Scientometrics
https://doi.org/10.1007/s11192-025-05491-z

Technological impact-guided technology opportunity analysis using a generative-predictive machine learning model

Gyumin Lee[1] · **Changyong Lee**[2]

Received: 23 December 2024 / Accepted: 14 November 2025

Abstract

Although patent analysis-based approaches to technology opportunity analysis have proven useful for discovering unexplored technological ideas, identifying opportunities to maximise the potential of existing technologies remains challenging. This study introduces an analytical framework for identifying new technological domains where existing technologies can exhibit greater technological impact. The proposed framework incorporates a generative–predictive machine learning model integrating the variational autoencoder and multi-layer perceptron architectures. The generative and predictive components of this model are jointly trained to construct an impact-centric technology landscape where technologies with similar domains and impacts are closely located. A gradient ascent search algorithm is used to explore this landscape and identify new technological domains that can maximise the potential technological impact of existing technologies. An empirical analysis covering 133,654 patents related to artificial intelligence technology verifies the reliability and feasibility of the framework in identifying domain-shift opportunities for existing technologies. The proposed analytical framework is expected to serve as a valuable tool that helps firms maximise their existing technological assets in the current era of open innovation.

Keywords Technology opportunity analysis · Technological impact · Generative machine learning model · Predictive machine learning model · Latent space exploration

✉ Changyong Lee
changyonglee@korea.ac.kr

Gyumin Lee
gyuminlee@dgu.ac.kr

[1] Department of Management Information Systems, Dongguk University, 30 Pildon-Ro 1-Gil, Jung-Gu, Seoul 04620, Republic of Korea

[2] Department of Public Administration, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea

Published online: 06 December 2025

Introduction

Technology opportunity analysis (TOA) has been gaining increasing interest from academics and practitioners as it represents a crucial requirement for firms to foster economic growth, sustain competitiveness, and mitigate the uncertainties involved in new businesses (Adams et al., 2006; Lee et al., 2014, 2018, 2023; Liu et al., 2023; Savino et al., 2017). Whereas early TOA-related research focused on developing expert-centric approaches (Jeong & Kim, 1997), recent literature highlights a shift towards developing data-driven approaches (Lee, 2021; Lee et al., 2018). Among such methods, patent analysis is considered the most scientific, with recent advancements in machine learning enabling the extraction of significant patterns and insights from extensive patent data across diverse fields (Lee et al., 2009, 2013a; Son et al., 2012; Wang et al., 2017; Yoon & Kim, 2012). However, while patent analysis-based approaches to TOA have proven useful for identifying new technological ideas that have not yet been fully explored, identifying opportunities that maximise the potential of existing technologies remains a key challenge.

The impact of technology varies depending on the domain in which it is implemented, even when its functionality remains unchanged. This suggests that a shift in technological domains can create promising technology opportunities (Dew et al., 2004). Historically, many significant inventions have emerged when the functions of existing technologies have been applied in different contexts (Andriani et al., 2017; Hur, 2017). For example, Autographer—developed as a hands-free, wearable intelligent camera with sensors that automatically determine when to capture photographs—was originally intended for use in healthcare, to assess memory recall in patients with memory impairments (Garud et al., 2016). Subsequently, it was repurposed for consumer use when its potential as a life-logging device became apparent. This domain shift expanded Autographer's customer base from a limited patient group to the broader public.

Building on this notion, exploring other technological domains where existing technologies could create a greater impact has become a key strategy for identifying innovative technology opportunities (Andriani et al., 2017; Dew et al., 2004). Regarding domain shifts as sources of technological innovation, many researchers have attempted to conceptualise the transition of technological domains in the context of technological change, investigating the effects of this process and the possibilities that it offers in the discovery of emerging technologies (Andriani & Carignani, 2014; Andriani et al., 2017; Ardito et al., 2021; Dew, 2009; Dew et al., 2004; Garud et al., 2016). Garud et al. (2016) introduced the concept of repurposing existing technologies by drawing analogies between biological evolution and technological change. Andriani et al. (2017) proposed an analytical framework to measure the frequency and impact of innovative domain shifts in the pharmaceutical industry. These endeavours have revealed the factors involved in technological innovation based on domain shifts and have yielded conceptual frameworks for modelling such shifts. We propose incorporating quantitative data and engineering knowledge into these frameworks in TOA contexts to offer clear guidance for organisations in systematically identifying potential technology opportunities through domain shifts. Notably, with recent advancements in machine learning, such as generative models, TOA can be enhanced by extracting trends in modern technologies from vast patent documents and generating insights beyond human capabilities, which can potentially be transformed into valuable business knowledge.

This study introduces an analytical framework based on a generative–predictive machine learning model for identifying new technological domains where existing technologies may create greater technological impact. A review of the literature on

TOA reveals three main issues that should be addressed for this challenge. First, suitable technology-related data should be utilised to measure the impact of technologies according to the domains in which they are applied. In this respect, patent classification codes, claims, and forward citations are considered valuable sources of data because (1) patent classification codes provide clues regarding the potential application areas of the technological knowledge encapsulated within a patented invention (Joo & Kim, 2010; Kang, 2012; Los & Verspagen, 2000); (2) patent claims detail the functions of the technology associated with a patented invention, outlining its intended purposes and distinguishing it from other existing technologies (Lee et al., 2013b; Noh et al., 2015; Park et al., 2012); and (3) the forward citations received by patents have been shown to be positively correlated with the future technological impact of the associated technologies (Hong et al., 2022; Lee et al., 2007, 2018; Woo et al., 2019). Second, concerning the development of machine learning models, examining how the impact of technology varies depending on the technological domains in which its functions are applied is crucial, considering the heterogeneity that exists across domains (Kim et al., 2022). While previous studies have assessed the potential impact of existing technologies within a particular technological domain, they have not considered the varying effects of implementing specific technologies in different domains (Hong et al., 2022; Lee & Lee, 2019; Lee et al., 2018; Woo et al., 2019). Therefore, the domain-specific technological impacts of existing technologies should be evaluated, considering their applications in diverse contexts. Finally, the technological components used to represent technological domains, such as patent classification codes, should be flexibly configurable to enable a more fine-grained characterisation of technology opportunities. The appropriate number and level of technological components required for representing technological domains may vary with the scope of the analysis. This necessitates a generative model with a broad representation space, which allows for technological domains to be depicted using an arbitrary number of components in a multi-level hierarchical structure.

Given these considerations, the proposed analytical framework is designed to be executed in four discrete steps. First, patent documents containing details about patent classes, claims, and citation relationships among existing technologies are collected to represent the domains, functions, and future impact of the associated technologies as machine-readable data. Second, an impact-centric technology landscape is constructed using a generative–predictive machine learning model that integrates the variational autoencoder (VAE) and multi-layer perceptron (MLP) architectures. Specifically, the VAE and MLP are jointly trained to reorganise the landscape such that technologies with similar domains and impacts are located in close proximity. Third, a structured exploration of the technology landscape based on a gradient ascent search algorithm reveals new technological domains as opportunities for maximising the impact of the existing technologies. Finally, the reliability and feasibility of the proposed analytical framework are examined by evaluating its generative and predictive performance using quantitative metrics and examining the changes in technological impacts resulting from the domain shifts.

We empirically analysed 133,654 patents granted in the field of artificial intelligence (AI) technology. Subsequently, the proposed analytical framework was employed to identify potential new technological domains where the technologies described in the collected patents are likely to have a greater impact. Upon evaluating these domain-shift opportunities, we found that the patents in the newly identified domains exhibited, on average, triple the technological impact of those that remained within their original domains. Additionally, they contained a higher proportion of breakthrough technologies—those with exceptionally high impact—compared with typical AI patents. These results confirm that the proposed analytical framework can be valuable for firms when formulating strategies to maximise the utilisation of their existing technologies in the era of open innovation.

The remainder of this paper is organised as follows: Related work summarises the related literature. Data and methodology explains the proposed analytical framework, which is then illustrated through an empirical analysis in Empirical analysis and results. Finally, Conclusions concludes the paper, discussing the limitations of our framework and outlining future research directions.

Related work

TOA is a line of research that involves systematic examination and evaluation of technological trends, market needs, and organisational capabilities to discover promising opportunities for innovation (Porter & Detampel, 1995). Over the past few decades, numerous studies have presented systematic approaches for TOA based on quantitative data and scientific methods (Lee, 2021). The key focus of TOA is the incorporation of factors associated with the impact of a technology. In this context, patent documents are informative as they encompass extensive and structured information—patent classification codes, abstracts, claims, descriptions, and prior-art citations—that can signify an invention's potential impact. Among these details, patent classification codes and claims have been considered pivotal in TOA as they clearly represent technological domains and functions, respectively. Patent classification codes are a set of symbols used to organise patent documents into specific technology groups, outlining an invention's scope of application. Based on this notion, previous studies have attempted to examine the relevance of the technological domains of a patent to its future impact. For instance, some studies have empirically shown that a broader set of technological domains is positively related to forward citations and economic value (Los & Verspagen, 2000; Trajtenberg, 1990). Additionally, novel combinations of technological domains, indicated by unique sets of patent classification codes, are frequently associated with high technological impact due to their ability to foster disruptive innovation (Arts & Veugelers, 2015; Kim et al., 2016). Patent claims define the legal boundaries of an invention through a set of numbered statements that specify the essential elements and functions necessary to protect the invention against infringement. The language of patent claims therefore captures the core functional features, capabilities, and novel aspects of an invention, differentiating the patented invention from other existing technologies. Previous research has utilised patent claims to represent the technological functions embedded in a patent, showing that these functions are correlated with the invention's future impact. Specifically, functional language that departs from prior art often signals breakthrough inventions, and patents having claims semantically distinct from existing technologies are more likely to achieve high technological impact (Hong et al., 2022; Lee et al., 2013b). Apart from technological domains and functions, market-related factors, such as commercial applicability, consumer acceptance, and competitive environment, can also highlight new technology opportunities. Since the market value of a technology often reflects its impact, these factors can play a critical role in TOA. However, compared with technology-related factors, which are well-documented in patent literature, market-related factors lack structured information and are often dispersed across various data sources. This hinders TOA research based solely on market data.

