Quantifying the progress of artificial intelligence subdomains using the patent citation network

Reza Rezazadegan[1,2] · Mahdi Sharifzadeh[1] · Christopher L. Magee[3]

Abstract

Even though Artificial Intelligence (AI) has been having a transformative effect on human life, there is currently no precise quantitative method for measuring and comparing the performance of different AI methods. Technology Improvement Rate (TIR) is a measure that describes a technology's rate of performance improvement, and is represented in a generalization of Moore's Law. Estimating TIR is important for R&D purposes to forecast which competing technologies have a higher chance of success in the future. The present contribution estimates the TIR for different subdomains of applied and industrial AI by quantifying each subdomain's centrality in the global flow of technology, as modeled by the Patent Citation Network and shown in previous work. The estimated TIR enables us to quantify and compare the performance improvement of different AI methods. We also discuss the influencing factors behind slower or faster improvement rates. Our results highlight the importance of Rule-based Machine Learning (not to be confused with Rule-based Systems), Multi-task Learning, Meta-Learning, and Knowledge Representation in the future advancement of AI and particularly in Deep Learning.

Keywords Artificial intelligence · Technological forecasting · Moore's law · Technology improvement rate · Complex networks · Centrality · Deep learning

Mathematics Subject Classification 05C82

JEL Classification E27 D85

✉ Mahdi Sharifzadeh
m.sharifzadeh@sharif.edu

1 Sharif Energy, Water and Environment Institute (SEWEI), Sharif University of Technology, Tehran, Iran

2 Department of Mathematics, College of Sciences, Shiraz University, Shiraz, Iran

3 Institute for Data, Systems, and Society, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Introduction

Artificial Intelligence is a disruptive technology that has had a transformative effect on human life (Baruffaldi et al., 2020; WIPO, 2019), to the extent that it has been considered comparableto electricity. AI is the most popular subfield of computer science (Taheri & Aliakbary, 2022) and is expected by some to surpass human intelligence in the near future. Despite these expectations and AI's current success, we still have not achieved General AI, i.e., artificial intelligence capable of solving a variety of different problems. The question of which AI method has the potential for achieving General AI is an old and controversial one. Symbolic AI, Machine Learning, and Neural Networks, as well as their subfields, each have their own achievements and adherents. Symbolic AI posits that intelligence can be reduced to the manipulation of symbols, and has Logic Programming as its main subdomain. Machine Learning is the study of algorithms that can automatically learn from data. Neural Networks and Deep Learning, on the other hand, try to mimic the working of neurons in the brain and reduce a given problem in a hierarchical manner.

Given the achievements and the flurry of activities in Deep Learning in recent years, many experts regard it as the most promising of AI methods (WIPO, 2019), to the extent that it is sometimes equated with AI in general discussions. However, the question remains as how to scientifically determine which AI subdomains have more potential for future advancement. In technological forecasting, such a problem is addressed by modeling the progress of technologies, with Moore's Law being one of the most empirically-supported ones:

$$q_t = q_0 e^{k(t-t_0)}. \tag{1}$$

here, q_t represents the technology's performance at time t and k is a constant called Technological Improvement Rate (TIR). The importance of such a model lies in the fact that even if an emerging technology has unsatisfactory performance, by estimating k we can forecast its future performance. This enables us to avoid being deceived by the current observed state of affairs of competing technologies. The present contribution estimates the TIR for different subdomains of AI, and thus, fills the long-lasting void for a scientific quantitative comparison of advancement rates of AI methods.

Note that the performance measure q_t, depends on the underlying technology. For example, for wireless technology, it is given by the amount of data transmitted per second. However, there is currently no all-around performance measure for AI methods, even though the development of such a methodology is underway by the US National Institute of Standards (NIST) (AIME Planning Team, 2021). Given the multifaceted nature of AI, a good AI performance measure should take the following into account (AIME Planning Team, 2021).

- Accuracy: This means how close to empirical reality a classification or forecasting method is in classifying or forecasting.
- Speed: Speed is important for real-time applications of AI. The increase in the adoption of the Internet of Things and the Metaverse have further highlighted the need for fast AI.
- Interpretability (or 'explainability'): For real-world applications, it is crucial to be able to understand how an AI method arrives at a conclusion (Barredo Arrieta et al., 2020).
- Ease of implementation: For industrial applications, it is important that an AI method is not difficult to implement and use (Khayyam et al., 2020).

- Ability to handle noise: Real-world data often comes with noise and uncertainties. Therefore, AI methods which can handle noise (such as probabilistic methods) are superior to those which cannot.
- Energy consumption: Unlike the human brain, AI methods such as large Deep Learning models require huge amounts of energy to train and leave a large carbon footprint (Strubell et al., 2019).
- Bias: AI methods can be biased towards specific demographic groups due to bias in the training data or the algorithms themselves. Such bias can cause problems in the social or medical applications of AI (Rejmaniak, 2021).

Many of these criteria are difficult to quantify. For example, even for accuracy, there are different measures such as accuracy, precision, recall, Receiver Operating Characteristic (ROC) curve, etc. Moreover, the results often depend on the benchmarking datasets used. There are currently some benchmarks for AI performance, such as the MLPerf developed by ML Commons (Mattson et al., 2020). MLPerf is, however, concerned solely with speed-related measures such as training time, latency, and throughput and not with other important aspects mentioned above. It also pertains only to some AI problems such as object detection, machine translation, and recommender systems and is focused on Deep Learning.

There are methods for assessing the performance of software, which can potentially be used for AI as well. For example (Leiserson et al., 2020) studied the effect of software optimization, algorithmic improvement, and hardware specialization on computing performance. Again, the research in (Leiserson et al., 2020) is concerned almost exclusively with speed, which is only one of the many factors that determine an AI method's performance. Other studies focus on computational complexity (Grace, 2013).

There have also been many scientometric studies of AI that use the (yearly) number of patents or publications in AI subdomains as a proxy measure for performance, see e.g. (Abadi & Pecht, 2020; Liu et al., 2021a, 2021b; Pandey et al., 2021; Tseng & Ting, 2013). Such studies, including the ones by the World Intellectual Property Organization (WIPO) (WIPO, 2019) and the Organization for Economic Cooperation and Development (OECD) (Baruffaldi et al., 2020), actually measure the amount of *activity* in the field which does not necessarily translate into progress and *improvement,* see Fig. 1 comparing wind turbines and MRI. This is because such studies implicitly assume that all patents are equal in their contribution to the field. The main difference between the current study and former scientomertic studies of AI, is our consideration of the network effects in patent analysis. Perhaps, the closest work to ours is the work of Jiang et al. (2022) which uses patent citation network analysis to quantify the international diffusion of AI inventions by country.

To circumvent the lack of a suitable performance measure for AI, we use an indirect method to estimate the TIR for AI subdomains. This method, first developed by Triulzi et al. in Singh et al., (2021) and Triulzi et al., (2020), is based on the empirically supported idea that technological domains that have a more central position in the global flow of technology, have a higher TIR (Benson & Magee, 2015a). Such positioning is measured by the network-theoretic notion of *centrality* applied to the domain's patents in the *Patent Citation Network* (PCN). The PCN is a directed network that has more than 6 million nodes representing US utility patents and approximately 85 million directed edges which represent citations between patents, see Fig. 2. We restrict ourselves to US patents only, because usually important innovations are eventually patented in the US as well. In (Lee et al., 2022) the notion of centrality was used to find the hubs in the
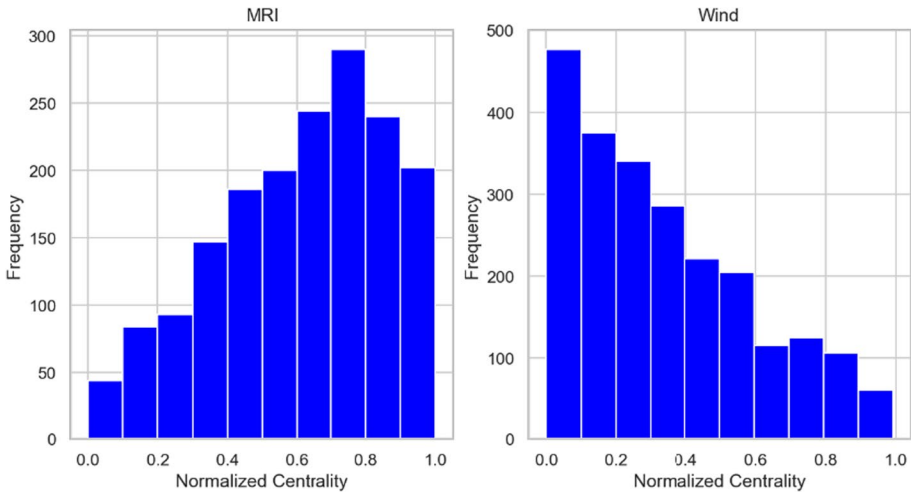
Fig. 1 Patenting activity does not necessarily translate into an improvement rate. Even though there are fewer MRI patents than wind energy patents (1769 and 2493 respectively), most of the patents belonging to the latter have a rather low centrality (importance). This translates into a much lower improvement rate (TIR) for wind technology, see Fig. 7

co-classification network of patents. This co-classification network considers the patents belonging to each pair of International Patent Classification classes and is different from the PCN we study here.

