DOI: 10.2478/orga-2025-0012

Open-Source Transformer-Based Information Retrieval System for Energy Efficient Robotics Related Literature

Tine BERTONCEL

University of Primorska, Faculty of Management, Koper, Slovenia, tine.bertoncel@fm-kp.si

Background and Purpose: This article employs the Hugging Face keyphrase-extraction-kbir-inspec machine learning model to analyze 654 abstracts on the topic of energy efficiency in systems and control, computer science and robotics.
Methods: This study targeted specific arXiv categories related to energy efficiency, scraping and processing abstracts with a state-of-the-art Transformer-based Hugging Face AI model to extract keyphrases, thereby enabling the creation of related keyphrase networks and the retrieval of relevant scientific preprints.
Results: The results demonstrate that state-of-the-art open-source machine learning models can extract valuable information from unstructured data, revealing prominent topics in the evolving field of energy-efficiency.
Conclusion: This showcases the current landscape and highlights the capability of such information systems to pinpoint both well researched and less researched areas, potentially serving as an information retrieval system or early warning system for emerging technologies that promote environmental sustainability and cost efficiency.

Keywords: *Energy efficiency, Keyphrase extraction, Early warning system, Information system, Semantic network, Transformers, Industry 4.0*

1 Introduction

The fourth industrial revolution, or Industry 4.0, propelled by advancements in cyber-physical systems, artificial intelligence (AI), big data, and the Internet of Things, demands innovative methodologies to evaluate the impacts of these technologies on society and industry (Boas et al., 2005; Jazdi, 2014; Warner & Wäger, 2019). Industry 4.0 integrates highly adaptable robotic systems and energy-efficient solutions to boost sustainability in manufacturing, thereby aiding environmental protection. This industrial paradigm holds significant potential for sustainable transformation through green production, smart digitization, and a commitment to environmental stewardship (Wang et al., 2017; Jena et al., 2019; Rocca et al., 2020; Dinu 2024).

Research on energy efficiency in robotics has utilized multiple innovative approaches to reduce energy consumption, from more efficient algorithms to energy-saving hardware design, among other things (Carabin et al., 2017; Wang et al., 2017; Ghungrad et al., 2023). Digital transformation, epitomized by Industry 4.0, acts as a disruptive force in small and medium manufacturing enterprises, enhancing organizational agility, improving competitiveness, and driving industry wide change. This transformation encourages the adoption of technologies that enhance the energy efficiency and technological sophistication of manufacturing processes through robotics. Such changes promote sustainability and innovation, positioning SMEs for competitive advantages in a rapidly evolving industrial

Received: 5th January 2025; Accepted: 12th April 2025

landscape (Ghobakhloo, 2020; Roblek et al., 2021; Philbin et al., 2022).

Text mining, a branch of AI, analyzes texts extensively, in order to extract valuable insights, often from unstructured text found in project documents, emails and social media posts, among other sources. This accelerates project management objectives and boosts digital strategies by providing deep insights into topics, while also tackling the challenges of traditional decision support systems amid increasing amounts of textual data, converting natural language into actionable insights and managing information overload efficiently (Froelich & Ananyan, 2008; Gajzler, 2010; Khan, 2018; Vasiliev & Goryachev, 2022).

Text mining has been shown to be a topic of interest for research in the field of information systems. Unstructured data accounts for 80 percent of today's data due to Web 2.0 and social media, particularly when manual analysis of these documents is too time consuming to be a worthwhile consideration. Text mining goes beyond information retrieval, aiming to discover relationships between texts, as well as create new information. Text mining covers several topics, all of which can help discover knowledge that would otherwise remain hidden or hard to find (Babu et al., 2014; Debortoli et al., 2016; Firoozeh et al., 2020).

One way to mine text is to use Automatic Term Extraction (ATE). Initially relying on handcrafted rules and NLP tools, ATE systems progressed to incorporate statistical measures and, later, hybrid approaches combining linguistic and statistical information. The latest advancements in ATE leverage neural techniques, particularly Transformer-based models, which offer automatic feature deduction and domain independence. These neural systems either utilize embedding representations for classification or fine-tune pretrained language models through transfer learning. Throughout this evolution, the core ATE process has remained consistent: extracting candidate terms and then determining their validity (Tran et al., 2023).

The goal of this article is to test a new open-source state-of-the-art (SOTA) natural language processing (NLP) model for ATE (henceforth referred to as keyphrase extraction), in order to do text mining and potentially reveal dominant research themes. Another goal is to construct a network that can be used to quickly retrieve relevant keyphrases and scientific articles related to energy efficiency within the field of systems and control, computer science and robotics. The results of the current study could guide future research and implementation strategies in a manner that prioritizes sustainability and societal well-being. Such a system could help scientists, engineers, managers in industry and policy makers, in cases when information needs to be retrieved as efficiently as possible for the purpose of quick decision making.

2 Method

2.1 Data retrieval (arXiv)

The arXiv preprint repository, initiated by physicist Paul Ginsparg in the early 1990s, has expanded from its origins in physics to include numerous disciplines, such as computer science and robotics (cs.RO). It serves as a platform for open access articles, accessible prior to peer review, and is supported by institutions like Cornell University and the Simons Foundation. The structured taxonomy of arXiv aids in the efficient organization and retrieval of over 1.5 million scientific article preprints uploaded to arXiv since 1991 until the end of 2024, approximately 0.5 million of these being in computer science and physics. It supports researchers in exploring a vast array of scholarly work and presents an intriguing prospect for benchmarking next-generation machine learning models (Clement et al., 2019; Rosenbloom, 2019; arXiv, 2024; Bagchi et al., 2024).

Also, several authors have found that authors that publish arXiv preprints receive more citations in the long run and is regarded as a contemporary counterpart to the conventional practice of manuscript sharing among peers for swift dissemination of findings (Davis & Fromerth, 2006; Moed, 2006; Sutton & Gong, 2017; Ferrer-Sapena et al., 2018; Bagchi et al., 2024). While arXiv e-print prevalence in computer science varies widely; it exceeds 60 percent in theoretical computer science and machine learning, but remains minimal in other areas, though generally on the rise. In addition, 23 percent of all papers in 2017, on the topic of computer science, were published on arXiv, compared to only 1 percent in 2007 (Sutton & Gong, 2017).

For the purpose of this study, which is to test the usefulness of current AI solutions for creating networks of related keyphrases on the topic of energy efficiency and as a means of retrieving scientific preprints related those keyphrases, the categories cs.RO (computer science and robotics), cs.SY (Systems and Control) and eess.SY (Systems and Control) were selected for web scraping (see Table 1).

The category cs.RO is the most direct and obvious choice, as energy efficiency and robotics are being researched. The category cs.SY and eess.SY were also scraped, as they are fields critical for robotics, CPS, and energy-related systems. They deal with the analysis, design, and optimization of control systems, which are essential for making robots and CPS operate effectively (arXiv, 2024).