Considering the aforementioned challenges, patent analysis is arguably the most scientific TOA technique as it (1) enables the examination of numerous patents accumulated over long periods, providing objective and reliable insights into technology opportunities, and (2) offers an empirical explanation of the multifaceted nature of technological innovation activities (Lee et al., 2011, 2020; Trajtenberg, 1990; Yoon & Kim, 2011). Specifically, patent analysis has benefitted from recent advancements in machine learning techniques that facilitate the efficient extraction of important patterns and insights from extensive patent data across various domains (Lee et al., 2009, 2013a; Son et al., 2012; Wang et al., 2017).

Modern TOA approaches based on patent analysis can be categorised into three main groups: patent mapping, morphological, and patent landscape approaches. Patent mapping involves the construction of a two-dimensional plot in which individual patents are mapped through text mining and dimension reduction (Song et al., 2017). This approach reveals patent vacuums—defined as regions with sparse patent density on the map—as unexplored technological domains. Lee et al. (2009) and Yoon et al. (2002) constructed patent maps by employing principal component analysis and self-organising feature map methods, respectively, to visualise the relationships among patents and identify potential technology opportunities. Son et al. (2012) developed a patent map using generative topographic mapping, a probabilistic method for creating a low-dimensional latent space from multidimensional data, to discover patent vacuums. The morphological approach involves using text-mining techniques to construct a morphological matrix based on patent information, which helps divide a complex technology system into simpler, separate dimensions, each representing various technological characteristics. This approach reveals the configurations occupied by existing technologies and allows technology opportunities to be identified by exploring unoccupied configuration territories. Yoon and Park (2005) developed a keyword-based morphological analysis technique that uses information from patents on thin-film transistor–liquid crystal display technologies. Yoon et al. (2008) conducted a morphological analysis using product manuals and patent documents to identify promising opportunities for developing new products or technologies. Lee et al. (2013a) developed a modified morphological matrix using patents on electronic shopping technology to organise different types of business models at the technological attribute level. The patent landscape approach is based on the notion of recombinant search, conceptualised using technology landscapes that serve as spatial metaphors, where each position indicates a particular configuration of technological components. Lee and Lee (2019) developed a patent landscape by employing a vector space model composed of patent classification codes, wherein each position—defined by a particular combination of patent classes—denotes a new technological idea and/or associated patented inventions. This approach allows promising technology opportunities to be identified by evaluating the novelty and value of ideas derived from the patent landscape using local outlier factors and naïve Bayes algorithms.

Although existing approaches in TOA have proven useful in identifying new technological ideas that have yet to be fully explored, they provide minimal insight into opportunities that can maximise the potential of existing technologies. Specifically, as firms tend to focus more on exploiting their existing technologies in the current era of open innovation, the practical utility of existing TOA approaches remains constrained (Ghemawat, 2003; Massa & Testa, 2008). This necessitates the development of an analytical framework that can identify new technological domains in which existing technologies may have a greater impact than in their current domains. In this regard, two key issues should be addressed, which the previous approaches do not comprehensively consider. First, modelling the variation in the impact of a technology when its functions are applied to different domains is crucial. However, patent mapping and morphological approaches typically focus on identifying new technology opportunities within a single narrow technological domain rather than exploring multiple domains. This restriction stems from the reliance of these techniques on expert knowledge and experience, which hinders the evaluation of a wide array of potential ideas (Lee et al., 2015; Song et al., 2017; Yoon & Park, 2007). Although the patent landscape approach enables extensive searches for technological ideas and rapid assessments of their novelty and value, it does not consider the differing impacts these technologies may have across various domains (Lee & Lee, 2019). Second, to represent technological domains, an expansive representation space that enables precise and detailed depictions of the application of each technology should be established. In patent analysis, technological domains are often represented by specific combinations of patent classes. Accordingly, the representation space for these domains is defined by the number and depth of the patent classes involved. However, the existing TOA approaches utilise high-level patent classification codes and a narrow, fixed array of patent classes owing to constraints on computational resources and the space that can be explored (Lee et al., 2009, 2013a; Son et al., 2012; Wang et al., 2017). To highlight possible avenues for methodological adaptation, we propose an analytical framework based on a generative–predictive machine learning model that can identify new technological domains in which existing technologies may yield greater technological impact. This study distinctly focuses on repurposing and increasing the value of existing, proven technologies by exploring their applications in novel contexts, rather than predicting completely new or emerging technologies. Table 1 summarises the differences between the existing approaches and the proposed approach.

Data and methodology

The proposed analytical framework uses a generative–predictive machine learning model to identify new technological domains in which existing technologies can have greater technological impact. The overall analytical procedure is designed to be executed in four discrete steps, as illustrated in Fig. 1. Specifically, four key methods underpin the framework: (1) a generative machine learning model based on the VAE architecture to represent existing technologies as fixed-size latent vectors; (2) a predictive machine learning model based on the MLP architecture to quantitatively assess the technological impact of existing technologies using forward-citation information; (3) a joint training technique to construct an impact-centric technology landscape, where technologies with similar domains and impacts are closely positioned together; and (4) a gradient ascent search algorithm that explores the technology landscape to identify new technological domains that can maximise the potential technological impact of existing technologies.

Data collection and pre-processing

We adopt the PatentsView database to collect patent documents on existing technologies. PatentsView is an open data platform developed by the United States Patent and Trademark Office (USPTO) to alleviate the complexity and inconvenience associated with using existing patent databases. PatentsView offers detailed information on patents granted since 1976, providing patent classes, claims, and citations as bulk data. Initially, we use the USPTO research datasets to identify the patent numbers of technologies relevant to a particular field of interest. Subsequently, we collect the basic information about these technologies from the PatentsView database, obtaining HTML-formatted

Table 1 Comparison between existing TOA approaches and the proposed approach

Aspect	Patent mapping	Morphological analysis	Patent landscape	Proposed approach
Approach	Hybrid (more qualitative)	Hybrid (more qualitative)	Hybrid (more quantitative)	Fully quantitative
Data sample size and analysis scope	Small amounts of data in a single domain	Small amounts of data in a single domain	Large amounts of data in a single domain	Large amounts of data across multiple domains
Input	Keyword vector	Morphological matrix	Configuration–value matrix constructed from details on patent class and forward citation among patents	Configuration–technological impact matrix constructed from details on patent class, claims, and forward citation among patents
Methods	Dimension reduction	Concept combination	Anomaly detection and predictive model	Generative–predictive machine learning model
Results and implications	Ideas represented by patent vacuums	Ideas identified by exploring unoccupied configuration of technological components	Ideas with high novelty and impact	New technological domains where existing technologies may yield greater technological impact

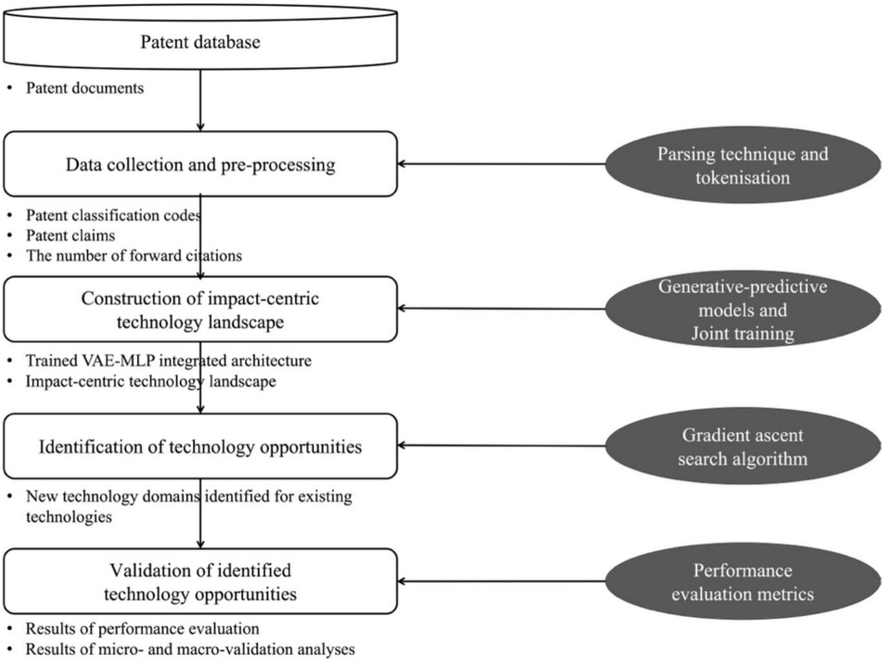

Fig. 1 Overall workflow of proposed analytical framework

patent documents that contain both structured data (e.g. patent number, grant date, and patent classification codes) and unstructured data (e.g. abstracts and claims). The collected patent documents are then parsed and stored according to the specific types of information that they contained.

The proposed analytical framework uses patent classes, patent claims, and forward citations from the collected data to represent the technological domains, functions, and impacts of existing technologies. Patent classes, which are assigned to patents for categorisation, specify the combination of technological components on which the patented inventions are based; they further indicate the potential application areas of the technological knowledge encapsulated within the inventions (Kang, 2012; Los & Verspagen, 2000). We utilise codes under the Cooperative Patent Classification (CPC), a scheme jointly developed by the European Patent Office (EPO) and USPTO, which offers global consistency and is regularly revised to ensure relevance and currentness (U.S. Patent and Trademark Office and European Patent Office, 2017). This allows us to capture the technological domains of existing technologies by assessing recent global technological trends. The CPC currently serves as a global standard as it facilitates patent searches, technology trend analysis, and international cooperation. The structure of the CPC mainly follows that of the International Patent Classification scheme, managed by the World Intellectual Property Organization. Specifically, CPC codes comprise sections, classes, subclasses, main groups, and subgroups based on a hierarchical structure. Considering the breadth of technological knowledge covered by each CPC level, the main group of CPC codes is employed to represent technological domains as it encompasses specific products, processes, and mechanisms related to patented inventions (Park & Yoon, 2017). Patent claims, featured in a patent document to define the technical boundaries of an invention's protection rights, describe the distinct technological functions that patented inventions can provide (Lee et al., 2013b). A patent document contains two types of claims: independent claims, which list the core functionalities of the patented invention, and dependent claims, which cover additional details; hence, the proposed analytical framework mainly utilises independent claims to capture the core technological functions of patented inventions in detail. Considering the positive correlation between the number of forward citations and the impact of patented inventions, as demonstrated by many empirical studies, the proposed analytical framework uses the forward citation counts of patents to evaluate the technological impact of existing technologies (Basberg, 1987; Kim et al., 2011; Lerner, 1994; Trajtenberg, 1990). Considering the temporal dynamics of technology, the framework focuses on the number of forward citations received by a patent within five years of grant approval. This ensures fair comparison of technological impact across patents granted at different times. Specifically, the five-year window is effective for identifying the potential of technology opportunities amid rapidly evolving technological trends.

