Narratives That Speak AI Lingua? AI Vocabulary in Listed Companies' Annual Reports

Dagmar Monett[1], Claudia Lemke[2], Liadan Anandarajah[1] and Tom Brandherm[1]
[1]Computer Science, Berlin School of Economics and Law, Germany
[2]Business Information Systems, Berlin School of Economics and Law, Germany
dagmar.monett-diaz@hwr-berlin.de
claudia.lemke@hwr-berlin.de
s_anandarajah19@stud.hwr-berlin.de
s_brandherm19@stud.hwr-berlin.de

Abstract: Narratives about intelligent artefacts have influenced both the public's imaginary and the actual development of the AI field since its foundation. Yet, in times where the field seems to be flourishing on the one hand, but rushing into an AI winter on the other, factual narratives about AI applications and advancements are more essential than ever. What is the gap between the actual capabilities of today's AI and the vocabulary used to report about them? In particular, what is the AI lingua used in official, legal documents in business? To find out, we analysed leading share index companies' annual reports from a representative fraction of the German economy (DAX 30), as a starting step in this direction. In this paper, we present a fact-based methodology for systematically assessing the true state of enterprise AI of those companies. Our initial empirical investigation covers only the annual reports of leading listed German enterprises in the DAX 30 as of May 2021 (i.e. before the DAX's expansion to 40 members). For this concrete example, we collected their annual reports from 2010 to 2020 (N=312). We then built upon previous work by extending natural language processing (NLP) algorithms we developed for these purposes. The idea is to systematically process and automatically detect the use of AI-related terminology in those annual reports. Such a terminology is part of a classification schema we introduce for differentiating concrete types of AI-related terms. We also compare different NLP libraries regarding their suitability and speculate on the reasons behind the poor performance of some of them. Furthermore, we look at relevant AI keywords and phrases, thereby conducting a human-based semantic analysis of the context – tasks that machines still cannot do effectively. We also give guidance on how to proceed in similar studies, i.e. on how to extend our methodology and the key findings to other national economies. This way, we are contributing not only to an informed perception about the state of enterprise AI, but also to filling the gap between the narratives it uses and the actual state of AI development.

Keywords: AI, AI adoption, AI narratives, annual reports, artificial intelligence, DAX 30, NLP

1. Introduction

Artificial intelligence (AI) is experiencing a boom in economy and society once again; at the same time, an exaggerated hype that promises fundamental organisational efficiency gains through automation and operational excellence, as well as an unprecedented resurge in innovation. Yet, the hype is increasingly synonymous with massive investments in both infrastructures and an AI workforce that must deal more with data (or the lack of it) and several pressing societal problems than with the actual AI techniques that are developed (Bender, 2022; Hao & Kruppa, 2022; Marcus, 2013; Vinsel and Funk, 2022). Some studies give the impression of constantly expanding application areas for AI solutions in companies (Chui et al., 2021; Rao & Greenstein, 2022), though without clarifying what the term "AI" actually means. On top of that, AI is often equated with its most prominent subfield from the last years, machine learning, which is largely seen as the sole technical solution to operational challenges—paraphrased as the *machinelearnization fallacy* in (Monett & Lemke, 2021). In spite of this, AI continues to open the doors for external investments, thereby increasing the company's image in public, also as an attractive employer, and suggests modern working methods and innovative technology use.

However, a closer look often reveals a misfit between public corporate communication about the maturity as well as variety of AI adoption, on the one hand, and the actual use or application of AI systems in various business processes, on the other hand. Among many, one of the reasons undoubtedly stems from the technology itself; for AI-based systems might require a comprehensible and sustainable problem definition together with proper use-case scenarios for their application, which is not always a given. This presupposes a clear understanding of the AI potential, the risks and limitations that might be present, as well as a strong business executive commitment, for instance. In addition, AI often relates to or is part of other technologies (like industry 5.0) which could make a transparent and realistic view of both the maturity level and actual scope of the AI landscape in companies very difficult. Nevertheless, the fact remains that both exaggerated expectations combined with hyperbole marketing communication misuse AI as a vehicle for companies' innovative capacity and modernity.

This paper looks closely at how exactly corporate communication reports about the actual level of AI adoption in financial and legal documents. The investigation is part of a project aimed to developing a fundamental basis for defining AI adoption's *technology readiness* in business. Intended is a contribution to a sustainable and robust index that reliably captures the actual potential, the existing diversity, and the range of AI applications in business. By doing so, it may provide a basis for comparison and a decision-making tool for future AI deployment sustained by a fact-based assessment of the situation in the field of enterprise AI.

Annual reports, as the flagship of a public company, seem to be particularly suitable for getting a first impression of the AI initiatives in the company. As a legal, financial document, it provides information about the company's health in the last fiscal year for stakeholders, shareholders, the media, and the interested public. The analysis of the annual reports, in our case of the German leading index DAX 30, is the subject of this paper. On the one hand, their examination in the context of AI initiatives gives an overview about how and to what extent the leading German public corporations communicate about AI. On the other, it is an ideal source for discovering and evaluating any linguistic misunderstandings in the use of the term AI. Coeckelbergh (2021) writes: "we need to scrutinize the grand narratives about AI that currently pervade AI discourse." We concur and start by asking: how do DAX 30 corporations "speak" the AI lingua and which one do they use concretely?