For scraping purposes, pandas, xml.etree.ElementTree, io, and requests libraries were used. The http request used the following boolean query: "cs.SY" OR "eess.SY" OR "cs.RO" AND "energy efficiency". Http requests were

Table 1: Subcategories of cs.RO, eess.SY / cs.SY

Category	Sub-categories and Focus Areas
Robotics (cs.RO)	This includes autonomous vehicles, commercial robotics and applications, kinematics / dynamics, manipulators, interfaces, propulsion, sensors, workspace organization (arXiv, 2024; Association for Computing Machinery, 2012)
Systems Engineering (eess.SY/cs.SY)	This includes automatic control systems, using robotics, reinforcement learning, sensor networks, cyber-physical and energy systems, among others. The category cs.SY is an alias for eess.SY (arXiv, 2024)

Source: Author's work

made until no new data was being returned. Each request returned an XML file, from which information of interest was extracted into a Pandas DataFrame. For the purpose of the study, the column 'abstracts' in the DataFrame was used to test the open-source SOTA NLP model, where keyphrases were extracted from each row of the column.

2.2 Data processing (Hugging Face AI Model)

A neural network model, employing the Transformer architecture, was utilized in this study. First introduced by Google in 2017, the Transformer has become a dominant architecture in natural language processing (NLP). It underpins several prominent commercial large language models (LLMs), including GPT-4, Claude, and Gemini. Additionally, open-source models developed by organizations such as GitHub, Google, Microsoft, Hugging Face, Facebook, and Salesforce, have also significantly contributed to this architecture and related open-source machine learning platforms (Gauci et al., 2018; Kochhar et al., 2021; Naveed et al., 2024).

The Transformer architecture surpasses convolutional and recurrent neural networks in language understanding and generation tasks. It scales effectively with both data and model size, facilitates efficient parallel training, offers multimodal representations, and has the ability of self-attention. The Transformers Python library, developed by Hugging Face, provides robust implementations suitable for both research and production environments. It includes comprehensive tools for tokenization, fine-tuning, and deployment, and offers compatibility with PyTorch and TensorFlow. Additionally, the library's Model Hub hosts an extensive array of pretrained models, enhancing accessibility to advanced NLP technologies and promoting community collaboration (M. Chen et al., 2019; Shin & Narihira, 2021; Yang et al., 2021; Bengesi et al., 2023).

Keyphrase extraction automates the extraction of representative phrases from documents, enhancing digital information systems with applications in semantic indexing, search, clustering, and classification. Keyphrases consist of multiple words, and serve a variety of purposes, such as identifying representative phrases from a document that succinctly summarize its content (Papagiannopoulou & Tsoumakas, 2019).

Keyphrase extraction tools, such as those found on Hugging Face, leverage deep learning techniques to pinpoint critical phrases in scientific documents. However, their capabilities are often confined to English-language documents and may falter in other linguistic or contextual settings, as is the case with keyphrase-extraction-kbir-inspec. Keyphrases are typically categorized as either extractive, derived directly from the text of the document, or abstractive, which, although not explicitly present in the document, effectively summarize its content.

For this study, the keyphrase-extraction-kbir-inspec model, developed by the ML6team and available on Hugging Face, was selected due to its state-of-the-art (SOTA) status. This extractive model is based on the Transformers architecture, finetuned on the Inspec dataset and demonstrates proficiency in extracting key phrases from scientific paper abstracts, achieving an F1 score of 62 percent on the Inspec dataset (Kulkarni et al., 2022; ML6Team, 2024; Zhu et al., 2024).

A Google Scholar search, using the keyword "keyphrase-extraction-kbir-inspec", yields 5 results showing that the tool has previously been used in research papers, since its release in March 2022 (ML6Team, 2024), to aid in the development of a hierarchical model for unraveling conspiracy theories (Melnick, 2024; Zhu et al., 2024), the development of the keyphrase extraction portion of a research project aimed at making scientific texts easier to understand for non-expert readers (Engelmann et al., 2023), finding NLP papers by asking multi-hop questions (Li & Takano, 2022), as well as a master's thesis on the automated selection of credible health information online (Bayani, 2024).

For this study, an array of Python libraries were utilized, including transformers, torch, sklearn, os, accelerate, re, concurrent.futures, psutil, pandas, nltk, gc, collections, ast, itertools, and matplotlib to facilitate tasks related to the Hugging Face model. Additionally, two custom modules were imported to enhance multiprocessing capabilities and streamline data cleaning. The code was run locally using an Nvidia Geforce 3070. Using Python, articles retrieved from the arXiv dataset underwent preprocessing.

First duplicates were removed, then each row in the 'abstract' column was processed to convert strings to lowercase and remove special characters. Stopwords were not removed, in order to make the text as close to the original as possible. The keyphrase extraction tool was then applied to extract keyphrases from each row. Subsequently, these keyphrases were lemmatized to reduce variations and decrease the complexity of the resultant networks. Finally, 50 example abstracts were selected and rated on how well the keyphrase extraction method determined keywords, in order to give a qualitative perspective on the resulting network.

The frequency of each keyword across all texts was computed, followed by the creation of tuples representing edges within a network graph, each tuple (edge) linking two keyphrases (nodes). Network graphs, which depict entities and their interconnections, can model diverse systems—from neuronal pathways to transportation networks (MATLAB, 2024). Analysis focused on the network's structure, specifically the number of edges and nodes. Additionally, a subnetwork was selected for detailed analysis to demonstrate what such an information retrieval system can do (see Figure 1 and Figure 2).

To visualize the data, the Python libraries collections, os, pandas, itertools, networkx, and matplotlib.pyplot were employed. Networks were designed to display up to ten edges per keyphrase to maintain clarity and prevent visual clutter in the representations. Such visualization can be used for many purposes, such as for networks of authors and publications across different academic fields (Kwon, 2022).