The collected data are pre-processed to ensure compatibility with our machine learning models. Specifically, since a given patent may belong to one or more patent classes defining its technological domains, we consider the set of CPC codes as a variable-length sequence. These codes are then tokenised into a discrete sequence to be utilised as model input. Similarly, given that patent claims are described in natural language, which comprises sequences of technological terms, the claims are tokenised using a predefined tokeniser developed for natural language processing. Additionally, the text in patent claims is cleaned by removing special symbols, stop words, and meaningless words, before being converted into sequences of discrete tokens. In terms of the technological impact, forward citation counts generally reflect impact differently across fields. For example, patents with more than five forward citations are rare in laser technology but common in biotechnology (Breschi et al., 2006). Furthermore, since breakthrough technologies seldom appear in practice, the distribution of forward citation counts follows a negative exponential pattern: most patents receive few citations, while a small minority receive many. Thus, to evaluate the impact of technology opportunities, distinguishing the typical citation counts that most patents receive from the exceptionally high counts that a few breakthrough patents receive is reasonable. Accordingly, recent studies have adopted quantised forward citation counts as a proxy for a patent's technological impact instead of raw citation counts (Lee & Lee, 2019; Lee et al., 2018; Woo et al., 2019). With reference to existing classification schemes, the proposed framework transforms forward citation counts into binary labels based on the percentile rank of the citation counts (specifically, the top 10%) within the entire set of collected patents. The collected patents are then labelled L1 (breakthrough) if they received more than the top 10% of forward citations or L2 (common) otherwise.

Construction of impact-centric technology landscape

An impact-centric technology landscape is constructed using a generative–predictive machine learning model based on the VAE and MLP architectures, as shown in Fig. 2. A VAE is a generative neural network architecture designed to learn a compressed representation of high-dimensional data in an unsupervised manner and to reconstruct the input data from the learned representation. VAEs comprise two main modules: an encoder and a decoder. The encoder converts the input data into latent vectors, while the decoder reconstructs the input based on these vectors. Through this reconstruction process, VAEs

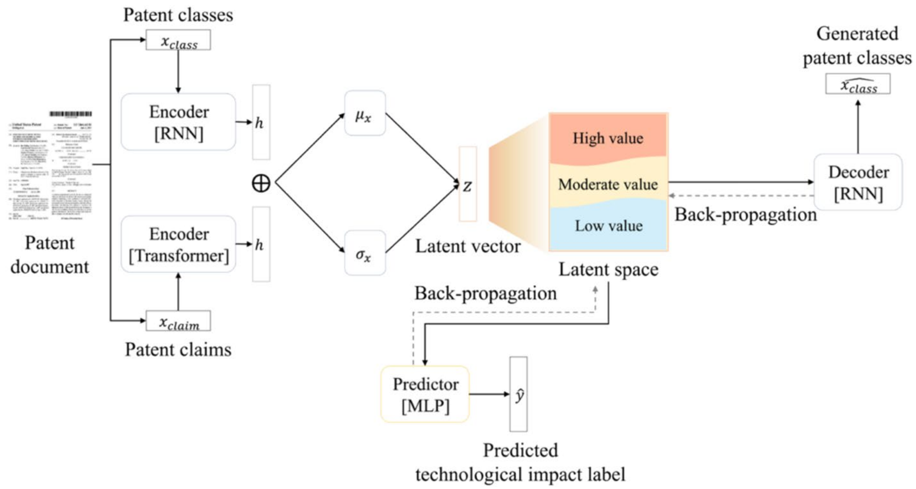

Fig. 2 Generative–predictive machine learning model based on VAE and MLP architectures

leverage latent vectors to capture the underlying features of the input, thereby constructing a latent space. Notably, each latent vector is derived through sampling from a probability distribution associated with its corresponding input, which results in slight variations even for identical inputs due to the probabilistic nature of the encoding process. This allows VAEs to generate novel and diverse outputs, even from unseen inputs.

The generative component of our model is configured to accept patent classes and claims and reconstruct the patent classes using recurrent neural networks (RNNs) and transformers. RNNs consist of interconnected artificial neural networks designed to process sequential data (Salehinejad et al., 2017). In our model, two distinct RNNs are used as the encoder and decoder, which accept sequences of CPC codes as input and generate CPC sequences as output, respectively. Transformers, which function based on self-attention mechanisms, efficiently handle long-term dependencies in sequential data by processing inputs in parallel (Vaswani et al., 2017). The transformer model in the proposed framework serves as another encoder, accepting patent claim text as input. The predictive component of our model employs an MLP to predict the technological impact label, which categorises the number of forward citations received by a patent. An MLP is a type of artificial neural network comprising multiple layers of interconnected neurons, typically used for classification and regression tasks.

Specifically, once a patent is input into the model, its CPC sequence (x_{class}) and claim text (x_{claim}) are encoded by the RNN and transformer, respectively, and then merged into a unified feature vector. This feature vector is subsequently transformed into mean (μ_x) and variance (σ_x) vectors that represent the probability distributions of latent variables for the input data. Thereafter, a continuous latent vector z is sampled from the multivariate probability distribution defined by μ_x and σ_x and then positioned within the latent space. Subsequently, the latent vector z is decoded into a sequence of CPC codes that mirrors the structure of the input data. This decoded sequence represents an artificial set of patent classes corresponding to the relevant technological domain. The stochastic sampling process of the VAE allows it to generate new yet semantically valid sequences of CPC codes that were not included in the training data. This is a crucial feature that addresses the limitations inherent in existing patent mapping or morphological approaches, which are confined to exploring existing technological configurations or rely on expert knowledge. The predictive component of the model then estimates the technological impact of existing technologies using the latent vector associated with each patent as input.

The generative and predictive components of our model are jointly trained using an integrated loss function that combines reconstruction, regularisation, and prediction losses, as detailed in Eq. (1):

$$TotalLoss = CrossEntropy\left(x_{class}, \widehat{x_{class}}\right) + D_{KL}\left(N\left(\mu_x, \sigma_x\right) \parallel N(0,1)\right) + CrossEntropy(y, \hat{y}) \tag{1}$$

where the first term represents the reconstruction loss, calculated as the cross-entropy loss between the true and reconstructed CPC sequences; the second term indicates the regularisation loss computed using the Kullback–Leibler divergence between the prior distribution of the latent vector z (assumed to be a Gaussian distribution) and the probability distribution parameterised by μ_x and σ_x; and the third term denotes the prediction loss, which is the cross-entropy loss between the true and predicted technological impact labels. During training, the model learns to synthesise valid CPC sequences while simultaneously estimating the number of forward citations received by existing technologies. This joint training of generative and predictive components of our model forces the latent space to encode features that are effective not only for reconstructing realistic CPC sequences, but for predicting technological impacts. Consequently, the latent vectors converted from patents that share similar classification information and technological impact are positioned closely together in the latent space. This arrangement forms an impact-centric technology landscape, enabling the potential impact of a technology to be predicted based on its specific technological function within a given domain.

Identification of technology opportunities

This step involves identifying technology opportunities by searching the impact-centric technology landscape for new technological domains where existing technologies may yield greater impact. To achieve this, the landscape is systematically explored using a gradient ascent search algorithm. In vector calculus, a non-zero gradient for a function at a given point indicates the direction of the steepest ascent, with its magnitude representing the rate of increase. Based on this notion, the algorithm iteratively updates the position of an initial point p_n by moving along the gradient, thereby finding a new point p_{n+1} that is guaranteed to lie on a contour corresponding to a higher function value, as shown in Eq. (2):

$$p_{n+1} = p_n + \alpha \cdot \nabla f\left(p_n\right) \tag{2}$$

where f is the objective function to be maximised, and α is a predefined step-size parameter. After each update, the gradient at the new position $\nabla f\left(p_{n+1}\right)$ is recalculated, and the process repeats until a termination condition is satisfied. This condition typically includes (1) achieving a sufficiently high function value; (2) a change in the function value below a predefined tolerance; or (3) reaching a predefined maximum number of iterations. Throughout this process, each position update guarantees an increase in the function value, ensuring monotonic ascent towards a local maximum.

Using the gradient ascent search algorithm, the proposed analytical framework explores the impact-centric technology landscape to identify new technological domains where existing technologies may create greater potential impact, as illustrated in Fig. 3. It is noteworthy that the optimisation process based on gradient ascent search should be performed over a continuous space because of the necessity of calculating derivatives of the objective function to find the gradient. Here, the latent space constructed by the VAE is composed of fixed-size continuous vectors, allowing for gradient ascent-based exploration, whereas the CPC sequences in the input space are discrete and variable in length, preventing their use in such continuous optimisation. The continuous nature of the latent space exploration allows for a shift to new domains with greater impact without deviating significantly from the original domain. This aligns with the fact that technological innovations often arise through the gradual expansion and transition of existing knowledge. Here, the predictive model – trained to estimate technological impact from the latent vectors of patents – serves as the objective function. First, each patent document is input into the generative–predictive model and converted into a latent vector based on its patent classes and claims. The model then outputs the predicted probability of the patent's impact label being L1 (i.e. high technological impact based on the forward citation count). Then, the gradient of this L1 probability with respect to the focal latent vector is calculated via backpropagation. In this step, the position of the focal latent vector is updated to maximise the L1 probability. With the latent vector being iteratively updated in this manner, the exploration process continues until the L1 probability exceeds a desired termination threshold (e.g. 0.8). As a result, we obtain a new CPC sequence that corresponds to a technological domain where the existing patented invention is expected to achieve greater impact. Notably, the position optimised for the focal latent vector might be a local optimum. While the global optimum can offer the highest L1 probability, the technological domain associated with it may vary greatly from the patent's original domain. In such cases, applying the technological functions inherent in the focal patent to the new domain may be unrealistic. Hence, despite the risk of getting stuck in local optima, the gradient ascent search algorithm can be useful for identifying new technology opportunities.

