TechNet 2.0: Expanding Technology Semantic Network with Qualitative Relations to Enhance Reasoning Capabilities

Serhad Sarica[1*] **and Jianxi Luo**[1]
[1]*Data-Driven Innovation Lab*
Singapore University of Technology & Design, Singapore
**Corresponding author: serhad_sarica@alumni.sutd.edu.sg*

This paper introduces a new semantic network knowledge base that can support both explicit and implicit inferences across engineering design concepts. Our approach is to merge the previously constructed Engineering Knowledge Graph (EKG) and Technology Semantic Network (TechNet). The terms in EKG are mapped to TechNet and then their qualitative relations are added into TechNet. We call the synthesized new knowledge base TechNet 2.0. TechNet 2.0 contains both qualitative and quantitative relations among elemental engineering design concepts. We exemplify the structure of TechNet 2.0 and its new capabilities to enrich and augment design knowledge representation and reasoning. The new knowledge base may fill the infrastructure gap in the process of creating artificial intelligence agents that can support various design understanding, inference, and generation tasks, especially in the very-early phases of the product development lifecycle.

Introduction

Semantic network representations of knowledge are increasingly employed in the design process to support various design activities [1], [2]. In common semantic networks, nodes represent specific knowledge pieces or concepts, which are known as semantic entities. The nodes are connected to one another via links that are knowledge semantic relations, via which knowledge can be accessed from one another. Employing semantic networks to represent design knowledge has several advantages, such as empowering the reasoning, analysis, and synthesis of the knowledge contained in design documents or data by enhancing design knowledge inferences.

One example of such semantic networks is Technology Semantic Network (TechNet) [3] while another one is Engineering Knowledge Graph (EKG) [4]. Both are trained on the entire USPTO patent text database while they employ different methodologies to retrieve and relate the engineering terms, which are elemental engineering design concepts. While TechNet employed statistical methods to retrieve terms up to 4 words long, EKG employs a set of rules based on Part-of-Speech (POS) tags to identify multi-word units. TechNet trains a language model on processed patent texts to derive quantitative relations, based on the cosine of the terms' high-dimensional embedding vectors. Such quantities can facilitate numerical computation and reasoning, but their meanings are explicit. EKG employs hard-coded rules based on POS tags and syntactic parse tree of the sentences to retrieve qualitative relations between terms, <entity, relationship, entity> triplets. These qualitative relations are explicit and can be understood easily by humans.

The differences of these two knowledge bases are summarized in Table 1 with statistics. Fig. 1 illustrates the terms and their semantic relations of a specific patent, retrieved from TechNet and EKG respectively.

Table 1. Database statistics of TechNet, EKG, and TechNet 2.0

	# Terms	**# Relations**	**# Relation Types**
TechNet	4,038,924	8.5×10^{12}	1*
EKG	288,807,731	794,956,771	234,955
TechNet 2.0	4,038,924	242,191,653**	4,776

* TechNet only supports quantitative relations. The only relation between terms is semantic similarity.

** The number only indicates the qualitative relations. TechNet 2.0 also inherits the quantitative relations represented in TechNet.

Although TechNet has proven to be useful in certain reasoning settings through the utilization of pairwise quantitative semantic similarity and graph theory based techniques [5]–[7], it is limited to explain the relations between the engineering concepts qualitatively. Despite the usefulness of quantitative relation representations in supporting knowledge computing, the communication of the knowledge encoded in an embedding space to designers is a tedious task and the explainability of the quantitative relations is limited.

In this study, our aim is to leverage complementary knowledge represented in TechNet and EKG to create a combined knowledge base, which is an expansion to TechNet, to provide a better medium that can support both qualitative and quantitative reasoning tasks together. This enhanced knowledge base, namely TechNet 2.0, can represent the EKG-derived qualitative relations between technical terms together with their quantitative relations provided in TechNet.

In next sections, we will introduce the steps we employed to merge the two prior knowledge bases, depict the new knowledge graph structure, and provide examples of the functions that TechNet 2.0 can have. Finally, we will conclude the paper with our plans for future work and a discussion on the research and application opportunities that this knowledge base may provide.