Search Path Node Pair (SPNP) centrality, is a measure of the importance of nodes in a network that was originally developed to uncover a "main path" of publications that led to the discovery of DNA (Hummon & Dereian, 1989). It is defined as the number of paths passing through a given node and here, as in (Triulzi et al., 2020), we apply it to the patents in the PCN. SPNP centrality is much more refined than other measures of importance, such as the number of citations or even PageRank. While PageRank is only concerned with how other patents (or more generally, nodes) cite a given one, SPNP centrality also measures the usage that the focal patent makes of other inventions and hence, provides a better quantification of technological flow.

By using centrality, one can quantify the contribution of different patents to a field and avoid the aforementioned problem with bibliometric studies. Moreover, as in (Triulzi et al., 2020) we normalize centrality to cancel out the effect of citation patterns which are caused by human practices and are not intrinsic to patents themselves. Normalization also evens out the effect of time so that newer patents, which have not had enough time to acquire citations, are only compared to each other, see "Estimation of Technological Improvement Rate for AI subdomains" section.

In Triulzi et al. (2020), it is shown that if μ_i, σ_i are the mean and standard deviation of normalized patent centrality for a technological domain, its rate of improvement K_i can be estimated by the following formula.

$$estimated\, K_i = e^{6.22 \cdot \mu_i - 4.97} \times e^{\sigma_i^2/2} \tag{2}$$

This formula is obtained by applying regression to the empirical performance data for 30 different technologies on one hand, and the mean centrality of patents belonging to
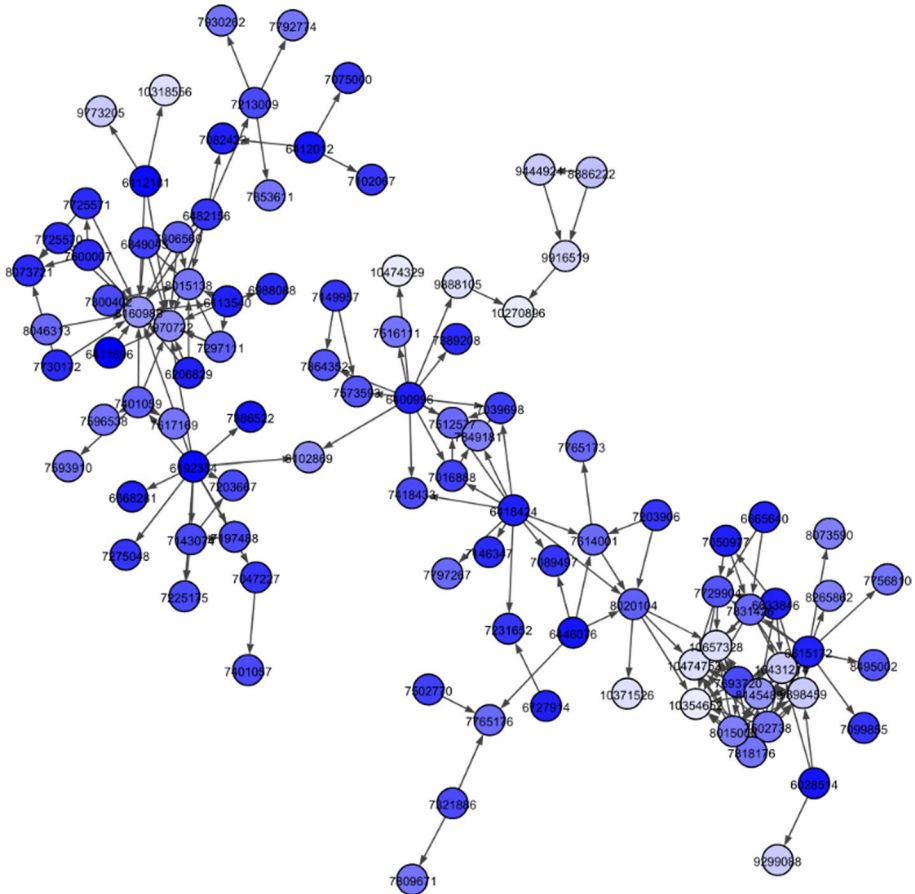
Fig. 2 A very small portion of the Patent Citation Network. Only AI patents of the top 5% centrality (among all patents) are depicted. Older patents are drawn with lighter colors. Citations between patents are represented by arrows. We observe that some patents are more central than others and their removal can severely affect network's connectivity

them on the other hand (Triulzi et al., 2020). In Triulzi et al., (2020), a detailed sensitivity analysis was performed including an analysis of predictor stability over time and a Monte Carlo cross-validation, ensuring the robustness of normalized centrality as a TIR predictor and its superiority to other predictors such as the number of citations. See "Model Validation and Evaluation" section for details on the model validation.

To summarize, we apply the method of (Singh et al., 2021; Triulzi et al., 2020) (further described in Methodology) to AI patents, to estimate TIR for AI subdomains. Regression model selection and evaluation for this work is largely influenced by the work of Triulzi et al. (2020) which is based on the former work of Benson and Magee (2015a) and Alstott et al. (2016). The unique contribution of our research lies in its focus on AI patents. We've applied the methodology of Triulzi et al., (2020) to these AI patents and contextualized the results within the broader AI model evaluation landscape.

Note that unlike (Triulzi et al., 2020) that uses only patent filed up to 2015, we consider patents granted up to 2020 and thus we had to compute patent centralities anew.

Even though scientific publications form a citation network as well, e.g. given by the Microsoft Academic Graph, we chose not to use them for the following reasons. The citation process for patents is a legal process, meaning that, by law, the filing party has to cite prior art. For research publications, however, citations are less rigorous and there are problems such as self-citations, and citations that do not imply usage (Meyer, 2000).

Methodology

Computations

As mentioned in the Introduction, we use the general method of Triulzi et al. (2020) for TIR estimation, which we further explain here. Estimating the TIR of a technological domain uses a regression model (Eq. 2) which is trained on the empirical TIRs of 30 different technological domains (Triulzi et al., 2020). The independent variable for the regression model is the average SPNP centrality for the patents in those domains. SPNP centrality is computed in the patent Citation Network (PCN) whose nodes are the patents and whose directed edges are given by patent citations.

In Triulzi et al., (2020), only patent data up to 2015 were used. Since AI is a fast-evolving technology, we used patent data up to 2020 and therefore, performed the centrality computations of Triulzi et al., (2020) anew for the updated patent data.

The *SPNP centrality* of a node in a directed acyclic graph (DAG), such as the PCN, is defined to be the number of all directed paths passing through the node. It equals the number of paths incoming to the node, multiplied by the number of paths outgoing from it. DAGs admit a so-called topological filtration in which nodes in the same layer have no links (citations). Using this filtration, one can compute the SPNP centralities of all the nodes in a network as large as the PCN in an efficient way using a Dynamic Programming algorithm (Batagelj, 2003). This means that the number of incoming paths is computed in an inductive way starting from the first layer and the number of outgoing paths is computed starting from the patents in the last layer.

Note, however, that the quantity used in the work of Triulzi et al. (2020) to estimate TIR is actually the *normalized centrality* of patents. The normalization process is done to offset the effect of biases in the patent citation process such as the following (Triulzi et al., 2020).

- Patents tend to cite patents in their own technological domain more often.
- Patents in some technological domains give more citations than patents in other domains.
- The age difference between the citing patent and the cited ones follows a specific frequency distribution.

The normalization process removes the effect of these biases and makes it possible to compare the centralities of patents in different domains, and granted in different years in a meaningful way. The process is done in two steps (Triulzi et al., 2020). In the first step, a frequency distribution is obtained of the centralities of patents in randomized PCNs. A randomized PCN has the same nodes as the original PCN but its edges (citations) are shuffled

in such a way that the number of citations between patents in each (patent classification class, year) pair is preserved.

As in the work of Triulzi et al. (2020), 1000 such randomized PCNs are constructed and for each patent, the z-score of raw centralities is computed in the distribution of centralities obtained from these 1000 randomizations. More precisely, the z-score is given by

$$z(v) = \frac{c(v) - \mu_{rand}(v)}{\sigma_{rand}(v)}$$

where $c(v)$ is the (raw) centrality of the node v, and $\mu_{rand}(v), \sigma_{rand}(v)$ are the mean and standard deviation of the distribution of its centralities in the 1000 randomized networks. A positive z-score indicates that the node's centrality is higher than expected in a random network.