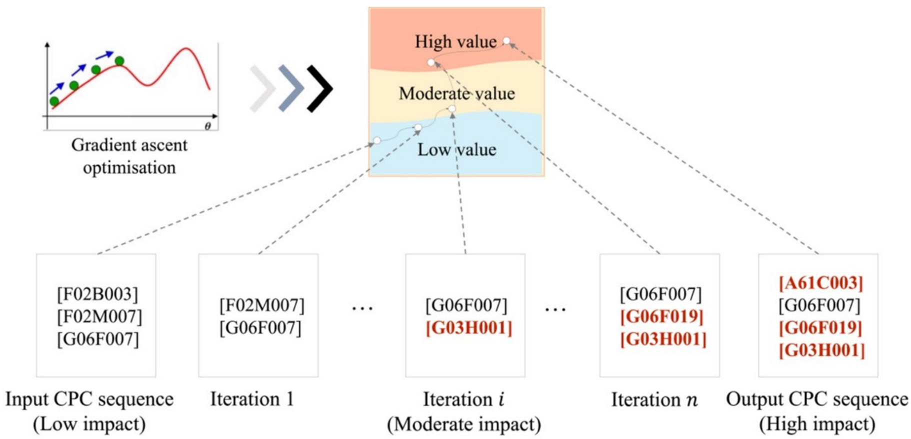

Fig. 3 Example of exploration of impact-centric technology landscape

Validation of identified technology opportunities

The reliability of the proposed analytical framework depends on our model's performance in generating valid CPC sequences and predicting technological impact labels for existing technologies. When a patent is input into the model, the generative performance is appraised based on the similarity between the patent's original CPC sequence and the CPC sequence generated by the model. Considering that all main groups of CPC codes assigned to a patent have equal importance, the technological domain of a patent can be viewed as a set of assigned CPC codes. Accordingly, the generative performance of our model is evaluated based on Jaccard similarity, which measures the number of shared elements between two finite sample sets, as shown in Eq. (3):

$$JaccardSimilarity = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

where A and B are the two sets being compared. This similarity measure ranges from 0 to 1, where 1 indicates that the two sets have the same elements, and 0 indicates no similarity between the two sets. The model's predictive performance is measured by comparing the actual and predicted technological impact labels. Several performance evaluation metrics are utilised for classification tasks: accuracy, precision, recall, and F1-score (Hossion & Sulaiman, 2015). Although accuracy comprehensively measures overall classification performance, it can yield misleading results when the dataset is imbalanced (Kim et al., 2019). Given the highly skewed distribution of forward citations in real-world scenarios, we additionally employ precision, recall, and F1-score to measure the overall effectiveness of our model's predictive component.

The feasibility of the proposed analytical framework lies in its ability to identify new technological domains where existing technologies are likely to achieve greater impact. To evaluate this ability, we compare the predicted technological impact when each patented invention is deployed within its existing domain against that in the new domain identified by the proposed analytical framework. Specifically, we devise two validation schemes at different levels: (1) a micro-validation analysis to determine whether the proposed analytical framework can identify practical and valuable technology opportunities at the individual patent level using patent citation networks, and (2) a macro-validation analysis to gauge how the identified opportunities influence broader technology trends by examining technological impact across a domain-wide group of patents.

A patent citation represents the portion of technological knowledge inherent in the citing patent that stems from the cited patent and, as such, explains the knowledge transfer process (Kim & Magee, 2017). Therefore, the citation links among patent classification codes offer clues regarding the historical development of technology based on domain shifts, revealing how technology opportunities emerged from existing technologies and were subsequently applied (Park & Yoon, 2018). In this regard, the micro-validation analysis is designed to examine the anticipated technological impact of applying existing patented inventions in new domains relative to their impact in their respective original domains. As the impact of a patented invention in a different domain cannot be observed directly, we compare the impacts of subsequent patents citing the focal patent in the same domain and in different domains. Figure 4 illustrates the overall structure of the citation relationships utilised for this analysis, while Table 2 presents the notations used to describe the analysis procedure. Specifically, we denote the set of patent classes assigned to each patent P in the collected data as $Class_{original}$ and the corresponding claims as $Claim_{original}$.

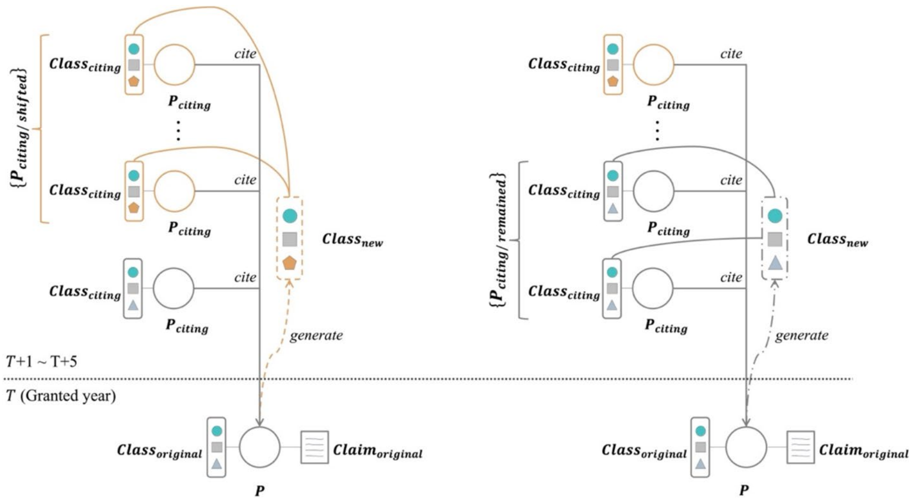

Fig. 4 Overall structure of citation relationships among patents

Table 2 Notations for micro-validation analysis

Term	Definition
P	A patent in the collected data (the focal patent)
$Class_{original}$	The set of patent classes assigned to P
$Claim_{original}$	The patent claim text of P
$Class_{new}$	The new technological domain for P identified using the proposed analytical framework
$\{P_{citing}\}$	The set of patents that cite P
$Class_{citing}$	The set of patent classes assigned to each citing patent in $\{P_{citing}\}$
$\{P_{citing/shifted}\}$	The set of citing patents belonging to $Class_{new}$
$\{P_{citing/remained}\}$	The set of citing patents belonging to $Class_{original}$

The new technological domain for the focal patent identified using the proposed analytical framework is then denoted as $Class_{new}$. Notably, $Class_{new}$ might be identical to $Class_{original}$ (see the right side of Fig. 4), which indicates that a domain shift may not enhance the focal patent's technological impact significantly. This implies that the focal patent has already been applied in an optimal technological domain and has achieved sufficient impact. The set of patents that cite the focal patent is defined as $\{P_{citing}\}$. Among the citing patents, we separate those belonging to the new technological domain and consider them as a reference set of patents indicating real examples of technological domain shifts; the set of those patents is denoted as $\{P_{citing/shifted}\}$. Conversely, the other citing patents, which belong to the same domain as the focal patent, are regarded as examples of within-domain technological improvement; this set of patents is denoted as $\{P_{citing/remained}\}$. Specifically, if the set of patent classes assigned to P_{citing} differs in even one CPC code from $Class_{original}$, it is considered an example of domain shift (i.e. $P_{citing/shifted}$). This strict criterion is established based on the understanding that increases in technological impact from a domain shift can arise not only from radical innovations, which are extremely rare, but also from incremental adaptations. In this context, we consider the main group-level CPC code suitable for monitoring shifts in technological domains, as it provides fine-grained coverage of technology applications. Finally, the practicality of the technology opportunities identified using the proposed analytical framework is assessed by examining the differences in technological impact among the patents in $\{P_{citing/shifted}\}$ and $\{P_{citing/remained}\}$. Hence, based on the patent citation networks, the micro-validation analysis confirms whether existing technologies can achieve greater impact when applied in the new domains identified using the proposed analytical framework.

The macro-validation analysis involves examining the influence of the identified technology opportunities on technology trends at a collective level. To this end, we again measure the general impact of technologies based on the forward citation count of a patent. Specifically, for each patent, we compute the average forward citation counts among patents in the same domain and among those in the new domain identified using the proposed analytical framework. We then determine the macro-level impact based on whether the average forward citation counts in the new domain exceed those in the original domain for each focal patent.

Empirical analysis and results

An empirical analysis of patents associated with AI was conducted to demonstrate the effectiveness of the proposed analytical framework. AI technology has advanced significantly over the past few decades and is now considered a general-purpose technology with broad applicability across various domains, such as manufacturing, healthcare, and robotics. Moreover, AI technologies currently have many cross-domain applications because despite being initially designed for a particular technological domain, they have been repurposed and innovatively utilised in other domains. For instance, deep learning–based object detection techniques, such as R-CNN (Girshick et al., 2014) and YOLO (Redmon et al., 2015), were originally developed to recognise common objects, including pedestrians, cars, and animals, in various types of images; however, they are currently being employed prominently in the healthcare sector for medical image analysis (Li et al., 2019). In this context, AI is at the forefront of patenting activities in most industries and has emerged as one of the most productive technologies, with a notably high rate of patent grants. Accordingly, patents related to AI technology are suitable as subjects of empirical analysis to validate the reliability and practicality of the proposed analytical framework.

Data collection and pre-processing

To construct the patent database, the patent numbers of patents related to AI technology were collected from the Artificial Intelligence Patent Dataset (AIPD), a publicly available dataset covering a broad range of AI patent documents provided by the USPTO (Pairolero et al., 2025). The AIPD contains about 15 million U.S. patent documents (including both patents and pre-grant publications) published from 1976 through 2023 in the field of AI. This dataset was created using machine learning techniques that identify patent documents containing one or more key components of AI technology, such as machine learning, natural language processing, computer vision, speech, knowledge processing, AI hardware, evolutionary computation, and planning and control. From the AIPD, we extracted the patent numbers of patents granted by the USPTO and predicted as having a high probability of being AI patents. Given that AI technology has rapidly developed over the past 20 years since the introduction of the deep belief network in 2006 (Hinton et al., 2006), we focused on technology opportunities in the AI field identified between 2006 and the present. Specifically, patents granted between 2006 and 2020 were identified considering that the proposed analytical framework employs a five-year window to calculate the forward citation counts. We obtained HTML-formatted documents containing the patent class and claim information corresponding to the identified patents from the bulk data provided by PatentsView. After parsing the documents and removing observations with missing values, the constructed patent database contained 133,654 AI-related patents, as shown in. Table 3.