Merging TechNet and EKG

Although they were both trained on the USPTO patent database, the two knowledge bases differ in two main aspects. First, while TechNet uses titles and abstracts of patents as the data source, EKG sources patent claims. Second, the methods to construct them differ. TechNet retrieves phrases up to 4-tokens long by relying on statistical significance of tokens appearing together in the patent text corpus. Despite the susceptibility of co-occurrence based statistical methods to noise, encoded filters in TechNet term-retrieval processes considerably reduced the noisy phrase formations. There exists a considerable number of noisy terms in TechNet. On the other hand, EKG builds its vocabulary by relying on Part-of-Speech (POS) tagging, syntactic parsing and hard-coded rules to derive <entity, relationship, entity> triplets. Despite rule-based methods' popularity on determining well-structured text within corpora of a limited size, they lack generalization. This issue can also be observed in EKG, where very long and diverse phrases, which sometimes do not constitute a meaningful multi-word unit, are treated as technically meaningful terms. In addition to multi-word unit, we can observe different kinds of noise in relation types, such as many misspelled words and multiple inflections of same words. Hence, both the vocabulary size and number of relations are extremely large and noisy.

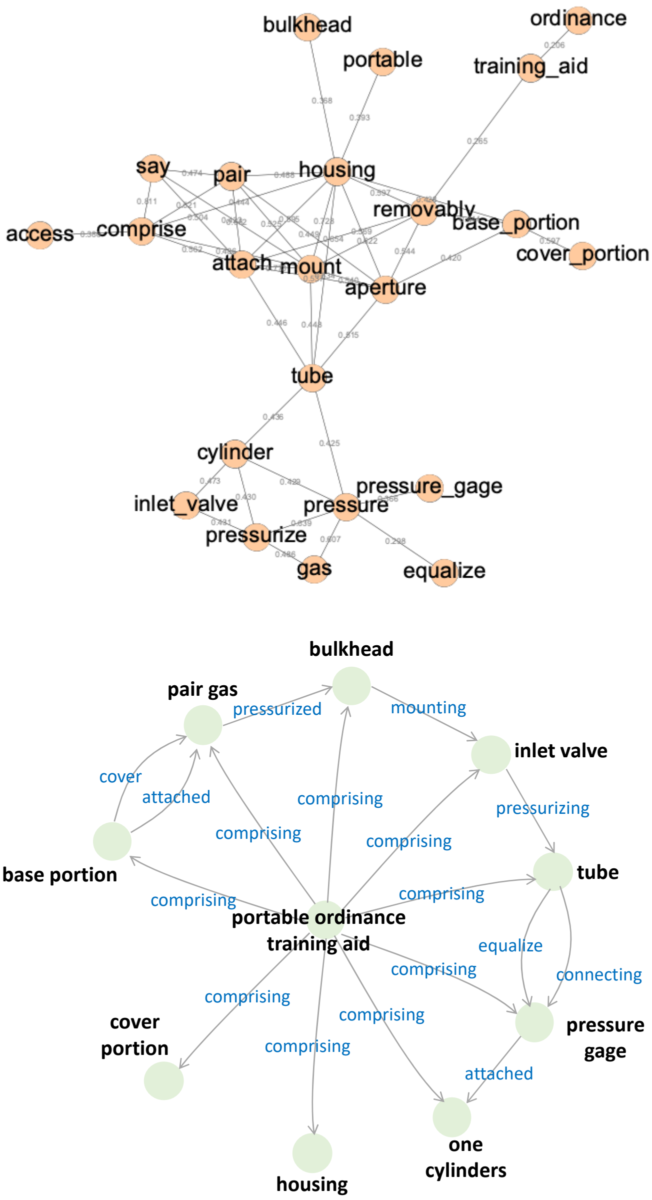


Fig 1. Visualizations from TechNet and EKG. *Top:* Terms and relations in claims of the patent US4014111 visualized by methods described in Sarica et al. [3] using TechNet. The numbers on edges are cosine similarities between terms *Bottom:* Terms and relations in claims of US401411 visualized by EKG (adopted from Siddharth et al. [4]).