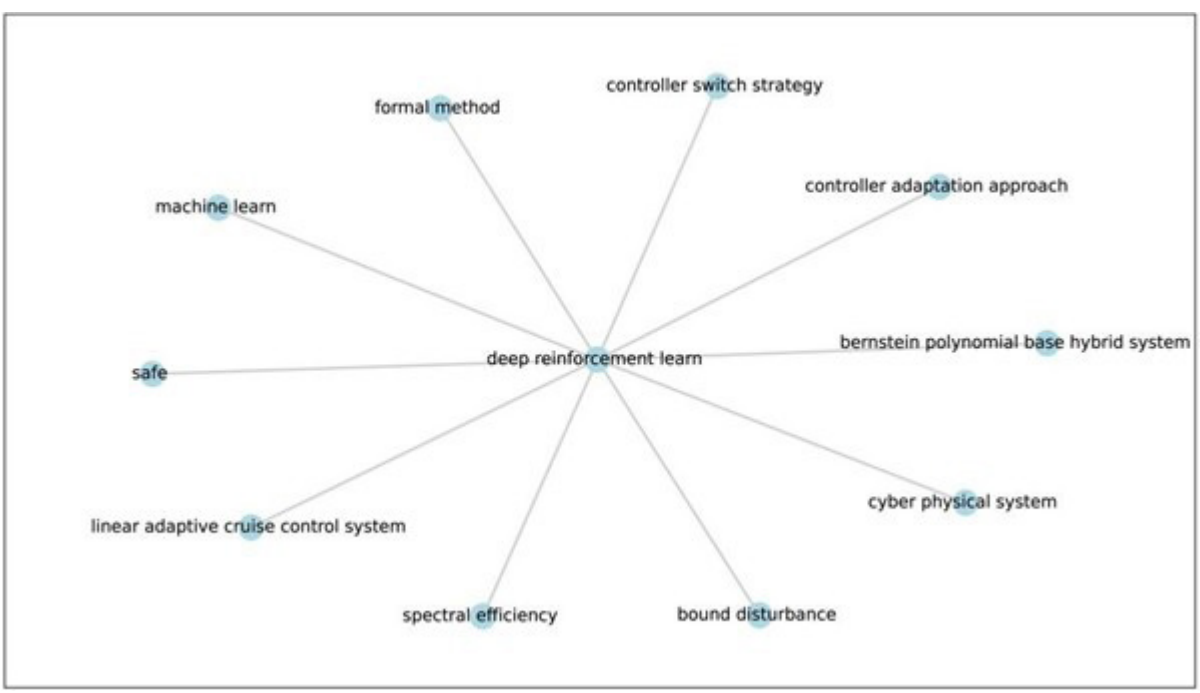

Figure 1: The initial keyphrase selected for traversing the network ("deep reinforcement learn.")
Source: Author's work

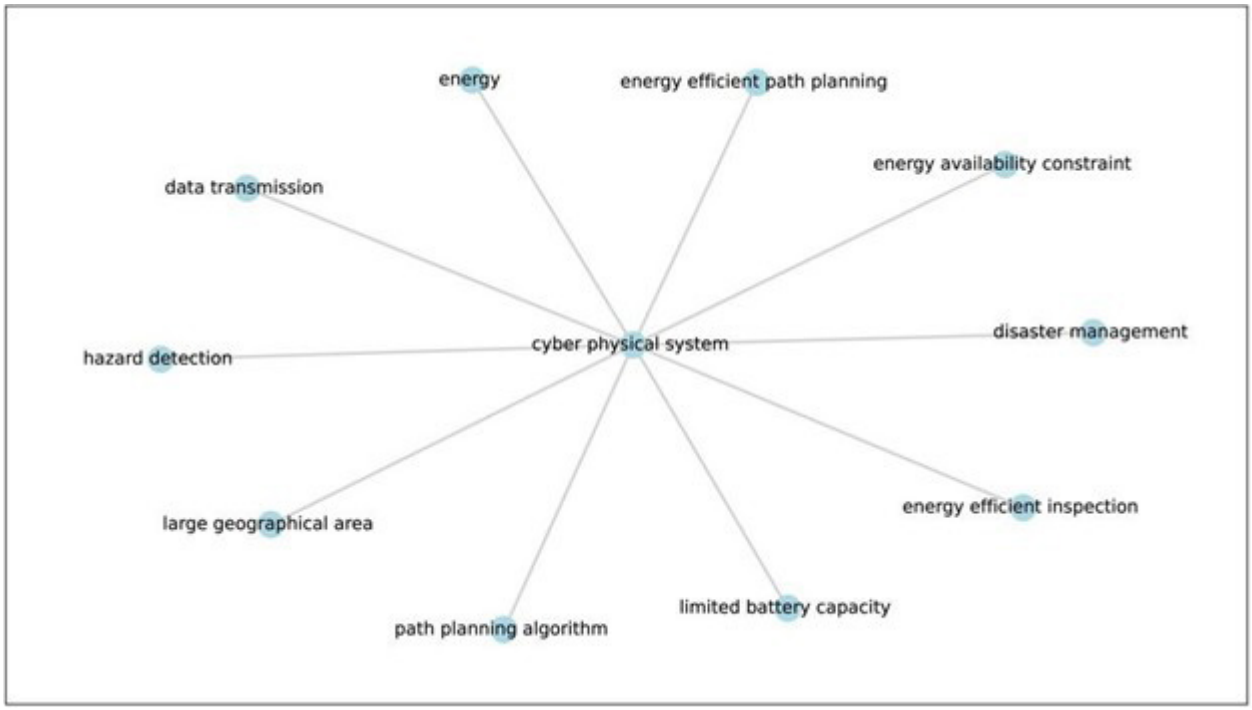

Figure 2: 10 edges of the keyphrases (node) "cyber physical system", expanding the network from Figure 1
Source: Author's work

Table 2: Keyphrase Frequencies Related to Energy Efficiency

(energy efficiency, 191), (reinforcement learn, 27), (deep reinforcement learn, 27), (unmanned aerial vehicle, 22), (simulation, 21), (energy consumption, 20), (machine learn, 18), (robustness, 17), (air conditioning, 17), (mobile robot, 17), (predictive control, 14), (efficiency, 14), (deep learn, 14), (model, 13), (legged robot, 13), (wireless network, 12), (automate vehicle, 12), (hvac system, 11), (autonomous vehicle, 11), (energy, 10), (sustainability, 10), (computational complexity, 10), (neuromorphic hardware, 10), (electric vehicle, 9), (hvac, 9), (occupant comfort, 9), (system, 9), (trajectory planning, 9), (trajectory optimization, 9), (heating ventilation, 8), (uncertainty, 8), (optimization, 8), (deep neural network, 8), (internet things iot, 8), (model predictive control, 8), (signalize intersection, 8), (power consumption, 8), (throughput, 8), (controller, 7), (privacy, 7), (fpga, 7), (security, 7), (simulation result, 7), (spectral efficiency, 7), (iot, 7), (cellular network, 7), (cost function, 7), (control, 7), (autonomous drive, 7), (locomotion, 7)

Note: The top 50 keyphrases are displayed above, each in a tuple showing the keyphrase and the frequency it appears in the abstracts.
Source: Author's work

3 Results and discussion

From a dataset of 1,092 articles retrieved from arXiv, 654 were found to be unique after merging the csRO, csSY, and eessSY dataframes and subsequently removing duplicates. The keyphrase extraction tool yielded 6,437 unique keyphrases, averaging approximately 10 keyphrases per abstract. An analysis of the Excel files for the categories csSY and eessSY revealed that csSY comprised of 449 preprints. Of these, 383 were also classified as eessSY preprints, which is the same number as there were eessSY preprints altogether. Additionally, there were 52 that were both categorized as eessSY and csRO. The category csRO, having 258 preprints, constituted about 39% of the total preprints.