The collected patent data were then pre-processed for application to the generative–predictive machine learning model developed in this study. First, we tokenised the CPC codes by treating each main group (e.g. G06F015 and H04L012) as a separate token. Next, because patent claims are hierarchical and since the first claim captures an invention's core functions (Yang & Soo, 2012), we extracted only the first claim to limit the computational cost. The patent claims were then tokenised using a WordPiece tokeniser, which has been widely used in BERT-based models. Finally, as described previously, we quantified the technological impact of each patented invention as the number of forward citations received within five years of grant approval. Thereafter, we categorised the patents into the breakthrough (L1) and common (L2) groups based on the previously defined percentile criterion (top 10%). In our patent database, the 90th percentile of citation counts was 15; hence, patents with more than 15 citations were labelled L1, whereas those with 15 or fewer were labelled L2.

To ensure the generalisability of our model, we split the constructed patent database into training, validation, and test sets. Following the commonly used dataset split criterion for training machine learning models, we allocated 70% of the data for training, 20% for validation, and 10% for testing through stratified random sampling. The resulting training, validation, and test sets contained 96,230, 24,058, and 13,366 samples, respectively.

Construction of impact-centric technology landscape

An impact-centric technology landscape was constructed by training the generative–predictive machine learning model on the patent database, using PyTorch as the main Python framework. Since our model includes deep and complex structures, various hyperparameters needed to be carefully determined based on the size and nature of the datasets. Pilot experiments were conducted in advance using a random search algorithm to determine the optimal set of hyperparameters. The hyperparameters selected for building and training the model are listed in (Appendix A). Moreover, because the proposed analytical framework is designed to jointly train the generative and predictive models using three disparate loss functions, balancing learning between the generation and prediction tasks is essential. For brevity, the complete details of the model training process are presented in Appendix A. The impact-centric technology landscape was constructed using the joint training of the generative and predictive part of our model, and parts of it are presented in Table 4.

Once a patent is input into the trained model, the CPC sequence is reconstructed, and the technological impact label is predicted for the corresponding patent, as illustrated in Table 5. For instance, the actual CPC sequence of the patent numbered 6,983,071 is (G06V010, G06V030), and its technological impact label is L2 because its forward citation

Table 3 Parts of collected patent database

Patent number	Patent claims	Patent classes (CPC sequence)	List of forward citations	Number of forward citations
6,983,071	1. A character segmentation device for removing…	G06V030, G06V010	–	0
6,983,272	1. a method of generating a result list in response…	G06Q030, G06Q040, G06Q050, …	7,870,158, 8,483,674, 8,484,234, …	38
6,981,417	1. An apparatus for use in acoustic micro-imaging…	G01N029	9,597,059, 9,857,338, 10,739,779, …	11
…	…	…	…	…
9,223,929	1. A method of processing affinity-based array…	G16B005, G16B025, G16B040, …	9,708,647, 10,106,839, 9,499,861, …	7
9,223,967	1. a microprocessor comprising …	G06F021, G06F009, G06F011	9,767,284, 10,324,795, 9,767,271, …	7
9,224,096	1. a method for a device to self – assess…	G06N003, H04W004, G06F021, …	10,710,239, 10,679,493, 10,656,923, …	22

Table 4 Parts of impact-centric technology landscape

Patent		Latent vector						
Patent number	Patent classes (CPC sequence)	v_1	v_2	v_3	…	v_{30}	v_{31}	v_{32}
6,983,071	G06V030, G06V010	− 1.172	0.349	− 0.503	…	1.103	0.473	0.186
6,983,272	G06Q030, G06Q040, G06Q050, …	0.529	− 0.402	0.433	…	− 0.291	0.242	− 1.772
…	…	…	…	…	…	…	…	…
9,223,967	G06F021, G06F009, G06F011	0.670	− 0.946	0.713	…	− 1.880	− 0.465	− 0.456
9,224,096	G06N003, H04W004, G06F021, …	− 1.343	− 5.160	− 4.010	…	1.129	− 9.250	− 0.190

count is 0. Our model successfully reconstructed the CPC sequence for this patent and correctly predicted its technological impact label. This result confirms that our model has adequately learnt the data required to construct an impact-centric technology landscape, in which technologies sharing similar domains and impacts are closely positioned.

Identification of technology opportunities

This step involves identifying new technological domains for existing technologies by exploring the impact-centric technology landscape using a gradient ascent search algorithm. First, the gradient of the L1 probability for the input patent's latent vector is computed. Next, the vector's position is updated in the gradient direction, and this procedure is repeated until the L1 probability exceeds a predetermined termination threshold. We set this threshold at 0.8 to ensure that the new domain would result in a significantly higher impact than the patent's original domain. With the latent vector at the final position, our model generates a CPC sequence that is expected to be the new technological domain where the patented invention may yield a greater technological impact.

Table 6 illustrates how the technology domain shifts during the landscape exploration. Notably, we continued the gradient ascent search beyond the termination threshold to track further changes. For example, Patent 6,983,071 was initially mapped to CPC codes G06V010 and G06V30, which cover 'Arrangements for image or video recognition or understanding' and 'Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition', respectively. It should be noted that, as the exploration progressed, a new CPC code G06F016 appeared, which covers 'Information retrieval; Database structures therefor; File system structures therefor'. This domain shift suggests that image-recognition or video-recognition methods have the potential to be useful in information retrieval. This observation aligns with the increasing demand for multi-modal inputs in search engines, as demonstrated by Google Lens.

The aforementioned findings demonstrate how the estimated L1 probabilities increase during the exploration process. They also highlight that as the potential for technological impact increases, the generated CPC sequences diverge incrementally from the original CPC sequences of the patents. These results confirm that the proposed analytical framework can methodically identify new technology domains that draw on the potential for innovation inherent in existing technologies by navigating the constructed technology

Table 5 Parts of generation and prediction results

Patent			Model outputs	
Patent number	Patent classes (CPC sequence)	True label (Number of forward citations)	Generated patent classes	Predicted label (L1 probability)
6,983,071	G06V010, G06V030	L2 (0)	G06V010, G06V030	L2 (0.3130)
6,983,272	G06Q030, G06Q040, G06Q050, …	L1 (38)	G06Q030, G06Q040, G06F016, …	L1 (0.6433)
…	…	…	…	…
9,223,967	G06F021, G06F009, G06F011	L2 (7)	G06F021, G06F009	L2 (0.4777)
9,224,096	G06N003, H04W004, G06F021, …	L1 (22)	G06K019, H04W004, G06Q010, …	L1 (0.9137)

Table 6 Results of exploring impact-centric technology landscape

Patent		Identified technology opportunities						
Patent number	Patent classes (True label)	Generated CPC sequence (L1 probability)						
		Iteration 1	Iteration 2	Iteration 3	…	Iteration (N – 2)	Iteration (N – 1)	Iteration N
6,983,071	G06V030, G06V010 (L2)	G06V010, G06V030 (0.3130)	G06V010, G06V030 (0.7168)	G06V010, G06V030 (0.8406)	…	G06V010, G06V030, **G06F016** (0.9852)	G06V010, G06V030, **G06F016** (0.9858)	G06V010, G06V030, **G06F016** (0.9863)
6,983,272	G06Q030, G06Q040, G06Q050, … (L1)	G06Q030, G06Q040, G06F016, H04L067, Y10S707 (0.6433)	G06Q030, G06Q040, G06F016, H04L067, Y10S707 (0.7508)	G06Q030, G06Q040, G06F016, H04L067, Y10S707 (0.8136)	…	G06Q030, G06Q040, G06Q050, G06F016 (0.9779)	G06Q030, G06Q050, G06F016 (0.9787)	**G06Q020**, G06Q030, G06Q040, G06Q050, G06F016 (0.9795)
…	…	…	…	…	…	…	…	…
9,223,967	G06F021, G06F009, G06F011 (L2)	G06F021, G06F009 (0.4777)	G06F021, G06F009, **G06F011** (0.5039)	G06F021, G06F009, **G06F011** (0.6887)	…	**G06F016**, G06F021, **G06Q010** (0.9768)	**G06F016**, G06F021, **G06Q010** (0.9821)	**G06F016**, G06F021, **G06Q010** (0.9859)
9,224,096	G06N003, H04W004, G06F021, … (L1)	G06K019, H04W004, G06Q010, G06F021 (0.9137)	G06K019, H04W004, G06Q010, G06F021 (0.9247)	G06K019, H04W004, G06Q010, G06F021, **Y02P090** (0.9671)	…	G06K019, G06Q010, H04W004, G06F021, **G07C005** (0.9872)	G06K019, G06Q010, H04W004, **Y02P090**, G06F021, **G07C005** (0.9872)	G06K019, G06Q010, H04W004, **G06K007**, G06F021, **G07C005** (0.9883)

landscape. Although the exploration results explain the process of identifying potential technology opportunities, further validation is required to quantitatively evaluate the reliability and feasibility of the proposed analytical framework.