Since TechNet covers a considerably smaller vocabulary than EKG, we determined to create a mapping from EKG and TechNet. We tried to directly map phrases up to 4-tokens long while we assumed that the last four tokens of the phrases longer than 4-token are the main technical term, while the tokens prior to the last 4-tokens considered as the modifiers of the phrase detected in last 4-tokens. Fig. 2 presents a graphical example of mapping a specific term's EKG representation to TechNet 2.0 representation. Since EKG has more and longer terms in its dictionary, the mappings are either one-to-one or many-to-one from EKG to TechNet. For example, both "infrared electromagnetic signal", and "jamming electromagnetic signal" terms are mapped to "electromagnetic signal". To retain the knowledge, we relate both "infrared" and "jamming" to "electromagnetic signal" through a "modified_by" relation.

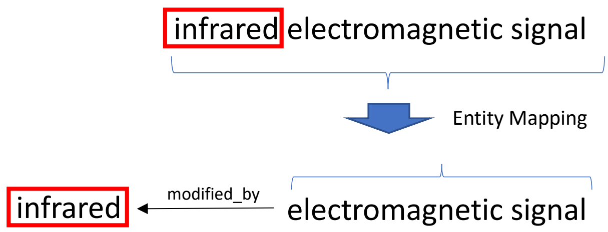


Fig 2. Example of the entity mapping procedure from EKG to TechNet.

Because the many-to-one mapping can create ambiguity and confusion in subsequent tasks such as design knowledge retrieval and representation, we retain the patent and classification information in TechNet 2.0. As a result, patent classes are on top of the knowledge graph hierarchy whereas patents are children of patent classes and terms are children of patents. Hence, the new knowledge graph database structure supports searching similar semantic relations in different technology domains. In addition, retaining the classification information in the semantic network will support merging information in other innovation and patent informatics studies, such as the network information covered in a series of studies by Luo et al. [8], [9].

We used the verbs contained in EKG and represented in TechNet as the relation types in TechNet 2.0. As a result, 5,462 verbs (or relation types) are detected. We mapped the relations of EKG if the lemmatized version of these relations appears in the set of TechNet 2.0 relation types. For example, EKG relations "expand, expanding, expanded" are all mapped to "expand" relation. As a result, we have 4,776 qualitative relations in TechNet 2.0.

Following the procedures described above, we mapped 54,286,973 EKG terms to 2,189,480 TechNet terms and produced 242,191,653 relations among which 191,972,609 stand for the SOA triplets that can be represented through TechNet terms and 27,876,516 are the "modified_by" relations which account for the noun and adjective modifiers of their parent terms (see Fig.1 for example). In brief, TechNet 2.0 contains 4,038,924 terms and 242,191,653 qualitative relations among the terms, together with the pair-wise quantitative relations among all terms.

Fig. 3 represents a sub-graph of TechNet 2.0 surrounding the term "autonomous vehicle". We collected 10 different terms that draw the strongest relations to "autonomous vehicle". The connection strength is based on the qualitative relation in TechNet 2.0. In addition, we collected next strongest relations among these 10 terms so that each node has at least two edges. In fact, there are many more relations between the presented nodes, but we kept the graph sparse for a better visualization. The term network is non-hierarchical as illustrated in Fig. 3. The network can be expanded through the strongest relations as well as through more structured queries supported by the qualitative relation-based reasoning capabilities of TechNet 2.0.

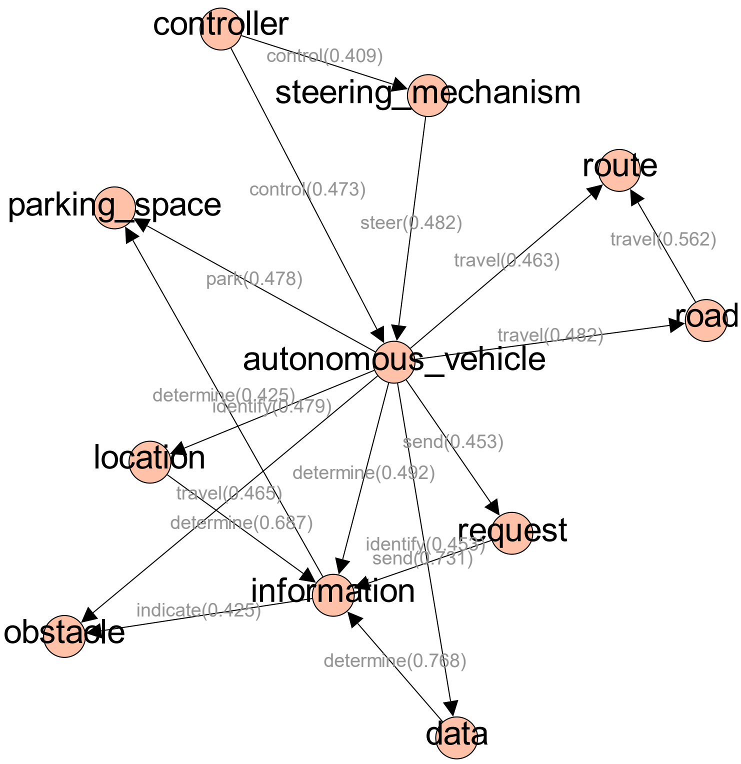
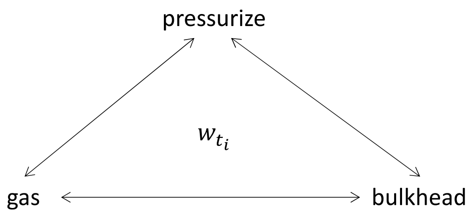