Table 3: Number of Node Edges for Each Keyphrase

(energy efficiency, 2327), (deep reinforcement learn, 332), (reinforcement learn, 286), (energy consumption, 285), (unmanned aerial vehicle, 284), (simulation, 259), (robustness, 229), (machine learn, 221), (mobile robot, 204), (air conditioning, 179), (deep learn, 173), (predictive control, 164), (wireless network, 162), (efficiency, 161), (model, 153), (computational complexity, 146), (energy, 131), (legged robot, 131), (hvac system, 129), (trajectory optimization, 126), (autonomous vehicle, 124), (automate vehicle, 123), (electric vehicle, 122), (neuromorphic hardware, 121), (system, 120), (uncertainty, 114), (sustainability, 111), (trajectory planning, 109), (security, 107), (simulation result, 105), (signalize intersection, 103), (classification, 103), (optimization, 102), (internet things iot, 100), (occupant comfort, 97), (power consumption, 96), (model predictive control, 95), (throughput, 95), (spectral efficiency, 95), (convex optimization, 95), (control, 95), (hvac, 94), (controller, 94),, 94), (cellular network, 94), (dynamic program, 91), (privacy, 91), (iot, 89), (smart grid, 86), (data transmission, 85)

Note: The top 50 keyphrases with the most edges are shown above, each represented as a tuple with the keyphrase and its associated edge count.
Source: Author's work

Two-word phrases were found to be the most frequent across a wide variety of datasets. This pattern is consistent not only in scientific domains but also in business, media, and bureaucratic contexts (Nomoto, 2022), as well as in the corpora analyzed for the purposes of this study (see Table 2). Handling multi-word terms, as well as distinguishing terms from general words, appears to be done well by the keyphrase-extraction-kbir-inspec algorithm. This allowed for the creation of networks of keyphrases, as seen in Figure 1.

The current study has shown that using the keyphrase-extraction-kbirinspec model is a reliable method for extracting keyphrases. The 50 examples of articles, where keyphrases were manually looked at, appeared to be of high quality and demonstrated proficiency in extracting keyphrases from abstracts in scientific papers, ignoring stopwords and other irrelevant keyphrases from the abstract, leaving only keywords that represented the abstract.

The high quality could be due to the fact that it was fine-tuned on the Inspec dataset, that is on a collection of 2,000 expert-annotated computer science papers with identified keyphrases, as well as the fact that most of the preprints were on the topic of computer science (Kulkarni et al., 2022; ML6Team, 2024; Zhu et al., 2024).

As illustrated by comparing Table 2 with Table 3, although keyphrases that appear more frequently in the abstracts tend to have more edges, higher frequency does not invariably correlate with a greater number of edges.

Text mining within the realm of data analytics is increasingly acknowledged as an efficient method for leveraging unstructured textual data. By analyzing data, text mining can reveal new knowledge and reveal significant patterns and correlations that would otherwise remain obscured (Hassani et al., 2020).

The networks created during this study, as detailed in Table 2 and Table 3, demonstrate that text-mining methods such as keyphrase extraction, using Transformer-based architecture, are useful for identifying related keyphrases (the nodes and edges of the network). These results show that such methods can effectively reduce the amount of time required to identify relevant or related keyphrases, as opposed to reading through all of the abstracts individually and identifying relevant keywords that way.

Terminology plays a crucial role in specialized knowledge, particularly in its development, representation, and communication through language (Leonardi, 2009). As such, terminology is useful for managers and employees that need a common up-to-date and representative source of information about specialized knowledge within their company. While the keyphrases found in Table 2 and Table 3 are likely not representative of the full literature, they could be expanded upon by scraping additional scientific articles or literature that is representative of the field of energy efficiency and are not found on the arXiv website (or by looking at other categories on arXiv), eventually leading to a more complete collection of keyphrases relevant to improving energy efficiency within a company or elsewhere. It should be noted that only the top 50 most frequent keyphrases were looked at and that there are thousands remaining that could potentially show less researched areas that might be gaining importance, which could act as an early warning system.

The keyword "deep reinforcement learn" from Table 3 was chosen to traverse the network and find subnetworks. As depicted in Figure 1, 10 keyphrases were selected that are related to the keyphrase "deep reinforcement learn" (out of a total of 332 keyphrases). Among these 10 keyphrases, "cyber physical system" was selected for further

exploration, in order to delve deeper into the network structure (see Figure 2).

Deep learning is an advanced subset of AI and machine learning that uses multi-layered neural networks to learn directly from raw data. Unlike traditional approaches, it automatically discovers patterns without extensive human intervention, using multiple processing layers to create increasingly abstract data representations. This allows deep learning to handle large datasets efficiently, with its effectiveness typically improving as data volume grows. By mimicking the brain's hierarchical learning process, deep learning models excel at solving complex problems across various domains, making it a core technology of the Fourth Industrial Revolution (Sarker, 2021). Reinforcement Learning, on the other hand, is one of three main machine learning paradigms, in addition to supervised learning and unsupervised learning. In reinforcement learning, agents learn optimal behavior through trial-and-error interactions with an environment, without requiring labeled data (Al-Mahamid & Grolinger, 2021).

Finally, within the "cyber physical system" network, the keyword "energy efficient path planning" was chosen (see Figure 2), which was a node that has only two prepublications associated with it. According to M. Chen et al. (2019), energy-efficient path planning is defined as "given a start location, a goal location, and a set of obstacles distributed in a workspace: find a safe and efficient path for the robot".

The preprint abstract by Monwar et al. (2018) has the nodes "energy efficient path planning" and "cyber physical system" associated with it; however, it should be noted that this preprint no longer has "deep reinforcement learn" associated with it. The preprint proposes an energy-efficient path planning algorithm for a swarm of unmanned aerial vehicles (UAVs) tasked with inspecting a large geographical area. The algorithm aims to minimize the overall energy consumption of the swarm, taking into account the energy required for flying, hovering, and data transmission by each individual UAV (see Table 4).

According to Hambarde & Proença (2023), Information Retrieval (IR) "is to identify and retrieve information that is related to a user's query. As multiple records may be relevant, the results are often ranked according to their relevance score to the user's query." The above results show that using the keyphrase-extraction-kbir-inspec model is effective in creating an IR system, where the user traverses a network of keyphrases, in order to find articles of interest. As seen above, there is no score based on relevance, which is often a part of IR systems (Hambarde & Proença, 2023; Jiang et al., 2023); however, the network could still be useful to do an exploratory search of articles, especially when there are already specific keywords of interest to search for. This method could complement other methods, such as clustering or Retrieval Augmented Generation (RAG), which has been shown to be a viable way to reduce hallucinations in LLMs (Jiang et al., 2023). However, using RAG, instead of (or in addition to) keyphrase extraction / network creation, could potentially be a much more costly alternative, as LLM models, particularly bigger ones, typically use A100 or V100 Nvidia graphics cards, whereas this study used one consumer Geforce 3070 card (Samsi et al., 2023).

Recent advancements in commercial and open-source machine learning algorithms have produced model's adept at extracting information pertinent to decision-makers, in the current study, for those with expertise in energy efficiency, robotics or systems and control research; however, the proposed network of relevant kephrases used for information retrieval is only a proof-of-concept. There need to be several improvements made before the network is practical.

Before starting to improve such a system, one approach could involve having several domain experts validate portions of the network and rating the usefulness of such a system for their work. If such a network is highly rated, then further improvements could be done to the network. According to Rosenbloom (2019), arXiv moderates submissions for content appropriateness rather than scientific validity; thus, for creating networks with scientific validity, a collection of published peer-reviewed articles would be necessary.

Nevertheless, the current network could lead to decision-makers discovering a preprint that has undergone peer review since its release on arXiv. In addition, other information could be added to the network for greater detail, such as citation count, year of publication etc.

Future enhancements to the proposed network could involve categorizing keyphrases under umbrella terms and linking synonyms within the same network. These improvements could significantly increase the speed at which users could browse technologies or other relevant information, enabling instant access to all scientific articles related to the selected keyphrases.