Validation of identified technology opportunities

To verify the reliability of the proposed analytical framework, we examined our model's performance on the generation and prediction tasks using the test set shown in Table 7. The generative performance was measured in terms of the Jaccard similarity between the original and generated CPC sequences, as shown inTable 7(a). The results show that our model can accurately reconstruct the actual sets of patent classification codes assigned to existing technologies. Table 7(b) presents the predictive performance of the model based on performance evaluation metrics relevant to classification. The evaluation involved comparing the actual and predicted technological impact labels of the collected patents. The values of the metrics for L1 and L2 were computed based on the estimated classification probability for each label. Despite the overall prediction accuracy being satisfactory, we focused more on other metrics due to the inherent imbalance in the patent database regarding technological impact labels. The precision, recall, and F1-score values indicate that our model successfully classified patents with the L2 label, whereas it performed modestly for those with the L1 label. Specifically, based on the F1-score, our model exhibited relatively poor predictive performance for the L1-labelled patents compared with the L2-labelled patents. This is because L1-labelled patents constitute only a minor proportion (10%) of the entire database. Importantly, for the L1 label, the recall value was much higher than the precision value. This result shows that our model is aggressive in predicting L1 labels, trying to classify high-impact technologies as accurately as possible. In other words, it focuses more on learning the distinct features of breakthrough technologies over common ones, based on their technological domains and the functions mentioned in the patent documents. Therefore, while this model may not be ideal for filtering out less significant technology opportunities, it could still be valuable for guiding the discovery of new domains where existing technologies may achieve greater impact. Based on these results, we conclude that the constructed impact-centric technology landscape sufficiently reflects real-world scenarios as it was appropriately reorganised through the joint training of our model.

Micro- and macro-validation analyses were conducted on the test set to verify the feasibility of the proposed analytical framework. The data used for model training encompassed patents granted between 2006 and 2015, along with their forward citation counts across five years after grant approval. The micro-validation analysis focused on the technological impact of the patents that cited the patents in the test set, i.e. the citing patents, based on their application domains. Therefore, we utilised AI-related patent documents granted between 2016 and 2020 to obtain information about the technological domains and impacts associated with the citing patents. By excluding patents without any forward citations, we identified a total of 6,929 pairs of patents, each consisting of a patent P in the test set and a corresponding set of patents P_{citing} that cited P. Among them, 1,930 citing patents were found to belong to the same domain as P, i.e. $Class_{original}$. We considered these patents as examples of within-domain improvement $\left\{P_{citing/remained}\right\}$. We then identified 61 citing patents belonging to the new technological domain identified by the proposed analytical framework, i.e. $Class_{new}$, and considered them as domain-shift examples $\left\{P_{citing/shifted}\right\}$.

Table 7 Performance evaluation results of generative–predictive model (a) Jaccard similarities of generated patent classes (b) Performance evaluation metrics based on predicted technological impact labels

Patent number	Patent claims (keywords)	Original CPC sequence	Generated CPC sequence	Jaccard similarity
6,981,379	power supply system supplies …	H01M008, Y02E060	H01M008	0.5000
6,983,099	information reproducing apparatus …	G11B027	G11B027	1.0000
…	…	…	…	…
9,225,879	method video sequential alignment …	G06F016, G06V020, G06T007, G11B027, H04N005	G06F016, G06V020, G11B027, H04N005	0.8000
9,225,944	method depicting coverage area …	H04N007, G05B023, H04N005, G05B023, H04N005	H04N007, G08B021, G08B013	0.2000
Overall (averaged)				0.7493

	Support	Accuracy	Precision	Recall	F1-score
L2	12,083	–	0.9445	0.5968	0.7314
L1	1283	–	0.1499	0.6695	0.2449
macro-averaged	13,366	0.6038	0.6038	0.6038	0.6038
micro-averaged	13,366		0.5472	0.6332	0.4882
weighted-averaged	13,366		0.8682	0.6038	0.6847

Table 8 Results of micro-validation analysis

Subset of citing patents	Forward citations					
	Count (exceeding L1 criterion)	Average	Standard deviation	Minimum	Median	Maximum
$\{\boldsymbol{P}_{citing/remained}\}$	1930	5.6662	13.4039	0	2	160.7
$\{\boldsymbol{P}_{citing/shifted}\}$	61 (18, 29.51%)	16.9016	30.7678	0	4	153

Table 8 presents descriptive statistics, including the average, standard deviation, minimum, median, and maximum values, of the forward citation counts computed for each patent in $\{P_{citing/remained}\}$ and $\{P_{citing/shifted}\}$. As observed, the average forward citation count for $\{P_{citing/remained}\}$ was only 5.6662, which is fairly low considering that the threshold to separate breakthrough (L1) and common (L2) technologies was set at 15 citations. Moreover, most of the patents in $\{P_{citing/remained}\}$ have fewer than three forward citations, which aligns with the general trends of patents relevant to AI technology. This result empirically confirms that most technological advances are incremental and that breakthrough inventions are rare. Compared with those in $\{P_{citing/remained}\}$, the patents in $\{P_{citing/shifted}\}$ exhibited a notably higher average forward citation count of 16.9016. Thus, the patents in $\{P_{citing/shifted}\}$ are likely to pertain to more innovative technological domains than the existing patents that they cite. Notably, $\{P_{citing/shifted}\}$ exhibited a relatively high standard deviation of about 30, which indicates that the forward citation counts varied considerably across patents. This observation aligns with the notion that while disruptive innovations in technologies resulting from domain shifts may have significant technological impacts, they often prove unviable (Adner & Levinthal, 2002). Moreover, among the examples of technological domain shifts in $\{P_{citing/shifted}\}$, 18 patents exceeded the L1 criterion, i.e. breakthrough inventions, accounting for approximately 30% of the 61 total patents. Given that the breakthrough patents were considered to be those with the top 10% of forward citation counts within the entire set of collected patents, this result demonstrates that the technology opportunities identified by our model are more likely to drive technological innovation than typically observed.

However, the results of the micro-validation analysis can vary with the termination threshold for the exploration since this threshold determines when to stop the gradient ascent search and obtain the modified CPC sequence that will serve as a new technological domain. To validate the proposed analytical framework more thoroughly, we performed a sensitivity analysis for five termination thresholds ranging from 0.5 to 0.9 with intervals of 0.1. Based on the sensitivity analysis results, we computed descriptive statistics of the forward citation counts for each version of $\{P_{citing/shifted}\}$ derived by exploring the impact-centric technology landscape under each termination threshold, as shown in Table 9. Among the five threshold values, 0.8, the value that we had originally set, yielded much higher average forward citation counts, as well as a significantly higher proportion of L1 patents in $\{P_{citing/shifted}\}$, compared with the other thresholds. Thus, the results of the micro-validation analysis confirm that the proposed analytical framework can identify new technology domains where existing technologies can have greater impact than in their current domains, at the individual patent level.

Table 9 Sensitivity analysis results for termination threshold during exploration

Termination threshold	Forward citations					
	Count (exceeding L1 criterion)	Average	Standard deviation	Minimum	Median	Maximum
0.5	57 (10, 17.54%)	13.1053	41.4577	0	2	276
0.6	57 (9, 15.79%)	7.1228	12.1068	0	3	71
0.7	56 (12, 21.43%)	11.4286	20.0725	0	4.5	102
0.8	61 (18, 29.51%)	16.9016	30.7678	0	4	153
0.9	48 (8, 16.67%)	7.9375	9.2216	0	4.5	40

For a more detailed examination of the citation patterns associated with the domain shifts that occur when the functions of existing technologies are applied in other domains, we investigated the identified technology opportunities in depth. Considering that $\{P_{citing/shifted}\}$ represents the real examples of technological domain shifts, this investigation offered additional clues regarding the practicality of the proposed analytical framework. Specifically, we examined individual patents in $\{P_{citing/shifted}\}$, particularly those with high forward citation counts, to assess whether the identified technology opportunities could represent genuine breakthroughs. Referring to the micro-validation analysis results presented earlier, 18 of the patents within $\{P_{citing/shifted}\}$ exceeded the L1 criterion (i.e. breakthrough examples). Table 10 summarises the original and shifted domains for each patent, alongside the corresponding technological impact.

Based on the obtained results, we first compared the technological domains of each citing patent with those of the focal patent. For instance, Patent 7,697,758 was originally assigned to $Class_{original}$ consisting of two CPC codes: 'G06K009: Methods or arrangements for recognising patterns' and 'G06V030: Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition'. For this patent, our model identified a new domain $Class_{new}$ which added 'G06V010: Arrangements for image or video recognition or understanding', 'G06V020: Scenes; Scene-specific elements', and 'G06F016: Information retrieval; Database structures therefor; File system structures therefor'. Among the patents citing the focal patent, Patent 8,953,886 was assigned to the same CPC codes as $Class_{new}$, demonstrating the existence of the domain shift identified by the proposed analytical framework. For another example, Patent 7,024,364 was originally assigned to $Class_{original}$ that covers two CPC codes: 'G10L015: Speech recognition' and 'H04M003: Automatic or semi-automatic exchanges in telephonic communication'. Our model then suggested replacing H04M003 with 'G06F003: Input arrangements for transferring data to be processed into a form capable of being handled by the computer'. Evidence of this domain shift comes from two patents, Patent 8,589,161 and 9,721,566, that cited Patent 7,024,364. Importantly, this example illustrates a more innovative domain shift from telephonic communication to general computing, spanning broader categories at the class-level CPC codes. In most instances of domain shifts identified by our model, the citing patents had higher forward citation counts than the cited patents. This demonstrates that the proposed analytical framework systematically captures technological domain shifts necessary to maximise technological impacts, whether they are incremental or radical.