Fig 3. A subgraph of TechNet 2.0 surrounding the term "autonomous vehicle".

TechNet 2.0: New Reasoning Capabilities

TechNet 2.0 will continue to provide the basic functions of TechNet (http://www.tech-net.org/) based on quantitative inter-term semantic similarity. This key feature is leveraged in TechNet 2.0 to assign relatedness measures to the relations within and among triplets. The relatedness of a relation triplet <term1, relation, term2>, w_{t_i}, where t_i is a specific triplet, is calculated as the mean of semantic similarities between the terms that constitute the triplet. This measurement is illustrated with an example in Fig 4. Using this method, we calculated every single triplet's relatedness. The relatedness of a triplet can indicate the overall commonness or novelty of the terms and their relation through the verb or action. Given the relatedness values of triplets, we can search among a single term's qualitative relations (actions) to other terms and sort them in terms of each triplet's novelty or commonness.



$$w_{t_i} = \frac{1}{3} \sum_{\forall w, w' \in t_i, w \neq w'} cosine\_sim(w, w')$$

Fig 4. A relation triplet example, <gas, pressurize, bulkhead> and how its relatedness measure is calculated.

Fig. 5 illustrates how the new features can be used to added graph inference. First, the claims of the patent US401411 are represented by blue and white nodes. We can visually compare this part of the visualizations with the EKG representation given in Fig. 1 to draw an understanding about the term and relation mapping methods employed to build TechNet 2.0. As an example of entity mapping, we can observe how “portable ordinance training aid” is mapped to “training_aid” along with new “modified_by” relation to form the <“training_aid”, “modified_by”, “portable”> and <“training_aid”, “modified_by”, ordinance> triplets while retaining the already encoded knowledge in EKG. In addition, multiple forms of qualitative relations such as “comprising” and “connecting” are mapped to their root forms such as “comprise”, “connect”.