In the future, it would be interesting to look at additional methodologies, such as clustering or RAG, or even more elaborate data wrangling approaches, as they could enhance the proposed system, particularly in regards to improving article retrieval, as well as provide a comparison to the current network.

An enhanced version of the current network could benefit managers in engineering and other decision-making roles by allowing them to spend less time searching for pertinent articles and more time analyzing articles that contain information related to potentially disruptive technologies. In a sense, it could act as an early warning system for aforementioned technologies. Additionally, it could help them discern connections that might not be immediately apparent.

Table 4: Preprint of article that had "cyber physical system" and "energy efficient path planning" in it

Title: Optimized Path Planning for Inspection by Unmanned Aerial Vehicles Swarm with Energy Constraints (DOI: 10.1109/GLOCOM.2018.8647342)
Abstract (snippet): Autonomous inspection of large geographical areas is a central requirement for efficient hazard detection and disaster management in future cyber-physical systems such as smart cities. In this regard, exploiting unmanned aerial vehicle (UAV) swarms is a promising solution to inspect vast areas efficiently and with low cost. In fact, UAVs can easily fly and reach inspection points, record surveillance data, and send this information to a wireless base station (BS). Nonetheless, in many cases, such as operations at remote areas, the UAVs cannot be guided directly by the BS in real-time to find their path. Moreover, another key challenge of inspection by UAVs is the limited battery capacity. Thus, realizing the vision of autonomous inspection via UAVs requires energy-efficient path planning that takes into account the energy constraint of each individual UAV...
The following are keyphrases that were extracted from the above abstract using Hugging Face model (keyphrase-extraction-kbir-inspec): ['autonomous inspection', 'cyber physical system', 'data transmission', 'disaster management', 'energy', 'energy availability constraint', 'energy efficient inspection', 'energy efficient path planning', 'hazard detection', 'large geographical area', 'limited battery capacity', 'path planning algorithm', 'polynomial time', 'smart city', 'surveillance data', 'unmanned aerial vehicle uav swarm', 'wireless base station']

Note: Example article and related information that a node can lead to using the proposed information system.
Source: Author's work

Also, in the current state, as was already mentioned, synonyms are not grouped within the same network, which means that it is possible not all relevant preprints related to "cyber-physical systems" and "energy efficient path planning" were identified within the subnetwork in this study. Moreover, the list of extracted keyphrases is extensive, making it challenging to define a few overarching categories for all of the keyphrases, though not impossible. Also, the network was traversed by starting with the keywords "deep reinforcement learn", which was associated with "cyber physical system" but was not associated with "energy efficient path planning". As a result, to improve the network, a feature could be added that informs the user when a keyphrase of interest is no longer associated with other keyphrases of interest downstream.

Such a final version of the network as is mentioned above, could help companies save money and be seen as environmentally friendly. Energy efficiency improvements are crucial for increasing product competitiveness in the global market, which can lead to decreased energy-related operating costs, increased return on equity, return on assets, return on investment, and return on sales, among other things (Backlund et al., 2012; Fan et al., 2017; Melnik & Ermolaev, 2020; de la Rue du Can et al., 2022; Knuutila et al., 2022)

The European Regional Development Fund and the Cohesion Fund were the primary EU funds targeting energy efficiency in enterprises, allocating €2.4 billion from 2014-2020. Estimates indicate that saving one unit of energy was cheaper than purchasing the same amount of electricity, suggesting that these investments were generally efficient (ECA, 2022).

Backlund et al. (2012) found that energy-intensive firms seem to be more successful when it comes to adopting energy management practices, e.g. an employed energy manager and the existence of an energy strategy. However, all companies should be thinking of increasing energy efficiency. For example, SMEs represent 99 percent of global businesses and 13 percent of world energy consumption. Despite barriers like high costs and lack of awareness, they have significant potential to improve energy efficiency. Low-cost measures and larger investments in processes and energy supply can lead to substantial savings and benefits, contributing to climate change mitigation and sustainable development (Gennitsaris et al., 2023).

Research indicates that maximum warming from CO2 emissions occurs about a decade after emission, with actions to reduce emissions potentially yielding benefits

within our lifetimes (Ricke & Caldeira, 2014). The Paris Agreement aims to limit global warming to 2°C above pre-industrial levels, with probabilistic analysis showing a 25 percent chance of staying below this threshold if cumulative CO2 emissions are limited to 1,000 Gt CO2 by 2050. The chance of staying below the threshold was predicted to be 50 percent if we reduce CO2 emissions by 1,440 Gt. To get a perspective of what that means, it is important to note that 234 Gt CO2 were emitted between 2000 and 2006 (Meinshausen et al., 2009). Additionally, Smith et al. (2018) predicted a 38 percent chance of exceeding the 1.5 °C threshold in a given month and a 10 percent chance in any given year between 2017 to 2021 (Smith et al., 2018). The Great Barrier Reef is already experiencing coral die-off when heat exposure surpasses critical thresholds, and by the 2030s, major crops will face extreme heat exposure, threatening food security. Such historical trends and future projections underscore the urgent need for robust, long-term climate strategies to mitigate ongoing warming (Hansen et al., 2006; Gourdji et al., 2013; Hughes et al., 2018). Overall, these studies amplify the need for quick decision making in regards to implementation of technologies that are energy efficient.

Text mining, particularly those advanced methods such as AI, and IT systems, such as Information Systems, use a lot of energy. Overall, the ICT sector's electricity consumption was estimated to be 4.7 percent of the global total, contributing approximately 1.7 percent of global CO2 emissions (National Research Council, 2011; The World Bank and ITU, 2024). However, implementing artificial intelligence could lower energy consumption and carbon emissions by about 8 to 19 percent by 2050. When combined with energy policies and low-carbon power generation, it could potentially reduce energy consumption by 40 percent and carbon emissions by 90 percent compared to business-as-usual scenarios (Ding et al., 2024). This is highly important, as energy efficiency does not only help organizations save money, but it also helps fight climate change. Climate change poses a significant threat, already damaging urban and natural systems and causing global economic losses exceeding $500 billion (L. Chen et al., 2023). Ritchie & Roser (2020) indicated that the energy sector accounts for 73.2 percent of all CO2 emissions, with industrial energy use contributing 24.2 percent of these emissions. Swifter implementation of new energy-efficient solutions is crucial for reducing greenhouse gas emissions, aligning with the EU's commitment to combating climate change.

Meeting the Paris Agreement targets is essential to limit global warming to less than 2°C above preindustrial levels, with an aspirational goal of not exceeding 1.5°C. Rapidly adopting these technologies is vital for achieving these objectives (Brugger et al., 2021; Dinu et al., 2023; Virjan et al., 2023).