Beyond comparing CPC sequences, we examined the differences in the technological knowledge embedded within the existing patented inventions and their citing patents,

Table 10 Breakthrough patents in $\{P_{citing/shifted}\}$

Patent number	$Class_{original}$	Forward citation count	$Class_{new}$	P_{citing}	$Class_{citing}$	Forward citation count
7,242,681	H04L063	10	H04L063, G06F021	9,875,344	H04L063, G06F021	20
8,549,579	H04L051, H04L063	8	H04L063, G06F021	9,087,215	H04L063, G06F021	23
8,307,177	G06F016, G06F011, G06F009	165	G06F016, G06F011, G06F009, G06F003	9,823,977	G06F016, G06F011, G06F009, G06F003	40
8,307,177	G06F016, G06F011, G06F009	165	G06F016, G06F011, G06F009, G06F003	10,474,542	G06F016, G06F011, G06F009, G06F003	20
9,089,966	B25J009	13	B25J009, Y10S901	9,486,921	B25J009, Y10S901	114
8,832,846	G06F008, G06F021	14	H04L063, G06F021	9,536,108	H04L063, G06F021	71
7,697,758	**G06K009, G06V030**	**33**	**G06V010, G06V030, G06F016, G06V020**	**8,953,886**	**G06V010, G06V030, G06F016, G06V020**	**39**
7,024,364	**G10L015, H04M003**	**10**	**G10L015, G06F003**	**8,589,161**	**G10L015, G06F003**	**33**
7,024,364	**G10L015, H04M003**	**10**	**G10L015, G06F003**	**9,721,566**	**G10L015, G06F003**	**102**
7,127,403	G10L015, H04M003	6	G10L015, G06F003	9,721,566	G10L015, G06F003	102
7,127,403	G10L015, H04M003	6	G10L015, G06F003	8,670,985	G10L015, G06F003	22
7,769,756	H04L065, H04N021, H04N019, H04H060	16	H04N021, H04H060, G06Q030	8,966,525	H04N021, H04H060, G06Q030	37
7,363,549	G06F011	28	G06F016, G06F011	7,979,441	G06F016, G06F011	18
7,363,549	G06F011	28	G06F016, G06F011	8,108,429	G06F016, G06F011	40
7,363,549	G06F011	28	G06F016, G06F011	7,904,913	G06F016, G06F011	24
7,363,549	G06F011	28	G06F016, G06F011	7,689,602	G06F016, G06F011	17
7,363,549	G06F011	28	G06F016, G06F011	8,131,723	G06F016, G06F011	29
8,832,594	G06F040, G06F003	161	G06F040, G06F016, G06F003	9,043,696	G06F040, G06F016, G06F003	153

focusing on the description section in the patent document. For example, Patent 7,697,758, titled ‘Shape clustering and cluster-level manual identification in post optical character recognition’, introduces a post-OCR (Optical Character Recognition) processing method that clusters similar character image snippets from an OCR output to detect and correct recognition errors. The patent for this invention was granted in 2010 and was cited in 2013 by Patent 8,953,886, titled ‘Method and system for character recognition’, which describes a more comprehensive character recognition system that extends beyond individual character shapes. In this example, the citing patent repurposed the clustering-based OCR post-processing concept from the cited patent within a larger, more versatile system that covers end-to-end document handling, from image capture and segmentation to information retrieval. Specifically, the application area of the cited patent (7,697,758) was relatively narrow as it was especially useful in digitising printed texts (e.g. scanned books or forms). However, the application scope of the citing patent (8,953,886) spans multiple domains, from page segmentation to information retrieval, as the invention not only captures text from a physical document but also searches for data related to the recognised text. In this process, the technological function of OCR post-processing specified by the cited patent (7,697,758) is utilised, enabling users to find the electronic version or metadata of a physical document. This example illustrates a clear technological transformation into modern multi-modal search systems, such as conversational AI services that process and understand images and text together. The example also showcases the practical realisation of a technology opportunity identified by the proposed analytical framework. Notably, the citing patent received more forward citations (39) than the focal patented invention (33), which highlights that the citing patent successfully realised the technology opportunity created by the innovative domain shift. The results of this in-depth investigation demonstrate the viability of the technology opportunities identified by the proposed analytical framework through real-world examples of technological domain shifts. Consequently, our framework is expected to serve as a key tool for identifying new technological domains where existing technologies can produce greater technological impact.

For the macro-validation analysis, we derived patent indicators for AI-related patents. For each patent in the compiled dataset, we compared the average forward citation count of the other patents in its original domain with that of the patents in the new domain identified using the proposed analytical framework. Table 11 presents the results of this comparison.

Table 11 Results of macro-validation analysis

Patent number	CPC sequence		Average forward citation count
8,386,498	Original patented invention	H04L063, H04L041	18.1667
	Identified technology opportunity	G06F016, H04L063, G06F021	9.0000
7,647,471	Original patented invention	G06F012	4.8472
	Identified technology opportunity	G06F016, G06F021	14.0000
…	…		
7,496,557	Original patented invention	Y10S707, G06F016	9.9010
	Identified technology opportunity	G06F016	10.1000
Overall (averaged)	Original patented inventions		10.3353
	Identified technology opportunities		13.4855

Overall, the patents in the new domains exhibited a greater technological impact than those in the original domains. This result demonstrates that the new technological domains identified using the proposed framework have, on average, a more positive influence on the overall technology trends than the domains where the existing patented inventions have already been applied.

Although the foregoing empirical analysis of AI technology underscores the reliability and feasibility of the proposed analytical framework, it covered only one technology field. Therefore, we assessed the applicability of the proposed analytical framework further through a robustness test considering a different technology field, i.e. semiconductors, as detailed in Appendix B.

Conclusions

This study introduces an analytical framework that uses a generative–predictive machine learning model based on the VAE and MLP architectures to identify new technological domains in which existing technologies may yield greater technological impact. Within the proposed framework, the generative and predictive components of the model are jointly trained to construct an impact-centric technology landscape where existing technologies with similar technological impacts are closely positioned. A gradient ascent search algorithm is used to explore this technology landscape and identify promising new technological domains for existing technologies. The empirical analysis conducted on 133,654 AI-related patents demonstrates the reliability and feasibility of the proposed framework for building a realistic technology landscape and identifying suitable new technology domains for existing technologies.

This study makes two major contributions to the existing literature. First, from a theoretical perspective, it extends previous research on TOA by introducing a fully operationalised framework that can support decision-making in identifying potential opportunities for technological domain shifts using quantitative data and scientific methods. This enables a systematic identification of opportunities arising from shifting existing technologies towards areas with high technological potential. Moreover, to the best of our knowledge, this study represents the first attempt to utilise a deep generative model for TOA research. The generative capability of the VAE enables the identification of realistic technology opportunities in a considerably large search space. Second, from a practical perspective, the proposed analytical framework presents more viable technology opportunities to enhance the utilisation of existing technologies by firms with limited resources and capabilities for TOA-based R&D. This method contrasts with previous approaches to TOA, which focused on discovering unexplored technological ideas. Furthermore, it offers an automated software system that allows even individuals without expertise in or relevant experience with complex machine learning models to identify potential technology opportunities. Therefore, the proposed framework can serve as a valuable decision-support system for firms to formulate strategies for maximising the utilisation of their existing technologies.

However, although this study has valuable implications, the proposed framework remains in the exploratory phase and requires further elaboration in terms of research scope, methodology, applicability, utility, and practicality. First, our framework is mainly designed to identify domain-shift opportunities for existing technologies, not to discover entirely nascent or emerging technologies that may lack direct existing technological precursors for repurposing. This poses a distinct and complementary challenge for future

TOA studies. Second, the framework employs main group-level CPC codes to represent the technological domains of existing patented inventions. In some cases, this level of the CPC code only roughly describes the domains of existing technologies, which necessitates a more detailed representation. In this regard, employing a lower level of patent classification codes (e.g. subgroup-level CPC codes) and more patent claims could enhance the expressive capacity for technology opportunities, albeit at the expense of requiring more complicated generative models (e.g. GPT) and far more data. Furthermore, the proposed generative-predictive model does not incorporate an explicit mechanism to reflect the structural and hierarchical aspects of CPC codes in the process of generating patent classes. This could result in identifying technology opportunities that are not likely to be viable when the technological functions of existing technologies cannot be applied in real-world settings. Future research may address this issue by adopting CPC code description texts as supplementary material to represent technological domains, which convey accurate structural meanings of patent classes. Third, although the proposed framework shows promising results in identifying technological domains where existing technologies can have greater impact, it still offers room for improvement, particularly in terms of predictive performance. This is largely because the patent database is highly imbalanced with respect to the number of forward citations. Nonetheless, the proposed framework serves as a valuable screening tool for potential application domains for existing technologies. Fourth, promising technology opportunities often emerge when existing technologies are applied to entirely different domains (Schoenmakers & Duysters, 2010). Therefore, the usefulness of the proposed analytical framework can be verified more thoroughly by performing validation analyses across multiple domains composed of diverse technology fields, rather than within a single domain. However, handling the extensive data involved would require more complex model architectures and stronger computational capabilities. This complexity can be mitigated in future research by employing advanced generative and predictive models coupled with high-performance computing devices. Finally, while the practicality of the proposed framework was verified through quantitative validations using patents granted over the past decade, whether technology opportunities identified from present-day patent data can be considered viable by human decision-makers remains to be confirmed. Future research could focus on a qualitative validation for practitioners to determine whether the identified technology opportunities have tangible significance and appear promising for future applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-025-05491-z.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00219212); the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5A2A03015926); and a Korea University Grant (K2403001).

Code availability All relevant Python code used in this study is publicly available at https://github.com/glee2/TOA-using-generative-predictive-models.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews, 8*, 21–47. https://doi.org/10.1111/j.1468-2370.2006.00119.x

Adner, R., & Levinthal, D. A. (2002). The emergence of emerging technologies. *California Management Review, 45*, 50–66.

Andriani, P., & Carignani, G. (2014). Modular exaptation: A missing link in the synthesis of artificial form. *Research Policy, 43*, 1608–1620. https://doi.org/10.1016/j.respol.2014.04.009

Andriani, P., Ali, A., & Mastrogiorgio, M. (2017). Measuring exaptation and its impact on innovation, search, and problem solving. *Organization Science, 28*, 320–338. https://doi.org/10.1287/orsc.2017.1116

Ardito, L., Coccia, M., & Petruzzelli, A. M. (2021). Technological exaptation and crisis management: Evidence from COVID-19 outbreaks. *R & D ManagemeNt, 51*, 381–392. https://doi.org/10.1111/radm.12455

Arts, S., & Veugelers, R. (2015). Technology familiarity, recombinant novelty, and breakthrough invention. *Industrial and Corporate Change, 24*, 1215–1246.

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of the literature. *Research Policy, 16*, 131–141. https://doi.org/10.1016/0048-7333(87)90027-8

Breschi, S., Tarasconi, G., Catalini, C., Novella, L., Guatta, P., & Johnson, H. (2006). Highly cited patents, highly cited publications, and research networks. European Commission.