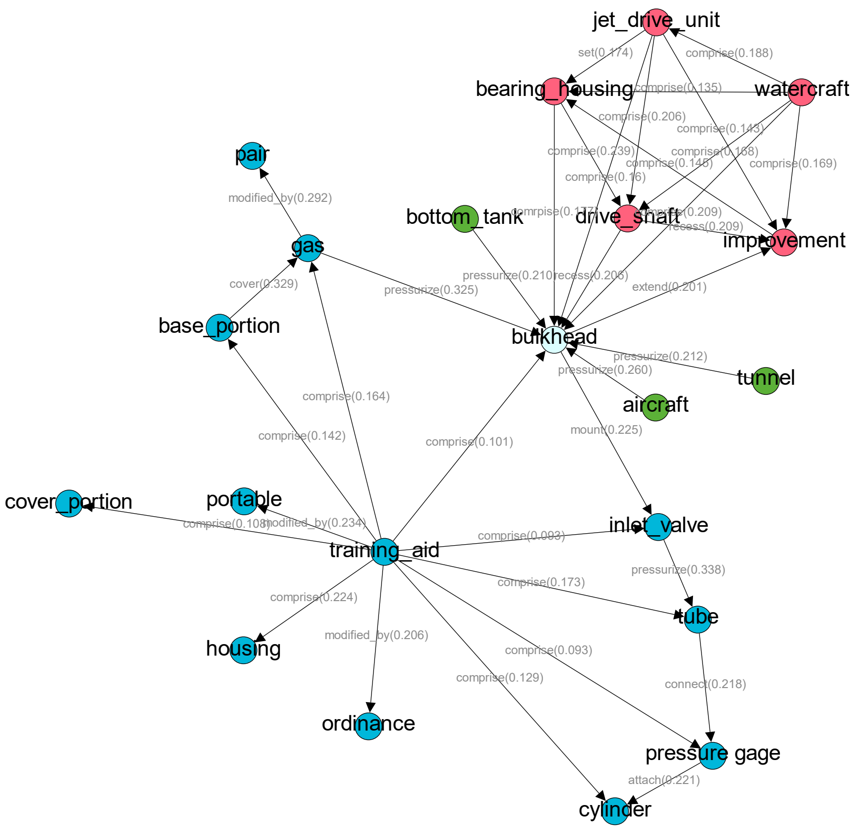


Fig 5. The terms and relations in the claims of the patent US401411 visualized by TechNet 2.0 (blue and white nodes), terms in US4813898 claims (red and white nodes) which are one hop away from the common ‘bulkhead’ term and other terms in TechNet 2.0 which have the ‘pressurize’ relation to the ‘bulkhead’ term (green nodes). The layout of the graph is constructed by Force Atlas algorithm [10]. The numbers in the brackets denote the triplet relatedness, i.e., w_{t_i}.

We used the “bulkhead” term as an anchor point to make inference and expanded the graph of the US401411 patent, which is primarily classified in F01D21/003 (Arrangements for testing or measuring) one hop towards another patent US4813898 which is classified in B63H23/34 (Propeller shafts; Paddle-wheel shafts; Attachment of propellers on shafts). In the previous version of TechNet, such structured explorations and queries were impossible since the semantic network was flat and fully connected. On the contrary, TechNet 2.0 is a multi-graph which consists of different types of nodes and relations.

In addition to targeted queries which let the graph grow towards other patents or technology domains, we also introduced a function to search for terms connected to a specific term via a specific relation. An example is given again for the "bulkhead" term in Fig. 5. When we queried the other terms which "pressurize" the "bulkhead" component, the query returned three terms, namely "aircraft", "bottom tank" and "tunnel".

Discussions and Outlook

The new knowledge base, TechNet 2.0, leverages and synthesizes two previously constructed large design knowledge bases to aid engineering design research and practice. TechNet 2.0 addresses the limitation of TechNet – it provides only quantitative relations between terms despite its fully-connected structure. TechNet 2.0 adds to the original TechNet with the compact and human-readable qualitative relational representation of the knowledge space created by EKG.

Meanwhile, those readily available public knowledge bases such as WordNet [11] and ConceptNet [12] can provide common-sense knowledge to support general knowledge inference applications. Our vision is to work towards developing such a knowledge base to capture comprehensive engineering design knowledge reliably and serve the design research community and industry. There have been increased inclination towards creating semantic representations of engineering design knowledge by the availability of state-of-the-art data science and database infrastructure [13]–[16]. The findings, successes or failures, datasets, and the know-how of such studies may be consolidated to introduce robust data retrieval and validation procedures to achieve our future goals.

In the context of computational methods in engineering design, a comprehensive and accurate knowledge infrastructure is essential for creating intelligent engineering design agents which can contextualize a problem conceptually and functionally and search for potential solutions. A very knowledgeable AI partner in design teams, actively involved in brainstorming activities, can increase the innovation speed and lead humanity to new horizons. Yet, our community could only bring forward either limited knowledge bases or comprehensive ones with limited capabilities. Our new knowledge base and upcoming iterations and updates can potentially fill this gap and provide the necessary infrastructure to AI applications in engineering design to support various design tasks, especially the ones in the early product development phases, such as knowledge exploration and representation, creative reasoning, gap finding, brainstorming, concept generation, and more. Such progresses and new capabilities further empower creative artificial intelligence for design and data-driven innovation [17].

Last but not the least, the comprehensiveness of the knowledge bases is directly related to the data source. One limitation of the current engineering design-based studies is that they mostly rely on textual data sources, especially on the patent database. Research on using other kinds of data sources, such as images or videos, to enable multi-modal representation and learning may provide additional opportunities to support more accurate, nuanced, and meaningful reasoning and decision-making tasks.