Such an information retrieval system can be useful to various kinds of SMEs, particularly in manufacturing and logistics, where highly energy intensive process benefit greatly from a Knowledge Management system. KM process, such as knowledge acquisition, dissemination, and application, significantly contribute to environmental, economic, and social sustainability, by increasing green innovation and organizational agility, which in turn enhance corporate sustainability performance. Overall, KM is vital for integrating sustainable strategies across firms, supporting both innovation and long-term sustainability goals (Abbas & Sağsan, 2019; López-Torres et al., 2019; Shahzad et al., 2020; Sharma, Jabbour, & Lopes de Sousa Jabbour, 2021). The proposed system in the current study could help organizations create a new KM system or add to their pre-existing one. As such, further research is need to determine the usefulness of such open-source models in achieving these goals.

4 Conclusion

This study demonstrates the effectiveness of keyphrase extraction techniques, particularly the keyphrase-extraction-kbir-inspec model, in efficiently and accurately categorizing scientific abstracts. The results reveal that keyphrase networks, can be valuable in developing information retrieval systems, as demonstrated with energy efficiency scientific abstracts. The ability to traverse networks of keyphrases can provide decision-makers with rapid access to relevant information.

However, further improvements are necessary to enhance the usefulness of such a network. These improvements include incorporating peer-reviewed articles, validating the network through domain experts, linking synonyms, exploring additional methodologies like clustering and RAG, incorporating broader datasets in other arXiv categories or sources other than arXiv to ensure a comprehensive representation of the field, among other things.

The study underscores the critical role of energy efficiency in improving business competitiveness and mitigating climate change. Despite their own energy demands, the integration of AI and text mining tools could contribute significantly to reducing global energy consumption and carbon emissions, aligning with broader sustainability goals.

Acknowledgement

The activities are implemented within the framework of the GREENTECH project, co-financed by the European Union - NextGenerationEU.

Literature

Abbas, J., & Sağsan, M. (2019). Impact of knowledge management practices on green innovation and corporate sustainable development: A structural analysis. *Journal of Cleaner Production, 229*, 611-620. https://doi.org/10.1016/j.jclepro.2019.05.024

AlMahamid, F., & Grolinger, K. (2021). Reinforcement learning algorithms: An overview and classification. *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. https://doi.org/10.1109/ccece53047.2021.9569056

arXiv. (2024). *arXiv Category Taxonomy*. https://arxiv.org/category_taxonomy

Association for Computing Machinery. (2012). *The 2012 acm computing classification system*. https://www.acm.org/publications/class-2012

Babu, M. S., Ali, A. A., & Rao, A. S. (2014). A study on information retrieval methods in text mining. *International Journal of Engineering Research & Technology (IJERT)*, *2*(15). https://doi.org/10.17577/IJERTCONV2IS15028

Backlund, S., Broberg, S., Ottosson, M., & Thollander, P. (2012). Energy efficiency potentials and energy management practices in swedish firms. *European Council for an Energy Efficient Economy*, 11–14

Bagchi, C., Malmi, E., & Grabowicz, P. (2024). Effects of research paper promotion via arxiv and x. *arXiv.* https://doi.org/10.48550/arXiv.2401.11116

Bayani, A. (2024). *Sélection automatisée d'informations crédibles sur la santé en ligne* (Publication No. 33068) [Mémoire, Université de Montréal]. Papyrus: Institutional Repository. https://hdl.handle.net/1866/33068

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2023). Advancements in generative ai: A comprehensive review of gans, gpt, autoencoders, diffusion model, and transformers. *IEEE Access, 12*. https://doi.org/10.48550/arXiv.2311.10242

Boas, T., Dunning, T., & Bussel, J. (2005). Will the digital revolution revolutionize development? drawing together the debate. *Studies in Comparative International Development*, *40*, 95–110. https://doi.org/10.1007/BF02686296

Brugger, H., Eichhammer, W., Mikova, N., & Dönitz, E. (2021). Energy efficiency vision 2050: How will new societal trends influence future energy demand in the european countries? *Energy Policy*, *152*, 112216. https://doi.org/10.1016/j.enpol.2021.112216

Carabin, G., Wehrle, E., & Vidoni, R. (2017). A review on energy-saving optimization methods for robotic and automatic systems. *Robotics, 6*(4), 39. https://doi.org/10.3390/robotics6040039

Chen, L., Chen, Z., Zhang, Y., Liu, Y., Osman, A. I., Farghali, M., Hua, J., Al-Fatesh, A., Ihara, I., Rooney, D. W., & Yap, P.-S. (2023). Artificial intelligence-based solutions for climate change: A review. *Environmental Chemistry Letters*, *21*(5), 2525–2557. https://doi.org/10.1007/s10311-023-01617-y

Chen, M., Li, Y., & Li, R. (2019). Research on neural machine translation model. *Journal of Physics: Conference Series*, *1237*, 052020. https://doi.org/10.1088/1742-6596/1237/5/052020

Clement, C., Bierbaum, M., O'Keeffe, K., & Alemi, A. (2019). On the use of arxiv as a dataset. *arXiv*. https://doi.org/10.48550/arXiv.1905.00075

Davis, P., & Fromerth, M. (2006). Does the arxiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*, 203–215. https://doi.org/10.1007/s11192-007-1661-8

de la Rue du Can, S., Letschert, V., Agarwal, S., Park, W. Y., & Kaggwa, U. (2022). Energy efficiency improves energy access affordability. *Energy for Sustainable Development, 70, 560–568.* https://doi.org/10.1016/j.esd.2022.09.003

Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems*, *39*. https://doi.org/10.17705/1CAIS.03907

Ding, C., Ke, J., Levine, M., & Zhou, N. (2024). Potential of artificial intelligence in reducing energy and carbon emissions of commercial buildings at scale. *Nature Communications*, *15*(1), 5916. https://doi.org/10.1038/s41467-024-50088-4

Dinu, V., (2024). Innovative Application of Artificial Intelligence in Business Impacting Socio-Economic Progress. *Amfiteatru Economic*, 26(66), 398-401. https://doi.org/10.24818/EA/2024/66/398.

Dinu, V., Baciu, L. E., Mortan, M., & Veres, V. A. (2023). Effect of economic, institutional and cultural factors on the implementation of eu energy policies. *Amfiteatru Economic Journal, 25*(63), 306–306. https://doi.org/10.24818/EA/2023/63/306

Engelmann, B., Haak, F., Kreutz, C. K., Khasmakhi, N. N., & Schaer, P. (2023). Text simplification of scientific texts for non-expert readers. *arXiv*. https://doi.org/10.48550/arXiv.2307.03569

Fan, L., Pan, S., Liu, G., & Zhou, P. (2017). Does energy efficiency affect financial performance? evidence from chinese energy-intensive firms. *Journal of Cleaner Production*, *151*, 53–59. https://doi.org/10.1016/j.jclepro.2017.03.044

Ferrer-Sapena, A., Aleixandre-Benavent, R., Peset, F., & Sánchez-Pérez, E. (2018). Citations to arxiv preprints by indexed journals and their impact on research evaluation. *Journal of Information Science: Theory and Practice*, *6*, 6–16. https://doi.org/10.1633/JISTAP.2018.6.4.1

Firoozeh, N., Nazarenko, A., Alizon, F., & Daille, B. (2020). Keyword extraction: Issues and methods.*Natural Language Engineering*, *26*(3), 259–291. https://doi.org/10.1017/S1351324919000457