The second step of the normalization process removes the effect of patent age on the number of citations given to patents i.e. that newer patents may receive fewer citations than mature ones. In this step, one takes the rank percentile of the aforementioned z-score among patents granted in the same year. This rank percentile is, for each patent, a number between 0 and 1 and is the sought-after normalized centrality.

Note that we still have to be careful with patents that have been filed in recent years, because they have not had enough time to acquire citations. For this reason, instead of using the normalized centrality of the focal patent, the average normalized centrality of the patents it cites is used as the final centrality measure (Triulzi et al., 2020). This is because the future centrality of a new patent can be estimated by the average centrality of the patents it cites (Triulzi et al., 2020). This quantity has the advantage that it does not change in time. In a more recent work (Singh et al., 2021), however, the actual normalized centrality of patents is used, and only patents older than three years are considered. This latter method is more suitable for older technological domains whose main activity does not belong to the last few years (Singh et al., 2021). We, therefore, use the first method i.e. the average centrality of the cited patents, because otherwise, we had to consider only patents filed in 2017 or earlier. The two methods have very close capabilities in predicting TIR, see (Singh et al., 2021).

We used the Python code provided by Alstott (https://github.com/jeffalstott/patent_centralities) and Triulzi (https://github.com/GiorgioTriulzi/TechnologyPerformanceImprovementEstimates) which are freely available on Github. We slightly modified the code to account for obsolete functions, unavailable libraries, or (small) missing steps.

Data collection: finding the right AI patents

We used the Patents View databases for patent information and citations which, at the time of our access, included the data for US patents up to September 2020. This includes over 6 million patents and 85 million citations.

There are a few widely adopted classification schemes for patents such as the Cooperative Patent Classification (CPC) or the International Patent Classification (IPC). However, these classifications, on their own, are not enough for finding AI patents. As detailed elsewhere (Baruffaldi et al., 2020; WIPO, 2019), finding AI-related patents, without including irrelevant ones is not an easy task, even for experts skilled in AI (Giczy et al., 2022). (See also (Liu et al., 2021a) for the case of publications.) This is because AI patents may be classified in non-AI classes and many AI subdomains do not have their corresponding CPC

class. Note that different studies on AI and its impact, use different methods and criteria for finding AI patents (Giczy et al., 2022).

In the WIPO report on AI (WIPO, 2019), a scheme for finding AI patents was developed which consisted of 3 different blocks. The first block consisted of patents in AI-related CPC classes. The second block was comprised of patents containing specific AI-related keywords. The third block consisted of patents satisfying both a CPC class and a keyword condition.

We initially used this approach; however, we realized that a lot of high-centrality patents found this way, were often not related to their corresponding subdomains. For this reason, we used the more stringent search condition used in the OECD report on AI (Baruffaldi et al., 2020). This method uses a set of CPC classes, a set of IPC classes, and a set of keywords. AI patents are taken to be the ones that satisfy one of the following conditions:

1. Belong to one of the specified AI-related IPC classes with no keyword condition.
2. Belong to one of the specified CPC classes related to AI, and contain one of the specified keywords.
3. Contain 3 of the specified keywords without a CPC class condition.

The IPC and CPC classes together with the keywords used in Baruffaldi et al., (2020) were chosen by a group of experts in Artificial Intelligence and patenting and they are reproduced in the Supplementary Tables S-I to S-III. In the OECD report (Baruffaldi et al., 2020) the keywords in the second clause are required to be either in the claims or the description of the patents. We, however, found many patents which had mentioned AI subdomains in their Background section without actually using them i.e. without mentioning the subdomain in the rest of the patent document. For this reason, we restricted our keyword search to the abstract and claims sections of patents only. This way we found 19,978 AI patents.

Note that the OECD report (Baruffaldi et al., 2020) does not consider dividing AI patents into subdomains as we do here. To perform this subdivision, we first needed a taxonomy of AI subdomains. We used a taxonomy similar to that of the WIPO report (WIPO, 2019), which is laid out in the Results section. We then map the keywords from (Baruffaldi et al., 2020) to their corresponding AI subdomain. Supplementary Table S3 lists the keywords belonging to each subdomain.

Since the start of this project, USPTO has published a dataset of AI patents (Giczy et al., 2022), based on the earlier work of Abood and Feltenberger (Abood & Feltenberger, 2018). Their methodology started with 959 patents and patent applications chosen by patent experts, as belonging to AI. A binary classifier was then trained on this set (on both patent text and patent forward and backward citations). The predictions of the model were examined by patent experts experienced in AI, and were also compared to the WIPO method mentioned above. However, the size of the examined sample was very tiny compared to the size of the whole dataset of patents used. Both the seed set and the model predictions for patents from 1976 to 2020 have been published. We compared the result of our method with that of Giczy et al. (2022) and found that more than 87% of our AI patents are classified as AI patents by their method too. The difference can be attributed to the fact that the method of Giczy et al. (2022) involves selecting patents manually for eight different AI technologies (knowledge processing, speech, AI hardware, evolutionary computation, natural language processing, machine learning,

computer vision and planning/control) and then training a classifier, while ours uses a whole-encompassing set of keywords and CPC classes as mentioned above.

Model validation and evaluation

Here we, for the sake of completeness, recall the model evaluation and selection methods used by Triulzi et al., (2020), based on the earlier work of Benson and Magee (Benson & Magee, 2015a). They used four criteria for choosing a robust predictor of TIR: (1) a high correlation with TIR, (2) the correlation must be independent of the time period used in evaluation, (3) correlation must not be restricted to any set of technologies used and 4) the regression coefficient and intercept must be independent of the time period and technological domains whose patents are used in training the model. They considered six different predictors including the number of citations received within 3 years of the patent publication, the average age of the patents the given patent cites and the mean centrality of the patents cited by the given patent, which was discussed above.

A Monte Carlo cross-validation was performed to examine the predictors. For each predictor and each given year (from 1975 to 2015), an ordinary least squares regression model with the sole predictor as independent variable and the logarithm of TIR as the target variable was trained. These models were trained using only patent data up to the given year, for a randomly selected training set of 15 out of the 30 domains. Domains with less than 100 patents (filed up to the year) were excluded. The trained models were then tested on the remaining 15 technological domains.

The quantities used for evaluating the predictors were the Pearson correlation between the predicted and the empirical TIR for the domains in the testing set (of 15 domains), the regression coefficient and intercept for the predictor. These quantities were recorded for each year for 100 different random samples of 15 out of the 30 technological domains.

The most accurate and reliable predictors of TIRs, were taken to be the ones with the highest and most stable correlation and the smallest variation in regression coefficient and intercept over time. It was concluded by (Triulzi et al., 2020) that the normalized centrality of patents has the highest and most stable correlation with TIR. It achieves a high predicting power with $R^2 = 0.63$ and strong stability of the estimated coefficients over time.

Triulzi et al. then study multiple-variable robust linear regression models (Yu & Yao, 2017) with the best predictors to see whether the goodness of fit of the prediction is improved by including more than one variable in the regression. In this case, all patent data for the 30 domains was used to train the regression model. They consider six different combinations of the predictors and observed that combining any of the predictors with normalized centrality did not significantly increase the predictive power of the model, as indicated by its R^2 and also the regression coefficient of the added predictors. Tests for homoscedasticity are also reported in the Supplementary Materials to (Triulzi et al., 2020).

Results

We use the taxonomy of AI subdomains used in the WIPO report (WIPO, 2019) with small modifications. This means that we subdivide AI subdomains into two categories:

- AI technical subdomains: The theoretical foundations of AI, with Machine Learning being the predominant subdomain in terms of the number of patents.
- AI functional application subdomains: The applied AI subdomains that are based on the AI techniques. Computer vision is the dominant one, mentioned in 49% of all AI patents (WIPO, 2019).

Estimation of technological improvement rate for AI subdomains

The estimated improvement rates (computed as described in Methodology), as well as the number of patents related to each AI technical subdomain, are reported in Fig. 3. We note from comparing the two parts of Fig. 3 that the TIR of a domain is not correlated with the number of its patents, something that has been discovered independently, for all technological domains, in (Singh et al., 2021). A concise description of each subdomain is given here and the reasons for high or low TIR are given in the Discussion.

AI technical subdomains

The highest estimated TIRs for technical subdomains belong to Rule-based Machine Learning (RBML) and Multi-task Learning. *Rule-based Machine Learning (RBML) (TIR = 0.92)*. It is a form of Machine Learning and should not be confused with Rule-based Systems, which are a type of expert system. As opposed to other AI methods which find a single model (hypothesis) that governs the whole given dataset, RBML finds local rules governing the training dataset (Lanzi & Stolzmann, 2000). This enables it to handle complex nonlinear datasets.