Dew, N. (2009). Serendipity in entrepreneurship. *Organization Studies, 30*, 735–753. https://doi.org/10.1177/0170840609104815

Dew, N., Sarasvathy, S. D., & Venkataraman, S. (2004). The economic implications of exaptation. *Journal of Evolutionary Economics, 14*, 69–84. https://doi.org/10.1007/s00191-003-0180-x

Garud, R., Gehman, J., & Giuliani, A. P. (2016). Technological exaptation: A narrative approach. *Industrial and Corporate Change, 25*, 149–166. https://doi.org/10.1093/icc/dtv050

Ghemawat, P. (2003). The forgotten strategy. *Harvard Business Review, 81*, 76–84.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/cvpr.2014.81.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*, 1527–1554.

Hong, S., Kim, J., Woo, H. G., Kim, Y. C., & Lee, C. (2022). Screening ideas in the early stages of technology development: A word2vec and convolutional neural network approach. *Technovation, 112*, 102407. https://doi.org/10.1016/j.technovation.2021.102407

Hossion, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining and Knowledge Management Process, 5*, 1–11. https://doi.org/10.5121/ijdkp.2015.5201

Hur, W. (2017). The patterns of knowledge spillovers across technology sectors evidenced in patent citation networks. *Scientometrics, 111*, 595–619. https://doi.org/10.1007/s11192-017-2329-7

Jeong, G. H., & Kim, S. H. (1997). A qualitative cross-impact approach to find the key technology. *Technological Forecasting and Social Change, 55*, 203–214. https://doi.org/10.1016/s0040-1625(96)00209-0

Joo, S. H., & Kim, Y. (2010). Measuring relatedness between technological fields. *Scientometrics, 83*, 435–454. https://doi.org/10.1007/s11192-009-0108-9

Kang, H. Y. (2012). Science inside law: The making of a new patent class in the international patent classification. *Science in Context, 25*, 551–594. https://doi.org/10.1017/s0269889712000233

Kim, J., & Magee, C. L. (2017). Dynamic patterns of knowledge flows across technological domains: Empirical results and link prediction. *arXiv*. https://doi.org/10.48550/arXiv.1706.07140

Kim, C., Lee, H., Seol, H., & Lee, C. (2011). Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. *Expert Systems with Applications, 38*, 12559–12564. https://doi.org/10.1016/j.eswa.2011.04.042

Kim, D., Cerigo, D. B., Jeong, H., & Youn, H. (2016). Technological novelty profile and invention's future impact. *EPJ Data Science, 5*, 8.

Kim, J., Kim, S., & Lee, C. (2019). Anticipating technological convergence: Link prediction using Wikipedia hyperlinks. *Technovation, 79*, 25–34. https://doi.org/10.1016/j.technovation.2018.06.008

Kim, J., Hong, S., Kang, Y., & Lee, C. (2022). Domain-specific valuation of university technologies using bibliometrics, Jonckheere-Terpstra tests, and data envelopment analysis. *Technovation*. https://doi.org/10.1016/j.technovation.2022.102664

Lee, C. (2021). A review of data analytics in technological forecasting. *Technological Forecasting and Social Change, 166*, 120646. https://doi.org/10.1016/j.techfore.2021.120646

Lee, C., & Lee, G. (2019). Technology opportunity analysis based on recombinant search: Patent landscape analysis for idea generation. *Scientometrics, 121*, 603–632. https://doi.org/10.1007/s11192-019-03224-7

Lee, Y.-G., Lee, J.-D., Song, Y.-I., & Lee, S.-J. (2007). An in-depth empirical analysis of patent citation counts using zero-inflated count data model: The case of KIST. *Scientometrics, 70*, 27–39. https://doi.org/10.1007/s11192-007-0102-z

Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation, 29*, 481–497. https://doi.org/10.1016/j.technovation.2008.10.006

Lee, C., Jeon, J., & Park, Y. (2011). Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach. *Technological Forecasting and Social Change, 78*, 690–702. https://doi.org/10.1016/j.techfore.2010.11.010

Lee, C., Park, H., & Park, Y. (2013a). Keeping abreast of technology-driven business model evolution: A dynamic patent analysis approach. *Technology Analysis and Strategic Management, 25*, 487–505. https://doi.org/10.1080/09537325.2013.785513

Lee, C., Song, B., & Park, Y. (2013b). How to assess patent infringement risks: A semantic patent claim analysis using dependency relationships. *Technology Analysis and Strategic Management, 25*, 23–38. https://doi.org/10.1080/09537325.2012.748893

Lee, Y., Kim, S. Y., Song, I., Park, Y., & Shin, J. (2014). Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. *Scientometrics, 100*, 227–244. https://doi.org/10.1007/s11192-013-1216-0

Lee, C., Kang, B., & Shin, J. (2015). Novelty-focused patent mapping for technology opportunity analysis. *Technological Forecasting and Social Change, 90*, 355–365. https://doi.org/10.1016/j.techfore.2014.05.010

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change, 127*, 291–303. https://doi.org/10.1016/j.techfore.2017.10.002

Lee, C., Jeon, D., Ahn, J., & Kwon, O. (2020). Navigating a product landscape for technology opportunity analysis: A word2vec approach using an integrated patent-product database. *Technovation, 96–97*, 102140. https://doi.org/10.1016/j.technovation.2020.102140

Lee, G., Lee, S., & Lee, C. (2023). Inventor–licensee matchmaking for university technology licensing: A fastText approach. *Technovation, 125*, 102765. https://doi.org/10.1016/j.technovation.2023.102765

Lerner, J. (1994). The importance of patent scope: An empirical analysis. *The Rand Journal of Economics, 25*, 319. https://doi.org/10.2307/2555833

Li, Z., Dong, M., Wen, S., Hu, X., Zhou, P., & Zeng, Z. (2019). CLU-CNNs: Object detection for medical images. *Neurocomputing, 350*, 53–59. https://doi.org/10.1016/j.neucom.2019.04.028

Liu, Z., Feng, J., & Uden, L. (2023). Technology opportunity analysis using hierarchical semantic networks and dual link prediction. *Technovation, 128*, 102872. https://doi.org/10.1016/j.technovation.2023.102872

Los, B., & Verspagen, B. (2000). R&d spillovers and productivity: Evidence from U.S. manufacturing microdata. *Empirical Economics, 25*, 127–148. https://doi.org/10.1007/s001810050007

Massa, S., & Testa, S. (2008). Innovation and SMEs: Misaligned perspectives and goals among entrepreneurs, academics, and policy makers. *Technovation, 28*, 393–407. https://doi.org/10.1016/j.technovation.2008.01.002

Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems with Applications, 42*, 4348–4360. https://doi.org/10.1016/j.eswa.2015.01.050

Pairolero, N. A., Giczy, A. V., Torres, G., Islam Erana, T., Finlayson, M. A., & Toole, A. A. (2025). The artificial intelligence patent dataset (AIPD) 2023 update. *The Journal of Technology Transfer*.

Park, Y., & Yoon, J. (2017). Application technology opportunity discovery from technology portfolios: Use of patent classification and collaborative filtering. *Technological Forecasting and Social Change, 118*, 170–183. https://doi.org/10.1016/j.techfore.2017.02.018

Park, I., & Yoon, B. (2018). Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *Journal of Informetrics, 12*, 1199–1222. https://doi.org/10.1016/j.joi.2018.09.007

Park, H., Yoon, J., & Kim, K. (2012). Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics, 90*, 515–529. https://doi.org/10.1007/s11192-011-0522-7

U.S. Patent and Trademark Office and European Patent Office (2017). *Guide to the CPC*. Retrieved June 11, 2025, from https://link.epo.org/cpc/Guide+to+the+CPC.pdf

Porter, A. L., & Detampel, M. J. (1995). Technology opportunities analysis. *Technological Forecasting and Social Change, 49*(3), 237–255. https://doi.org/10.1016/0040-1625(95)00022-3
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You only look once: Unified, real-time object detection. *arXiv*. https://doi.org/10.48550/arxiv.1506.02640
Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv*. https://doi.org/10.48550/arxiv.1801.01078
Savino, T., Petruzzelli, A. M., & Albino, V. (2017). Search and recombination process to innovate: A review of the empirical evidence and a research agenda. *International Journal of Management Reviews, 19*, 54–75. https://doi.org/10.1111/ijmr.12081
Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy, 39*, 1051–1059. https://doi.org/10.1016/j.respol.2010.05.013
Son, C., Suh, Y., Jeon, J., & Park, Y. (2012). Development of a GTM-based patent map for identifying patent vacuums. *Expert Systems with Applications, 39*, 2489–2500. https://doi.org/10.1016/j.eswa.2011.08.101
Song, K., Kim, K., & Lee, S. (2017). Discovering new technology opportunities based on patents: Text-mining and F-term analysis. *Technovation, 60*, 1–14. https://doi.org/10.1016/j.technovation.2017.03.001
Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics, 21*, 172. https://doi.org/10.2307/2555502
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*.
Wang, X., Ma, P., Huang, Y., Guo, J., Zhu, D., Porter, A. L., & Wang, Z. (2017). Combining SAO semantic analysis and morphology analysis to identify technology opportunities. *Scientometrics, 111*, 3–24. https://doi.org/10.1007/s11192-017-2260-y
Woo, H.-G., Yeom, J., & Lee, C. (2019). Screening early stage ideas in technology development processes: A text mining and k-nearest neighbours approach using patent information. *Technology Analysis and Strategic Management, 31*, 532–545. https://doi.org/10.1080/09537325.2018.1523386
Yang, S.-Y., & Soo, V.-W. (2012). Extract conceptual graphs from plain texts in patent claims. *Engineering Applications of Artificial Intelligence, 25*, 874–887. https://doi.org/10.1016/j.engappai.2011.11.006
Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&d planning using SAO-based semantic patent networks. *Scientometrics, 88*, 213–228. https://doi.org/10.1007/s11192-011-0383-0
Yoon, J., & Kim, K. (2012). Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics, 90*, 445–461. https://doi.org/10.1007/s11192-011-0543-2
Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change, 72*, 145–160. https://doi.org/10.1016/j.techfore.2004.08.011
Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management, 54*, 588–599. https://doi.org/10.1109/tem.2007.900796
Yoon, B., Yoon, C., & Park, Y. (2002). On the development and application of a self–organizing feature map–based patent map. *R & D ManagemeNt, 32*, 291–300. https://doi.org/10.1111/1467-9310.00261
Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: Application of text mining. *R & D ManagemeNt, 38*, 51–68. https://doi.org/10.1111/j.1467-9310.2007.00493.x