Froelich, J., Ananyan, S. (2008). Decision Support via Text Mining. In*: Handbook on Decision Support Systems 1. International Handbooks Information System* (pp 609–635). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-48713-5_28 Frontier Group. (2023, October 5). *Fact file: Computing is using more energy than ever*. https://frontiergroup.org/resources/fact-file-computing-is-using-more-energy-than-ever/

Gajzler, M. (2010). Hybrid advisory system for the process of industrial flooring repairs. *Technological and Economic Development of Economy*, *16*, 219–232. https://doi.org/10.1016/j.proeng.2013.03.056

Gauci, J., Conti, E., Liang, Y., Virochsiri, K., He, Y., Kaden, Z., Narayanan, V., Ye, X., & Fujimoto, S. (2018). Horizon: Facebook's open source applied reinforcement learning platform. *arXiv*. https://doi.org/10.48550/arXiv.1811.00260

Gennitsaris, S., Oliveira, M. C., Vris, G., Bofilios, A., Ntinou, T., Frutuoso, A. R., Queiroga, C., Giannatsis, J., Sofianopoulou, S., & Dedoussis, V. (2023). Energy efficiency management in small and medium-sized enterprises: Current situation, case studies and best practices. *Sustainability, 15*(4). https://doi.org/10.3390/su15043727

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. J*ournal of Cleaner Production*, *252*, 119869. https://doi.org/10.1016/j.jclepro.2019.119869

Ghungrad, S., Mohammed, A., & Haghighi, A. (2023). Energy-efficient and quality-aware part placement in robotic additive manufacturing. *Journal of Manufacturing Systems, 68*, 536–559. https://doi.org/10.1016/j.jmsy.2023.05.019

Gourdji, S., Sibley, A., & Lobell, D. (2013). Global crop exposure to critical high temperatures in the reproductive period: Historical trends and future projections. *Environmental Research Letters*, *8*. https://doi.org/10.1088/1748-9326/8/2/024041

Hambarde, K., & Proença, H. (2023). Information retrieval: Recent advances and beyond. *Journal of Information Retrieval*, *15*(3), 200–215. https://doi.org/10.1109/ACCESS.2023.3295776

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D., & Medina-Elizade, M. (2006). Global temperature change. *Proceedings of the National Academy of Sciences*, *103*, 14288–14293. https://doi.org/10.1073/pnas.0606291103

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, *4*(1). https://doi.org/10.3390/bdcc4010001

Hughes, T., Kerry, J., Baird, A., Connolly, S., Dietzel, A., Eakin, C., Heron, S., Hoey, A., Hoogenboom, M., Liu, G., McWilliam, M., Pears, R., Pratchett, M., Skirving, W., Stella, J., & Torda, G. (2018). Global warming transforms coral reef assemblages. *Nature, 556*, 492–496. https://doi.org/10.1038/s41586-018-0041-2

Jazdi, N. (2014). Cyber physical systems in the context of industry 4.0*. 2014 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, 1–4. https://doi.org/10.1109/AQTR.2014.6857843

Jena, M., Mishra, S., & Moharana, H. (2019). Application of industry 4.0 to enhance sustainable manufacturing. *Environmental Progress & Sustainable Energy*, *39*, 1–11. https://doi.org/10.1002/ep.13360

Jiang, Z., Xu, F. F., Gao, L., Sun, Z., Liu, Q., Dwivedi-Yu, J., Yang, Y., Callan, J., & Neubig, G. (2023). Active retrieval augmented generation. *arXiv*. https://doi.org/10.48550/arXiv.2305.06983

Khan, S. (2018). Text mining methodology for effective online marketing. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *3*(8), 465–469. https://doi.org//10.32628/CSEIT12283129

Knuutila, M., Kosonen, A., Jaatinen-Värri, A., & Laaksonen, P. (2022). Profitability comparison of active and passive energy efficiency improvements in public buildings. *Energy Efficiency*, *15*(6), 38. https://doi.org/10.1007/s12053-022-10046-9

Kochhar, P., Kalliamvakou, E., Nagappan, N., Zimmermann, T., & Bird, C. (2021). Moving from closed to open source: Observations from six transitioned projects to github. *IEEE Transactions on Software Engineering*, *47*, 1838–1856. https://doi.org/10.1109/TSE.2019.2937025

Kulkarni, M., Mahata, D., Arora, R., & Bhowmik, R. (2022). Learning rich representation of keyphrases from text. *arXiv*. https://doi.org/10.48550/arXiv.2112.08547

Kwon, H. (2022). Visualization methods of information regarding academic publications, research topics, and authors*. Proceedings, 81*(1). https://doi.org/10.3390/proceedings2022081154

Leonardi, N. (2009, January 1). *Terminology as a system of knowledge representation: An overview*. https://hdl.handle.net/11393/44305

Li, X., & Takano, T. (2022). Finding NLP papers by asking a multi-hop question. *Proceedings of the Annual Conference of JSAI*. https://doi.org/10.11517/jsaislud.95.0_59

López-Torres, G. C., Garza-Reyes, J. A., Maldonado-Guzmán, G., Kumar, V., Rocha-Lona, L., & Cherrafi, A. (2019). Knowledge management for sustainability in operations. *Production Planning & Control*, *30*(10–12), 813–826. https://doi.org/10.1080/09537287.2019.1582091

MATLAB. (2024). *Graph and network algorithms - matlab & simulink*. https://nl.mathworks.com/help/matlab/graph-and-network-algorithms.html

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S., Frieler, K., Knutti, R., Frame, D., & Allen, M. (2009). Greenhouse-gas emission targets for limiting global warming to 2°c. *Nature, 458*, 1158–1162. https://doi.org/10.1038/nature08017

Melnick, S. (2024). *A Computational Journey Through Conspiracy Theories: A Genealogical Approach* (Publication No. 1841) [Master's thesis, University of Vermont]. ScholarWorks. https://scholarworks.uvm.edu/graddis/1841/

Melnik, A., & Ermolaev, K. (2020). Strategy context of decision making for improved energy efficiency in industrial energy systems. *Energies, 13*(7). https://doi.org/10.3390/en13071540

ML6Team. (2023). *Keyphrase extraction model: Kbir-inspec*. https://huggingface.co/ml6team/keyphrase-extraction-kbir-inspec

Moed, H. (2006). The effect of 'open access' upon citation impact: An analysis of arxiv's condensed matter section. *arXiv.* https://doi.org/10.48550/arXiv.cs/0611060

Monwar, M., Semiari, O., & Saad, W. (2018). Optimized path planning for inspection by unmanned aerial vehicles swarm with energy constraints. *Proceedings of the IEEE Global Communications Conference and Ad Hoc and Sensor Networks Symposium*. https://digitalcommons.georgiasouthern.edu/electrical-eng-facpubs/138/

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A comprehensive overview of large language models. *arXiv.* https://doi.org/10.48550/arXiv.2307.06435

National Research Council (2011). The future of computing performance: Game over or next level. The National Academies Press. https://doi.org/10.17226/12980

Nomoto, T. (2022). Keyword extraction: A modern perspective. *SN Computer Science*, *4*(1), 92. https://doi.org/10.1007/s42979-022-01481-7