Multi-task Learning (MTL) (TIR = 0.86) involves solving several ML problems at once (Zhang & Yang, 2021). *Cognitive Modeling/Computing (TIR = 0.73)* aims at automating cognitive tasks and producing algorithms and software which can learn without reprogramming (Hurwitz et al., 2012). A cognitive system typically emulates the large-scale working of the mammalian brain. *Logic Programming (TIR = 0.73)* is one of the oldest AI methods and it aims at producing intelligent machines by encoding human logic into computers. *Meta-Learning (TIR = 0.69 = 1)* refers to methods that can automatically design or fine-tune learning algorithms (Huisman et al., 2021; Vanschoren, 2018). Meta-Learning enables the use of the experience gained in one AI problem in other problems and therefore improves performance. It is believed to be essential for achieving General AI. *Semi-supervised Learning (TIR = 0.65)* enables one to make use of unlabeled data to improve the performance of Supervised Learning methods such as classification (van Engelen & Hoos, 2020). Its importance stems from the fact that labeling data is expensive. It can also be used in conjunction with various Supervised Learning methods.

Among technical AI subdomains, average and low TIRs belong to Graphical Models, Decision Trees, Transfer Learning, Bayesian Learning, Reinforcement Learning, Fuzzy Logic, High dimensional Data, Feature Learning, Clustering, Neural Networks and Deep Learning, Instance-based Learning, Ensemble Learning, Latent Representations, Bio-inspired Methods, Sparse Representations, and Neuromorphic Computing. The discussion of these subdomains is presented in the Supplementary Text, except for Neural Networks and Deep Learning which are discussed in the Discussion section.
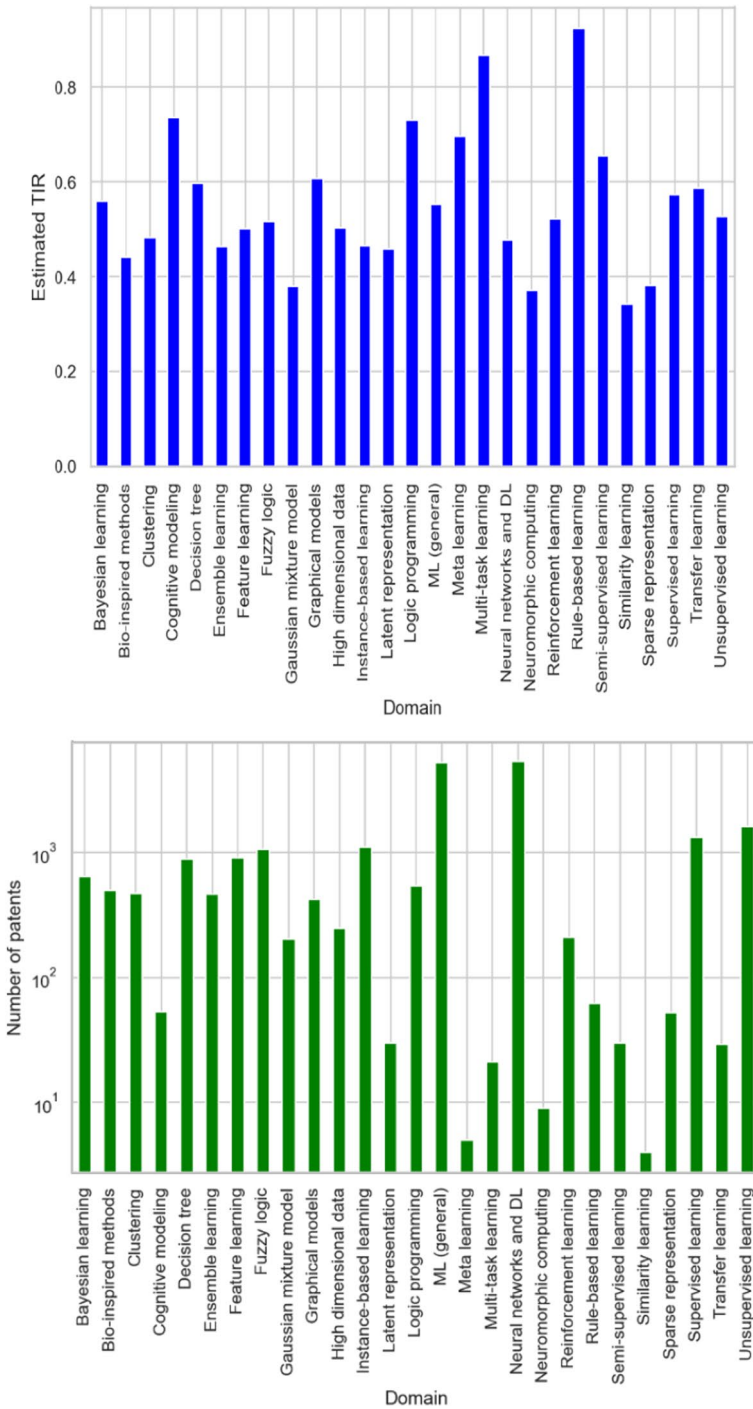
Fig. 3 The estimated improvement rate (TIR) for AI Technical Subdomains (top figure), and the number of patents found in each (bottom figure). The number of patents found in a domain reflects the activity in the subdomain and does not necessarily correlate with its improvement rate
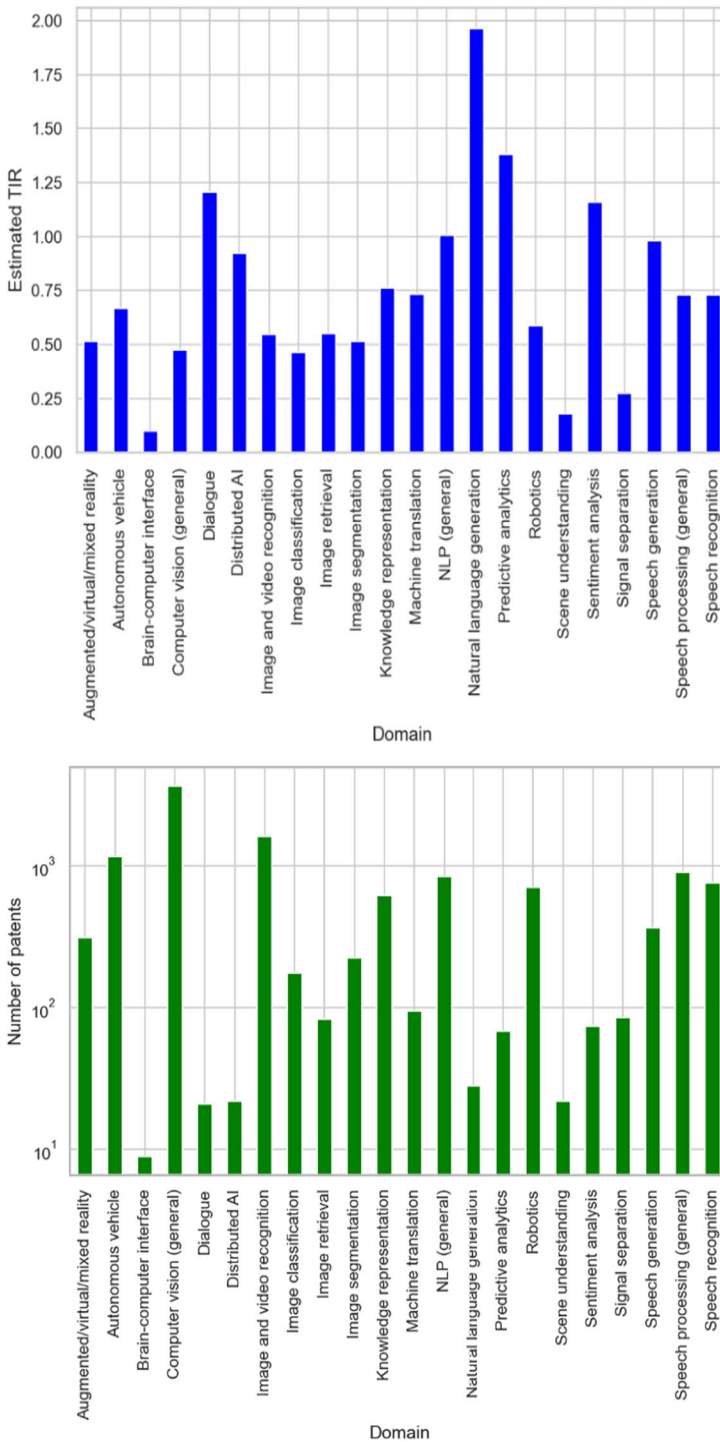
Fig. 4 The estimated TIR (top) and the number of patents (bottom) in AI Functional Application subdomains

Functional application subdomains

These subdomains generally have higher TIRs compared to technical domains presumably because they use the best foundations and they are more applied, see Fig. 4. This figure again shows the lack of correlation between improvement rate and patent activity.

