514–525.
Wang, L., Alexander, C.A., 2015]. Big data in design and manufacturing engineering. Am. J. Eng. Appl. Sci. 8 (2), 223.
Wang, C.B., Chen, Y.J., Chen, Y.M., Chu, H.C., 2005a]. Application of ART neural network to development of technology for functional feature-based reference design retrieval. Comput. Ind. 56 (5), 428–441.
Wang, B.P., Han, Z.X., Xu, L., Reinikainen, T., April, EuroSimE 2005 2005b]. A novel response surface method for design optimization of electronic packages. In: Proceedings of the 6th International Conference on Thermal, Mechanial and Multi-Physics Simulation and Experiments in Micro-Electronics and Micro-Systems, 2005, IEEE, pp. 175–181.
Wang, C., Yan, C.Q., Wang, J.J., Chen, L., Li, G.J., 2014]. Application of dual-adaptive niched genetic algorithm in optimal design of nuclear power components. In: International Conference on Nuclear Engineering, American Society of Mechanical Engineers, July (Vol. 45905, p. V02AT09A013).
Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., Chen, W., 2016]. Forecasting technological impacts on customers' co-consideration behaviors: a data-driven network analysis approach. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, August (Vol. 50107, p. V02AT03A040).
Whitfield, R.I., Duffy, A.H., Coates, G., Hills, W., 2003]. Efficient process optimization. Concurr. Eng. 11 (2), 83–92.
Wu, D., Rosen, D.W., Wang, L., Schaefer, D., 2015]. Cloud-based design and manufacturing: a new paradigm in digital manufacturing and design innovation. Comput. Des. 59, 1–14.
Xie, C., Zhang, Z., Nourian, S., Pallant, A., Bailey, S., 2014]. On the instructional sensitivity of CAD logs. Int. J. Eng. Educ. 30 (4), 760–778.
Xie, C., Schimpf, C., Chao, J., Nourian, S., Massicotte, J., 2018a]. Learning and teaching engineering design through modeling and simulation on a CAD platform. Comput. Appl. Eng. Educ. 26 (4), 824–840.
Xie, T., Jiang, P., Zhou, Q., Shu, L., Zhang, Y., Meng, X., Wei, H., 2018b]. Advanced multi-objective robust optimization under interval uncertainty using Kriging model and support vector machine. J. Comput. Inf. Sci. Eng. 18 (4).
Xie, Y., Dervieux, C., Riederer, R., 2020]. R Markdown Cookbook. Chapman and Hall/CRC.
Xue, D., Dong, Z., 1997]. Optimal fuel cell system design considering functional performance and production costs. In: In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.
Yan, X.T., Sawada, H., 2006]. A framework for supporting multidisciplinary engineering design exploration and life-cycle design using underconstrained problem solving. Artif. Intell. Eng. Des. Anal. Manuf.: AI EDAM 20 (4), 329.
Ye, X., Thevenot, H.J., Alizon, F., Gershenson, J.K., Khadke, K., Simpson, T.W., Shooter, S.B., 2009]. Using product family evaluation graphs in product family design. Int. J. Prod. Res. 47 (13), 3559–3585.
Yin, H., Fang, H., Wen, G., Gutowski, M., Xiao, Y., 2018]. On the ensemble of metamodels with multiple regional optimized weight factors. Struct. Multidiscip. Optim. 58 (1), 245–263.
Yu, T.L., Yassine, A.A., Goldberg, D.E., 2007]. An information theoretic method for developing modular architectures using genetic algorithms. Res. Eng. Des. 18 (2), 91–109.
Zhang, K., Althoefer, K., July 2019]. Designing origami-adapted deployable modules for soft continuum arms. In: Annual Conference Towards Autonomous Robotic Systems, Springer, Cham, pp. 138–147.
Zhang, Y., Du, X., 2014]. Automatic field data analyzer for closed-loop vehicle design. Inf. Sci. 259, 321–334.
Zhang, J., Chowdhury, S., Messac, A., 2012a]. Domain segmentation based on uncertainty in the surrogate (DSUS). 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA, p. 1929.
Zhang, X., Simpson, T., Frecker, M., Lesieutre, G., 2012b]. Supporting knowledge exploration and discovery in multi-dimensional data with interactive multiscale visualisation. J. Eng. Des. 23 (1), 23–47.
Zheng, J., Xing, W., Zhu, G., Chen, G., Zhao, H., Xie, C., 2020]. Profiling self-regulation behaviors in STEM learning of engineering design. Comput. Educ. 143, 103669.
Zhou, N., Kisselburgh, L., Chandrasegaran, S., Badam, S.K., Elmqvist, N., Peppler, K., Ramani, K., 2015]. Using real-time trace data to predict collaboration quality and creative fluency in design teams. In: Proceedings of the 11th International Conference on Computer Supported Collaborative Learning CSCL, Gothenburg, Sweden, June 7–11.
Zhou, Q., Wang, Y., Jiang, P., Shao, X., Choi, S.K., Hu, J., Cao, L., Meng, X., 2017]. An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems. Knowl.-Based Syst. 131, 10–27.