ECA (2022). *Energy efficiency in enterprises*. https://www.eca.europa.eu/Lists/ECADocuments/SR22_02/SR_Energy-effic-enterpr_en.pdf

Papagiannopoulou, E., & Tsoumakas, G. (2019). A review of keyphrase extraction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(2). https://doi.org/10.1002/widm.1339

Philbin, S., Viswanathan, R., & Telukdarie, A. (2022). Understanding how digital transformation can enable SMEs to achieve sustainable development: A systematic literature review. *Small Business International Review*, *6*(2), 1–12. https://doi.org/10.26784/sbir.v6i1.473

Ricke, M., & Caldeira, K. (2014). Maximum warming occurs about one decade after a carbon dioxide emission. *Environmental Research Letters*, *9*(12). https://doi.org/10.1088/1748-9326/9/12/124002

Ritchie, H., & Roser, M. (2020, September 18). *Greenhouse gas emissions by sector*. https://ourworldindata.org/ghg-emissions-by-sector

Roblek, V., Meško, M., Pušavec, F., & Likar, B. (2021). The role and meaning of the digital transformation as a disruptive innovation on small and medium manufacturing enterprises. *Frontiers in Psychology*, *12*, 592528. https://doi.org/10.3389/fpsyg.2021.592528

Rocca, R., Rosa, P., Sassanelli, C., Fumagalli, L., & Terzi, S. (2020). Integrating virtual reality and digital twin in circular economy practices: A laboratory application case. *Sustainability*, (6), 1–8. https://doi.org/10.3390/su12062286

Rosenbloom, L. (2019). Arxiv.org. *Charleston Advisor, 21*(2), 8–10. https://doi.org/10.5260/chara.21.2.8

Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., & Gadepally, V. (2023). From words to watts: Benchmarking the energy costs of large language model inference. *arXiv*. https://doi.org/10.48550/arXiv.2310.03003

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science, 2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Shahzad, M., Qu, Y., Zafar, A. U., Rehman, S. U., & Islam, T. (2020). Exploring the influence of knowledge management process on corporate sustainable performance through green innovation. *Journal of Knowledge Management, 24*(9), 2079-2106. https://doi.org/10.1108/JKM-11-2019-0624

Sharma, R., Jabbour, C. J. C., & Lopes de Sousa Jabbour, A. B. (2021). Sustainable manufacturing and industry 4.0: What we know and what we don't. *Journal of Enterprise Information Management, 34*(1), 230-266. https://doi.org/10.1108/JEIM-01-2020-0024

Shin, A., & Narihira, T. (2021). Transformer-exclusive cross-modal representation for vision and language. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2719–2725. https://doi.org/10.18653/v1/2021.findings-acl.240

Smith, D., Scaife, A., Hawkins, E., Bilbao, R., Boer, G., Caian, M., Caron, L., Danabasoglu, G., Delworth, T., Doblas-Reyes, F., Doescher, R., Dunstone, N., Hermanson, R., Ishii, M., Kharin, V., Kimoto, M., Koenigk, T., Kushnir, Y., Matei, D., ... Yeager, S. (2018). Predicted chance that global warming will temporarily exceed 1.5 °c. *Geophysical Research Letters*, *45*(21), 11, 895–11, 903. https://doi.org/10.1029/2018GL079362

Sutton, C., & Gong, L. (2017). Popularity of arxiv.org within computer science. *arXiv*. https://doi.org/10.48550/arXiv.1710.05225

The World Bank and ITU. (2024). *Document of the world bank: Project appraisal document*

on a proposed loan. https://documents1.worldbank.org/curated/en/099121223165540890/pdf/P17859712a98880541a4b71d57876048abb.pdf

Tran, H. T. H., Martinc, M., Caporusso, J., Doucet, A., & Pollak, S. (2023). The recent advances in automatic term extraction: A survey. *arXiv.* https://doi.org/10.48550/arXiv.2301.06767

Vasiliev, A., & Goryachev, A. (2022). Application of text mining technology to solve project management problems. *2022 XXV International Conference on Soft Computing and Measurements (SCM)*, 202–205. https://doi.org/10.1109/SCM55405.2022.9794858

Virjan, D., Popescu, C. R., Pop, I., & Popescu, D. (2023). Energy transition and sustainable development at the level of the European Union. *Amfiteatru Economic Journal, Bucharest University of Economic Studies, Bucharest*, *25*(63), 429–446. https://doi.org/http://dx.doi.org/10.24818/EA/2023/63/429

Wang, L., Mohammed, A., Wang, X., & Schmidt, B. (2017). Energy-efficient robot applications towards sustainable manufacturing. *International Journal of Computer Integrated Manufacturing*, *31*(8), 692–700. https://doi.org/10.1080/0951192X.2017.1379099

Warner, K., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning, 52*(3), 326–349. https://doi.org/10.1016/j.lrp.2018.12.001

Yang, G., Tang, H., Ding, M., Sebe, N., & Ricci, E. (2021). Transformer-based attention networks for continuous pixel-wise prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 16249–16259. https://doi.org/10.48550/arXiv.2103.12091

Zhu, Y., Zhang, X., Hu, S., Xue, Y., & Liang, J. (2024). Enhancing keyphrase extraction with data augmentation using large language models. *EPJ Data Science, 13*(1), 5. https://doi.org/10.1140/epjds/s13688-024-00470-5

Dr. Tine Bertoncel is an assistant professor of business informatics at the University of Primorska's Faculty of Management. With a diverse professional background spanning data analysis, data science, and digital marketing, he brings a wealth of experience from various industries to his academic role.

Odprtokodni Transformer sistem za iskanje informacij v literaturi povezani z energetsko učinkovito robotiko

Ozadje in namen: Uporabili smo strojno učenje po modelu Hugging Face za analizo 654 povzetkov na temo energetske učinkovitosti v sistemih, nadzoru, računalništvu in robotiki.
Metode: V raziskavi so bile izbrane specifične kategorije arXiv, ki so povezane z energetsko učinkovitostjo in zajemanjem ter obdelavo povzetkov s sodobnim odprtokodnim Hugging Face keyphrase-extraction-kbir-inspec modelom za ekstrakcijo ključnih besed. Na ta način smo oblikovali povezana omrežja ključnih besed za pridobivanje relevantnih znanstvenih predpublikacij.
Rezultati: Rezultati raziskave kažejo, da sodobni odprtokodni modeli strojnega učenja iz nestrukturiranih podatkov lahko izvlečejo relevantne informacije o pomembnih temah na še vedno premalo raziskanem področju energetske učinkovitosti.
Zaključek: Prikazali smo trenutno stanje in možnosti za nadaljnje raziskovanje informacijskih sistemov za iskanje relevantnih informacij, ki lahko služijo odločevalcem kot managerski sistem zgodnjega obveščanja z uporabo sodobnih digitalnih tehnologij, ki spodbujajo okoljsko trajnost in izboljšujejo energetsko učinkovitost.

Ključne besede: *Energetska učinkovitost, Ekstrakcija ključnih besed, Sistemi zgodnjega obveščanja, Informacijski sistem, Semantično omrežje, Transformerji, Industrija 4.0*