Among functional application subdomains, the highest TIRs belong to *Predictive Analytics* and *Natural Language Processing (NLP)* subdomains, i.e., Natural Language Generation (TIR = 1.96), Dialogue (TIR = 1.2), and Sentiment Analysis (TIR = 1.15). *Predictive Analytics* (TIR = 1.38) is involved in virtually any prediction problem in science, engineering, finance, business, and medicine. Distributed AI (TIR = 0.92) and Knowledge Representation (TIR = 0.76) have low or average TIRs among AI functional application subdomains, even though their TIRs are quite high. The importance of *Distributed AI* stems from the decentralization trends such as Decentralized Finance, Spatial Web, Edge Computing, and Blockchain technology which are expected to transform various aspects of human life. Distributed AI is even referred to as *AI 2.0* by some and is marked by changing goals of Artificial Intelligence from making computers mimic human intelligence to building complex intelligent systems (Pan, 2016). *Knowledge Representation* is very important in AI as it provides a way to encode human knowledge about the world onto computers, which is in turn needed for computers to be able to solve problems related to the outside world (Davis et al., 1993).

Speech Processing (TIR = 0.72) subdomains together with *Autonomous Vehicles* (TIR = 0.66 = − 0.2), *Robotics* (TIR = 0.59) and *Computer Vision* subdomains (TIR = 0.47) have below-average TIR. Signal Separation (TIR = 0.27), and Brain-Computer Interface (TIR = 0.1) have a low TIR. *Signal Separation* aims at separating simultaneous signals without prior knowledge about the signals themselves. *Brain-Computer Interface* makes use of different technological domains i.e. electronics, medicine, neurology as well as ML.

Overlap analysis for AI patents

We analyze the overlaps between patent sets of different AI subdomains. Such analysis shows which patents are shared between different AI subdomains and thus gives a strong indication of important combinations of two or more AI methods. The overlaps are plotted in Supplementary Figs. S1 to S3. Precisely, the numbers in the off-diagonal cells of the heat maps are given by the fraction of the patents in the focal domain that are shared with the other domain (Sharifzadeh et al., 2019). The numbers in the diagonal cells are the fractions of patents in the focal domain that are not shared with other domains.

There are some obvious overlaps; for example, in Fig. S1, Machine Learning (ML) has a lot of overlap with other subdomains. Neural networks and Deep Learning (DL) also overlap with most other technical subdomains except for Cognitive modeling, Sparse Representation, and Rule-based Learning. Unsupervised learning has a significant overlap with High-Dimensional Data and Clustering as dimensionality reduction and clustering are important methods in Unsupervised Learning. The largest overlap of Supervised Learning is with Decision Trees. Other trivial overlaps include Decision Tree and Ensemble Learning, and Bayesian Learning and Graphical Models.

Other than such trivial overlaps, interesting overlaps include Bio-inspired methods and Meta-Learning, Decision Tree and Rule-based Learning, Feature Learning and Neuromorphic Computing, Instance-based Learning, Decision Trees/Ensemble Learning, Sparse

Representations, and Neuromorphic Computing. Unsupervised learning has much more overlap with Meta-Learning, compared to Supervised Learning.

The most significant overlaps between Neural Networks and DL are related to Multi-task Learning, Transfer Learning, Neuromorphic Computing, and Reinforcement Learning (RL). This is to be expected as Transfer Learning is frequently used in large-scale Neural Network models. Neuromorphic computing also mimics the work of biological neurons, both on software (through Spiking Neural Networks) and hardware level (through memristors and neuromorphic devices), more than Neural Networks do. Finally, it is through the use of Deep Learning in RL that the latter has been able to achieve significant milestones such as beating humans in the games of Go and Atari.

We observe that dark diagonal cells in Fig. S1, signifying subdomains that have small overlaps with other subdomains, are rare except for Bio-inspired Methods, Cognitive Modeling, Fuzzy Logic, Neural Networks and DL, and Rule-based Learning. Thus, except for these sub-domains, AI's overall foundations are highly integrated.

For Functional Application subdomains (Fig. S2), most overlaps are with related sub-fields except for Speech Recognition's overlap with NLP subdomains, such as Machine Translation and Natural Language Generation. There is also a small but nontrivial overlap between Computer Vision (general) and NLP on one hand and Speech Processing (general) on the other.

We also consider overlaps between technical and functional application subdomains, see Fig. S3. As can be seen from the Figure's legend, these overlaps are much smaller than the ones in the last two plots. Natural Language Processing (NLP) has significant overlaps with Cognitive Modeling and Latent Representations. Latent Representations are used in converting words and texts to vectors (Liu et al., 2020).

Computer Vision overlaps with most technical subdomains, highlighting its use of various AI techniques. It overlaps most significantly with Cognitive Modeling, Feature Learning, Multi-task Learning, Neuromorphic Computing, Similarity Learning, and Sparse Representations. It however has insignificant overlaps with Meta-Learning, Rule-based Learning, Latent Representations, Bio-inspired Methods, Fuzzy Logic, and Logic Programming.

Robotics does not have much of an overlap with technical subdomains except for Meta-Learning. Sentiment Analysis overlaps with Cognitive Modeling. Speech Processing (general) and Speech Recognition overlap with Gaussian Mixture Models, Multi-task Learning and Bayesian Learning.

Knowledge flow between different AI subdomains

To understand the knowledge flow between different areas of AI, we computed the normalized number of citations between every two subdomains. The normalization process is done in a similar way as for normalizing centrality by taking the z-score of the observed number of citations in the distribution of such citations in randomized networks. The results are depicted in supplementary Figs. S4–S6.

Nontrivial citation flows include Bio-inspired Methods and Feature Learning citing Fuzzy logic and Neural Networks; Fuzzy logic citing Bio-inspired methods, Neural Networks, and Logic programming; Logic programming citing Fuzzy logic, Bayesian Learning, and Neural networks. Neural networks and DL cite many other technical domains such as Bio-inspired methods, Fuzzy logic, Feature Learning, Logic programming,

Instance-based Learning, and Reinforcement Learning. On the contrary, we observe that Reinforcement Learning, Transfer Learning, and Sparse Representations do not cite other technical subdomains much indicating low knowledge flow. Latent Representations is the only technical subdomain cited by Meta-Learning.

Among Functional Application subdomains (Fig. S5), NLP cites Speech Processing subdomains; Augmented/Virtual/Mixed Reality cites Computer Vision and Image and Video Recognition (with a small z-score). Autonomous Vehicle cites the same two subdomains together with Robotics. There is a small but nontrivial citation z-score from both Computer Vision and NLP to Speech Processing. Note that Speech Processing cites NLP and Computer Vision as well (with a higher z-score). Robotics mainly cites Computer Vision Subdomains.

Brain-computer Interface and Distributed AI do not significantly cite Technical subdomains or other Functional Application subdomains. Dialogue and Signal Separation do not cite others except Sentiment Analysis.

We also studied citations from Functional Application subdomains to the Technical ones (Fig. S6). Machine Learning (general) and Neural Networks receive a lot of citations from Functional Application subdomains. Other than these two subdomains, Computer Vision (general) uses Feature Learning and Instance-based Learning the most. Distributed AI uses Logic Programming, and Robotics uses Bio-inspired Methods and Reinforcement Learning. Finally, Speech Processing uses Bayesian Learning, Clustering, and Fuzzy Logic the most. Speech Processing (general) uses Unsupervised Learning more than Supervised Learning, something which is reversed for Image Classification and Robotics.

Discussion

The assumption behind our results is that domains with more central patents have higher improvement rates. This assumption has been validated for a set of 30 technological domains by (Triulzi et al., 2020), see "Model Validation and Evaluation" section. The absence of an agreed-upon performance measure for different AI methods (as discussed in the Introduction), makes validating our results tricky. For this reason, we rely on NIST performance criteria (AIME Planning Team, 2021), mentioned in the introduction, and the following general rules for validating our results. Note that, as mentioned in the Introduction, other similar studies mostly rely on the count of patents in each AI subdomain, and thus their results are not in the same vein as ours.

To understand the reasons for low or high TIRs, note that there are a few principles that govern a technology's growth rate such as the following:

- Principle 1: Technologies that involve interacting components (e.g. combustion engines or robotics) have slower improvement rates than technologies with less internal interaction (Basnet & Magee, 2017). See Figs. 1 and 7.
- Principle 2: A technology's TIR is largely determined by how well it can make use of developments in other technological fields rather than the efforts within the technology (such as expenditure by governments) (Basnet & Magee, 2017).