References

[1] J. M. Sowa, 'Semantic Networks, Encyclopedia of Artificial Intelligence, Wiley .', 1992.

[2] J. Han, S. Sarica, F. Shi, and J. Luo, 'Semantic Networks for Engineering Design: State of the Art and Future Directions', *J. Mech. Des.*, pp. 1–45, Aug. 2021, doi: 10.1115/1.4052148.

[3] S. Sarica, J. Luo, and K. L. Wood, 'TechNet: Technology semantic network based on patent data', *Expert Syst. Appl.*, vol. 142, p. 112995, Mar. 2020, doi: 10.1016/j.eswa.2019.112995.

[4] L. Siddharth, L. T. M. Blessing, K. L. Wood, and J. Luo, 'Engineering Knowledge Graph From Patent Database', *J. Comput. Inf. Sci. Eng.*, vol. 22, no. 2, Apr. 2022, doi: 10.1115/1.4052293.

[5] S. Sarica, B. Song, E. Low, and J. Luo, 'Engineering Knowledge Graph for Keyword Discovery in Patent Search', in *Proceedings of the Design Society*, Delft, 2019, vol. 1, pp. 2249–2258. doi: 10.1017/dsi.2019.231.

[6] S. Sarica and J. Luo, 'DESIGN KNOWLEDGE REPRESENTATION WITH TECHNOLOGY SEMANTIC NETWORK', *Proc. Des. Soc.*, vol. 1, pp. 1043–1052, Aug. 2021, doi: 10.1017/pds.2021.104.

[7] S. Sarica, B. Song, J. Luo, and K. L. Wood, 'Idea generation with Technology Semantic Network', *AI EDAM*, pp. 1–19, 2021, doi: 10.1017/S0890060421000020.

[8] J. Luo, B. Yan, and K. Wood, 'InnoGPS for Data-Driven Exploration of Design Opportunities and Directions: The Case of Google Driverless Car Project', *J. Mech. Des.*, vol. 139, no. 11, p. 111416, Oct. 2017, doi: 10.1115/1.4037680.

[9] J. Luo, S. Sarica, and K. L. Wood, 'Guiding data-driven design ideation by knowledge distance', *Knowl.-Based Syst.*, vol. 218, p. 106873, Apr. 2021, doi: 10.1016/j.knosys.2021.106873.

[10] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, 'ForceAtlas2 , a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', vol. 9, no. 6, pp. 1–12, 2014, doi: 10.1371/journal.pone.0098679.

[11] G. A. Miller, ‘WordNet: a lexical database for English’, *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995, doi: 10.1145/219717.219748.
[12] R. Speer, J. Chin, and C. Havasi, ‘ConceptNet 5.5: An Open Multilingual Graph of General Knowledge’, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1612.03975
[13] J. Hao, L. Zhao, J. Milisavljevic-Syed, and Z. Ming, ‘Integrating and navigating engineering design decision-related knowledge using decision knowledge graph’, *Adv. Eng. Inform.*, vol. 50, p. 101366, Oct. 2021, doi: 10.1016/j.aei.2021.101366.
[14] H. Zuo, Y. Yin, and P. Childs, ‘Patent-KG: Patent Knowledge Graph Use for Engineering Design’, *ArXiv210811899 Cs*, Sep. 2021, Accessed: Nov. 03, 2021. [Online]. Available: http://arxiv.org/abs/2108.11899
[15] A. Huet, R. Pinquié, P. Véron, A. Mallet, and F. Segonds, ‘CACDA: A knowledge graph for a context-aware cognitive design assistant’, *Comput. Ind.*, vol. 125, p. 103377, Feb. 2021, doi: 10.1016/j.compind.2020.103377.
[16] W. Jiang, Y. Wang, J. Hu, L. Guan, and Z. Zhu, ‘Construction of Substation Engineering Design Knowledge Graph Based on “Ontology Seven-step Method”’, in *2021 4th International Conference on Energy, Electrical and Power Engineering (CEEPE)*, Apr. 2021, pp. 957–962. doi: 10.1109/CEEPE51765.2021.9475682.
[17] J. Luo, ‘Data-Driven Innovation: What is it?’, *IEEE Trans. Eng. Manag.*, pp. 1–7, 2022, doi: 10.1109/TEM.2022.3145231.