Perhaps the most striking result of this research is that Neural Networks and Deep Learning have a below-average TIR. Figure 5 suggests that several AI technical subdomains are likely to improve faster than Deep Learning (DL) and could be even more significant in

the future. Figure 3 compares the TIRs and number of patents of technical (or theoretical domains including Rule-based Machine Learning with that of Deep Learning. Figure 6 demonstrates an overall higher number of patents for the latter and a higher fraction of high-centrality patents for the former. First of all, note that this result pertains to Deep Learning as a theoretical AI subdomain, and its applications in NLP and Computer Vision are assessed in the context of the latter subdomains, among the functional application subdomains of AI, in Fig. 4. There are many reasons for believing that the current progress rate of Deep Learning may not continue into the future (Marcus, 2018; Rudin & Carlson, 2019). For example, the computational power required for Deep Learning has been increasing at a rate far exceeding the rate of increase in computing power (OpenAI, 2022; Thompson et al., 2020). Said differently, increased parallel processing power, whose use enabled in Deep Learning through the transformer architecture (Vaswani et al., 2017), has been one of the main drivers of DL advancement in last few years and may not continue at current levels (Leiserson et al., 2020; Lohn & Musser, 2022; OpenAI, 2022). Deep Learning consumes inordinate amounts of energy resulting in carbon emissions equivalent to that of 5 cars in their lifetime, for training a large model (Strubell et al., 2019). The cost of training a large DL model such as Open AI's GPT-3 is estimated to be around 4.6 million dollars (Li, 2020). It has an edge over other AI methods only in problem categories that involve raw numerical data such as Computer Vision (Rudin & Carlson, 2019).
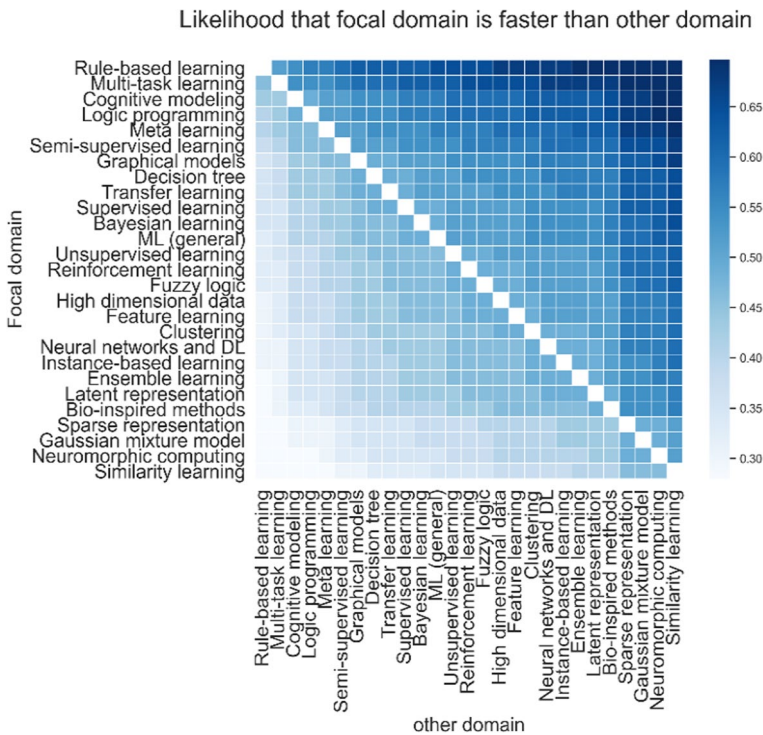

Fig. 5 The likelihood that an AI technical subdomain is improving faster than other subdomains. These likelihoods are computed as in (Triulzi et al., 2020), by assuming a normal probability distribution for the TIR of a domain. A similar heat map for AI functional application subdomains is provided in Fig. S7
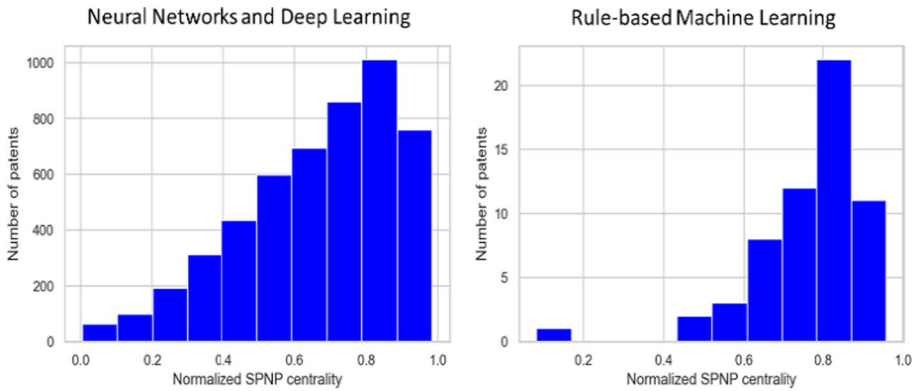
Fig. 6 The distribution of normalized centrality for Neural Networks and Deep Learning (left) vs Rule-based Machine Learning (right). Even though the former has many more patents, the latter has a larger fraction of high-centrality patents, resulting in a higher estimated TIR

There are other reasons behind the low estimated TIR of Deep Learning as well. Thompson et al. note that from a statistical point of view, a k-fold improvement in the performance of a model requires at least k^2 more datapoints; however because DL models are over-parametrized, this factor is at least k^4. Moreover, in practice, one has to increase the computational power by a factor of k^9 (Thompson et al., 2021). Deep Learning is also not very explainable, meaning that one cannot easily understand how a Deep Learning model arrives at a conclusion. The implications of our finding are discussed in the Conclusion. Our analysis also highlights the dangers hidden in using the raw number of patents or publications to assess a domain's progress.

As for other AI subdomains, *Rule-based Machine Learning* makes use of Genetic Algorithms, probability density estimation, Fuzzy Logic, as well as Bayesian Networks in its rule discovery, and it is used in classification, regression, and reinforcement learning. Its high TIR (about 2.6 standard deviations more than the mean TIR for AI technical subdomains) is therefore justified by Principle 2, above. It is also by design very interpretable.

Multi-task Learning (MTL) reduces training time and the needed computationsl power, compared to solving the problems separately. MTL effectively increases the sample size of a problem and therefore, is useful for under-sampled problems (limited data) which constitute an important subclass of Machine Learning problems. This and the fact that it is built on top of other AI methods result in the high TIR of MTL methods, see Rule 2, above. Even though *Logic Programming* has lost its spot as the dominant AI method, and it is argued that logic and other simple principles are not able to solve complex problems in Computer Vision and NLP (Crevier, 1993), logic still plays a significant role in intelligence (Darwiche, 2020). Logic Programming has been able to adapt to modern learning methods such as probabilistic learning (*Statistical Relational Learning* (Natarajan et al., 2014)) and Neural Networks (*Neuro-symbolic Computing* (Sarker et al., 2021)).

Theoretical AI subdomains with below-average estimated TIRs are discussed in the Supplementary Materials. Here we turn our attention to functional application subdomains shown in Fig. 5. We note that the unsupervised nature of NLP methods such as word representations (embeddings), and the availability of enormous amounts of training text data

have a great impact on the NLP's high improvement rate. This means that NLP algorithms can learn the meaning of words by studying their co-occurrences in corpora of texts (Devlin et al., 2018; Mikolov et al., 2013). *Predictive Analytics* also makes use of almost all available ML methods. The lower than average estimated TIR of *Autonomous Vehicles* and *Robotics* is attributed to their use of physical components (Rule 1), being more expensive to develop, as well as safety and legal concerns. We observe that NLP methods have a higher TIR than *Speech Processing* and that has a higher TIR compared to *Computer Vision*. This is because the complexity of data (e.g. in terms of size and representation dimension) increases from text to speech to image and video.

We also note that *Generative AI* subdomains, such as Natural Language Generation and Speech Generation, have very high estimated TIRs, and this is validated by a recent flurry of activity in Generative AI, especially text generation (Zhao et al., 2023). *Signal Separation* has low estimated TIR and this can be attributed to the fact that it is a highly underdetermined problem meaning that there are many more unknowns than constraints. As per *Human–Computer Interface*, its progress needs experiments on live mammals which is expensive and raises legal issues. Therefore, its low TIR is justified by Rule 1.

Data limitations

Using patents to assess technological progress has its own shortcomings. Patents do not include all advancements within a specific field, e.g. because of interest in keeping an invention secret, or a lack of inclination towards safeguarding novel inventions. Additionally, inventions do not encompass every facet of a technological domain (Benson & Magee, 2015b).

Using patents to measure technological progress in Artificial Intelligence has extra caveats as follows. Since the evaluation process for patent applications is a time consuming process, patents typically fall behind the pace of technology and do not necessarily reflect the most recent trends. Verendel notes that patenting is less useful for technologies in which innovation cycle is shorter than patenting process (Verendel, 2023). Some AI breakthrough ideas, such as backpropagation are not patented, even though others such as the transformer neural network are (Vaswani et al., 2017). The papers on such inventions are usually posted as preprints on the Arxiv preprint server, and then are presented at computer science and AI conferences. Note that, there are restriction on patenting abstract inventions both in the US (Patent & Trademark Office, 2019) and Europe (European Patent Office, 2019). Moreover, many AI tools are released as open source software by companies such as Meta, Google and others and are not patented. Examples include the PyTorch deep learning libraries and the Llama large language model (Touvron et al., 2023). Such inventions are not detected, at least in their pure form, by the sole analysis of patent datasets. As such, our analysis pertains more to the industrial aspect of AI.

It should also be noted that patents often combine more than one technology, and a patent which is identified as an AI patent, using keyword or classification criteria, usually involves the application of AI to other technologies. For this reason, the centrality of an AI patent may not be solely due to its AI aspect, and methods for quantifying the relevance of different technologies in a patent can be helpful in performing a more refined analysis.

Conclusion

In this study, we quantify the importance of patented innovations in Artificial Intelligence, using the notion of centrality in the Patent Citation Network, and this goes beyond the count of patents and even patent citations. We were thus able to provide estimates for the performance improvement rates of AI subdomains, based on the method of (Triulzi et al., 2020). As discussed in the Introduction, in the absence of a general-purpose performance measure for AI methods, our estimated TIR may be used as a proxy for such a measure. We also provide evidence, based on the literature, for our results. Despite the limitations of the method we use (discussed in the Discussion), it indeed quantifies which AI subdomains are more central in the large-scale flow of technology, and hence are likely to improve faster.

Overall, AI has a very high estimated TIR and is second only to one field (Optical Telecommunications) in this regard, (Fig. 7). The fact that software domains have one of the highest TIRs among technological domains has been established independently as well (Singh et al., 2021). The fact that the highest estimated TIR among applied AI subdomains belongs to Generative AI subdomains such as text generation, is highlighted by the recent boom in Large Language Models such as ChatGPT (Zhao et al., 2023).

We demonstrated the importance of research in, and the application of Rule-based Machine Learning, Multi-task Learning, Cognitive Modeling, Meta-Learning, Logic Programming, Knowledge Representations, and Distributed AI in the future progress of AI, as demonstrated by their high TIRs. It is interesting that our estimates show that more explainable AI methods such as Rule-based Machine Learning have a higher TIR than less comprehensible ones such as Deep Learning and Ensemble Learning.

Note that our results pertain to industrial and patented AI and may not reflect the academic significance of AI research or open-source AI innovations. Note also that TIR is intrinsic to a technological domain, meaning that external factors such as increased

Fig. 7 Comparison of the estimated TIR of Artificial Intelligence to 30 technological domains for which empirical TIR is known. AI is second only to Optical Telecommunication. Data for the 30 subdomains is taken from (Triulzi & Magee, 2020)
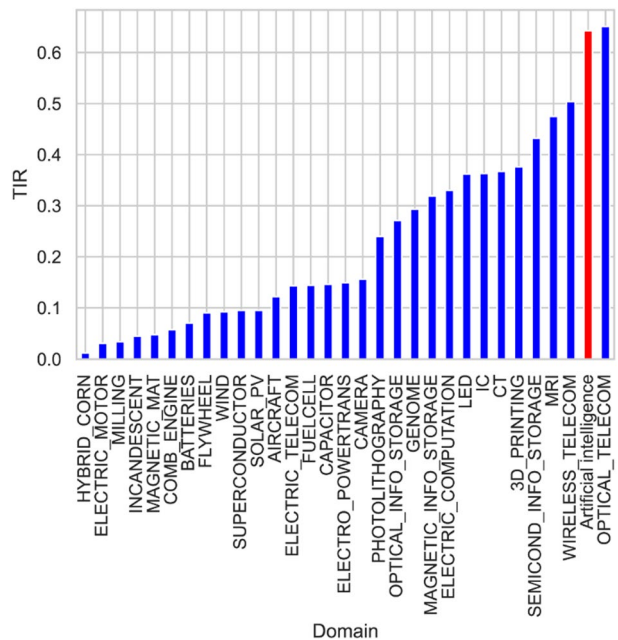
spending, do not accelerate the pace of a low-TIR domain (Basnet & Magee, 2017). Thus, the results of this paper illustrate which AI subdomains are more worthy of investment for future advancement of AI, and therefore, are of importance for policymaking and R&D purposes.

Specifically, contrasting the lower-than-average estimated TIR for Deep Learning with its recent history of achievements, we arrive at the conclusion that its progress can hit a bottleneck related to the needed computational power. Prominent AI research labs such as Open AI state that the trend of the rapid increase in computation needed for Deep Learning is likely to continue (OpenAI, 2022; Thompson et al., 2020). Given the coming end to Moore's Law for computer chips (Leiserson et al., 2020), OpenAI suggests that "policymakers should consider increasing funding for academic research into AI, as it's clear that some types of AI research are becoming more computationally intensive and therefore expensive" (OpenAI, 2022).

The results of the present study also suggest that Deep Learning research should aim at incorporating the advantages of other, high TIR AI subdomains, such as Rule-based Machine Learning, to mitigate the problems mentioned in the Discussion. There are currently two schools of thought concerning how Neural Networks and Symbolic AI can merge. The first one suggests that the two should be combined in hybrid AI models. Hybrid AI is an AI paradigm that has recently garnered more attention (Mao et al., 2019; Marcus, 2020). The other school of thought, held by the pioneers of Deep Learning, holds that instead of merging Symbolic AI and Deep Learning in hybrid systems, the latter can emulate the strengths of the former, in its own way. "We would like to design neural networks which can do all these things [what Symbolic AI aims at] while working with real-valued vectors so as to preserve the strengths of deep learning", write the three pioneers of Deep Learning, Y. Bengio, Y. LeCun and G. Hinton (Bengio et al., 2021). Interestingly, the Neural Production Systems (Goyal et al., 2021) developed by the Bengio team is a form of Rule-based Machine Learning in which the learning part is done by means of neural networks. These recent developments are all in accord with the results of our research. As (Thompson et al., 2021) notes, "We must either adapt how we do deep learning or face a future of much slower progress."

An important future step in this line of research is to enhance the TIR estimation method by including AI research publications. As mentioned in Discussion, many AI inventions are open source and are announced as preprints on the ArXiv preprint server, before being presented in computer science conferences. ArXiv preprints form a citation network and one can study their (normalized) centrality. Another refinement is related to the fact that patents often combine different technologies and, patents considered as AI patents usually use AI methods to aid in other endeavors. As such, we need a method of quantifying the belonging of a patent to different technologies. Finally, the regression model used by Triulzi et al. has room for improvement. The high correlation between a technology's TIR and the average centrality of its patents, raises the question of whether TIR has a network-theoretic counterpart. The regression model used here, ignores the temporal structure of the citation network. In other words, averaging centrality over patents granted in different years, ignores the fact that PCN is a network whose structure is dependent on time. It is also important to note that data on the performance improvement (in terms of accuracy, power consumption, etc.) is available for some AI models (Patterson et al., 2022) and one can use them to obtain empirical measures of those models' improvement rates. As mentioned in the introduction, there is ambiguity in how to combine these performance measure into a single indicator, and comparison to our estimates can be helpful in this regard.

References

Abadi, H. H. N., & Pecht, M. (2020). Artificial intelligence trends based on the patents granted by the united states patent and trademark office. *IEEE Access, 8*, 81633–81643.

Abood, A., & Feltenberger, D. (2018). Automated patent landscaping. *Artifical Intelligence and Law, 26*, 103–125.

AIME Planning Team. (2021). Artificial intelligence measurement and evaluation at the national institute of standards and technology

Alstott, J., Triulzi, G., & Yan, B. (2016). Mapping technology space by normalizing patent networks. *Scientometrics*. https://doi.org/10.1007/s11192-016-2107-y

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baruffaldi, S.H., Baruffaldi, S., Rao, N. (2020). Identifying and measuring developments in artificial intelligence

Basnet, S., & Magee, C. L. (2017). Artifact interactions retard technological improvement: An empirical study. *PLoS ONE, 12*, 1–17. https://doi.org/10.1371/journal.pone.0179596

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv Prepr. 1–27

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*, 58–65.

Benson, C. L., & Magee, C. L. (2015a). Quantitative determination of technological improvement from patent data. *PLoS ONE, 10*, 1–23. https://doi.org/10.1371/journal.pone.0121635

Benson, C. L., & Magee, C. L. (2015b). Technology structural implications from the extension of a patent search method. *Scientometrics, 102*, 1965–1985.

Crevier, D. (1993). *AI: The tumultuous history of the search for artificial intelligence*. Basic Books Inc.

Darwiche, A. (2020). Three modern roles for logic in AI. In: *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. pp. 229–243

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *AI Magazine, 14*, 17.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv Prepr. arXiv1810.04805

European Patent Office (2019) Technical character of an invention, https://www.epo.org/en/legal/case-law/2019/clr_i_d_9_1_1.html

Giczy, A. V., Pairolero, N. A., & Toole, A. A. (2022). Identifying artificial intelligence (AI) invention: A novel AI patent dataset. *The Journal of Technology Transfer, 47*, 476–505. https://doi.org/10.1007/s10961-021-09900-2

Goyal, A., Didolkar, A., Ke, N. R., Blundell, C., Beaudoin, P., Heess, N., Mozer, M. C., & Bengio, Y. (2021). Neural production systems. *Advances in Neural Information Processing Systems., 34*, 25673–25687.

Grace, K. (2013). Algorithmic progress in six domains. Machine Intelligence Research Institute.

Huisman, M., van Rijn, J. N., & Plaat, A. (2021). A survey of deep meta-learning. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-021-10004-4

Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*, 39–63.

Hurwitz, J., Kaufman, M., & Bowles, A. (2012). The foundation of cognitive computing. *Cognitive computing and big data analytics* (pp. 1–20). Wiley.

Jiang, L., Chen, J., Bao, Y., & Zou, F. (2022). Exploring the patterns of international technology diffusion in AI from the perspective of patent citations. *Scientometrics, 127*, 5307–5323. https://doi.org/10.1007/s11192-021-04134-3

Khayyam, H., Jamali, A., Bab-Hadiashar, A., Esch, T., Ramakrishna, S., Jalili, M., & Naebe, M. (2020). A novel hybrid machine learning algorithm for limited and big data modeling with application in industry 4.0. *IEEE Access, 8*, 111381–111393. https://doi.org/10.1109/ACCESS.2020.2999898

Lanzi, P., & Stolzmann, W. (2000). *Learning classifier systems: From foundations to applications*. Springer.

Lee, S., Hwang, J., & Cho, E. (2022). Comparing technology convergence of artificial intelligence on the industrial sectors: Two-way approaches on network analysis and clustering analysis. *Scientometrics, 127*, 407–452. https://doi.org/10.1007/s11192-021-04170-z

Leiserson, C. E., Thompson, N. C., Emer, J. S., Kuszmaul, B. C., Lampson, B. W., Sanchez, D., & Schardl, T. B. (2020). There's plenty of room at the Top: What will drive computer performance after Moore's law? *Science, 80*, 368. https://doi.org/10.1126/science.aam9744

Li, C. OpenAI's GPT-3 Language model: A technical overview, https://lambdalabs.com/blog/demystifying-gpt-3/

Liu, N., Shapira, P., & Yue, X. (2021a). Tracking developments in artificial intelligence research: constructing and applying a new search strategy. *Scientometrics, 126*, 3153–3192. https://doi.org/10.1007/s11192-021-03868-4

Liu, N., Shapira, P., Yue, X., & Guan, J. (2021b). Mapping technological innovation dynamics in artificial intelligence domains: Evidence from a global patent analysis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0262050

Liu, Z., Lin, Y., & Sun, M. (2020). *Representation learning for natural language processing*. Springer.

Lohn, A.J., Musser, M. (2022). AI and compute: How much longer can computing power drive artificial intelligence progress?

Mao, J., Gan, C., Kohli, P., Tenenbaum, J.B., Wu, J. (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. arXiv Prepr. arXiv1904.12584

Marcus, G. (2018) Deep learning: A critical appraisal. arXiv Prepr. arXiv1801.00631

Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. arXiv e-prints

Mattson, P., Reddi, V. J., Cheng, C., Coleman, C., Diamos, G., Kanter, D., Micikevicius, P., Patterson, D., Schmuelling, G., Tang, H., Wei, G.-Y., & Wu, C.-J. (2020). MLPerf: An industry standard benchmark suite for machine learning performance. *IEEE Micro, 40*, 8–16. https://doi.org/10.1109/MM.2020.2974843

Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics, 49*, 93–123. https://doi.org/10.1023/A:1005613325648

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv Prepr. arXiv1301.3781

Natarajan, S., Kersting, K., Khot, T., & Shavlik, J. (2014). Statistical relational learning. *SpringerBriefs Computer Science*. https://doi.org/10.1007/978-3-319-13644-8_2

OpenAI. (2022). AI and Compute. Blog Open AI. 1–11

Pan, Y. (2016). Heading toward artificial intelligence 20. *Engineering, 2*, 409–413. https://doi.org/10.1016/J.ENG.2016.04.018

Pandey, S., Verma, M. K., & Shukla, R. (2021). A scientometric analysis of scientific productivity of artificial intelligence research in India. *Journal of Scientometric Research, 10*, 245–250. https://doi.org/10.5530/JSCIRES.10.2.38

Patent and trademark office. (2019). 2019 revised patent subject matter eligibility guidance, https://www.federalregister.gov/documents/2019/01/07/2018-28282/2019-revised-patent-subject-matter-eligibility-guidance

Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D. R., Texier, M., & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. *Computer, 55*, 18–28. https://doi.org/10.1109/MC.2022.3148714

Rejmaniak, R. (2021). Bias in artificial intelligence systems. *Białostockie Studia Prawnicze, 3*, 25–42.

Rudin, C., & Carlson, D. (2019). The secrets of machine learning: ten things you wish you had known earlier to be more effective at data analysis. *Operations research & management science in the age of analytics management science in the age of analytics* (pp. 44–72). INFORMS.

Sarker, M.K., Zhou, L., Eberhart, A., Hitzler, P. (2021). Neuro-symbolic artificial intelligence: current trends. arXiv e-prints

Sharifzadeh, M., Triulzi, G., & Magee, C. L. (2019). Quantification of technological progress in greenhouse gas (GHG) capture and mitigation using patent data. *Energy & Environmental Science, 12*, 2789–2805. https://doi.org/10.1039/c9ee01526d

Singh, A., Triulzi, G., & Magee, C. L. (2021). Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description. *Research Policy., 50*, 104294. https://doi.org/10.1016/j.respol.2021.104294

Strubell, E., Ganesh, A., McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *ACL 2019—Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. pp. 3645–3650

Taheri, S., & Aliakbary, S. (2022). Research trend prediction in computer science publications: A deep neural network approach. *Scientometrics, 127*, 849–869. https://doi.org/10.1007/s11192-021-04240-2

Thompson, N.C., Greenewald, K., Lee, K., Manso, G.F. (2020) The computational limits of deep learning. arXiv e-prints

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2021). Deep learning's diminishing returns: the cost of improvement is becoming unsustainable. *IEEE Spectrum, 58*, 50–55. https://doi.org/10.1109/MSPEC.2021.9563954

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M., Subramanian, R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., Scialom, T. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv e-prints

Triulzi, G., Alstott, J., & Magee, C. L. (2020). Estimating technology performance improvement rates by mining patent data. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2020.120100

Triulzi, G., & Magee, C. L. (2020). Functional performance improvement data and patent sets for 30 technology domains with measurements of patent centrality and estimations of the improvement rate. *Data in Brief, 32*, 106257. https://doi.org/10.1016/j.dib.2020.106257

Tseng, C.-Y., & Ting, P.-H. (2013). Patent analysis for technology development of artificial intelligence: A country-level comparative study. *Innovation., 15*, 463–475. https://doi.org/10.5172/impp.2013.15.4.463

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning, 109*, 373–440. https://doi.org/10.1007/s10994-019-05855-6

Vanschoren, J. (2018). Meta-learning: a survey. 1–29

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł, & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5999–6009.

Verendel, V. (2023). Tracking artificial intelligence in climate inventions with patent data. *Nature Clinical Practice Endocrinology & Metabolism, 13*, 40–47. https://doi.org/10.1038/s41558-022-01536-w

(2019). WIPO: WIPO Technology Trends 2019: Artificial Intelligence. World Intellectual Property Organization

Yu, C., & Yao, W. (2017). Robust linear regression: A review and comparison. *Communications in Statistics - Simulation and Computation, 46*, 6261–6282. https://doi.org/10.1080/03610918.2016.1202271

Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2021.3070203

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z. (2023). A survey of large language models. arXiv Prepr. arXiv2303.18223