1.1 Related work

In addition to the fundamentally necessary balance-sheet information, the profit and loss account, and the management report, annual reports are ideal as a corporate image platform as they contain several voluntary disclosures by a company. This is why researchers are always interested in the analysis of such reports.

Bowman (1984) carried out a content analysis of annual reports to study companies' strategic decision making and discover new ideas about how companies deal with risks related to corporate strategy. Many works have subjected annual reports to fundamental information analysis, e.g. to determine communication structures about the business development of a specific company over time (Thomas 1997), to discover specific price and trading response strategies (Cready & Mynatt, 1991) or investment behaviour, as well as the handling of intellectual capital as intangible assets (Brennan, 2001). Other works investigate annual reports' design of individual graphics and images in order to get a better idea of which corporate values are conveyed through these forms of representation (David, 2011). Quintana Pelayo (2020) also analyses information contained in annual reports, thereby focusing on chances in the company regarding products and people. And Beattie et al. (2004) analyse and evaluate the narratives used in annual reports to investigate how they deal with corporate transparency and accountability. Summarised, several authors have carried out different forms of analysis and interpretations of information from annual reports, which makes it ideal for identifying information also about terms related to AI (see e.g. (Bonsón et al., 2021)).

Thus, it makes sense to analyse annual reports to look for basic AI narratives that can help with investigating whether AI is actually used in those companies and how. For this purpose, we selected the annual reports from 2010 to 2020 of the German leading index DAX 30. Germany was in fourth in the ranking of industrialised nations with the world's largest GDP in 2021 (Statista, 2022). A temporal examination of the development of AI narratives in the economically most vital corporations in Germany should provide information about how the adoption of this transformative technology in business is taking shape in Germany. We developed a natural language processing (NLP) algorithm to conduct an analysis of those annual reports regarding AI-related terminology.

2. Methodology

2.1 Research design

The following list describes the phases we considered in both the methodological approach and the corresponding research:

- Literature review: systematic literature review and synthesis of selected works on relevant concepts and topics.

- Identification of the main research objectives and definition of the research questions.
- Classification: development of a notation for the classification and storage of data from the test objects.
- Data collection: systematic data collection (DAX 30 annual reports, documents on annual, quarterly and monthly financial statements).
- Data processing: pre- and post-processing of the collected data by using suitable NLP algorithms, both from previous work and newly developed for these purposes.
- Analysis: Mining and quantitative analysis of the selected documents regarding AI-based content and defined keywords, by using the algorithms and processes from the literature as suggested in (Monett et al., 2020) and (Elsevier, 2018).
- Discussion: interpretation of the results.
- Documentation: systematic documentation of the results and research tasks.

The next sections introduce most of these research phases, briefly.

2.2 Research questions

On a general note, our study aims to investigate whether and when DAX 30 companies use AI-related terminology in their annual reports. The following fundamental questions are of interest, for instance: In which part of an annual report is AI-related terminology used? Which terminology in particular? Are there specific organisational units involved? Are concrete products and/or services mentioned, which ones in particular? Is the temporal course in the reporting similar to those common in the media coverage of AI?

In order to analyse these and similar inquiries, we subdivided our research into three major groups of research questions (RQs):

RQs group 1—Linguistic Analysis:

- When did leading German stock corporations start using AI-related terms in their annual reports?
- Has the number and variety of mentions of AI-related terminology increased with time, especially over the last ten years?
- Which concrete AI-related terminology is used and which are its characteristics?

RQs group 2—Semantic Analysis:

- In which business context are AI-related terms used?
- Are there AI-related terms that are hyped in the media also used in the annual reports and for the same purposes?
- How are they used and in which contexts?

RQs group 3—Business context:

- Is there a connection between AI-relevant terms and concrete measures such as projects or initiatives in the respective companies that have been increasingly used over the last ten years?
- Do fundamental statements of an annual report (such as the corporate vision, business model, product(s), or service portfolio) change under the influence of the current media hype of AI?

We tackle some of them in the rest of this paper.

3. Data collection and analysis

One of the first decisions made concerned the type of information about the DAX 30-listed companies, since both the freely accessible data and information from them include a huge variety of document types and formats. The choice was made to analyse, in a first exploratory analysis, the annual reports of those companies by using natural language processing techniques, with which selected information can be extracted for further automated processing. Table 1 shows the Python libraries and packages that were used when working with and analysing the annual reports. Each report was available in the form of a PDF (Portable Document Format) document, which was downloaded from the website of the corresponding DAX 30 company. These are official documents that are (compulsorily) openly accessible, as was already mentioned in Section 1. We collected all DAX 30 annual reports spanning from 2010 to 2020 (N=312) that were available as of May 2021, i.e. all the corresponding documents before the DAX's expansion to 40 members. For listed companies like Adidas, Allianz, SAP, Siemens or Volkswagen, all eleven annual reports from 11-year period were considered; however, only six in the case of Covestro, seven from Delivery Hero, and two from Siemens Energy, determined by the years they entered the DAX 30 (Chen, 2022).

Table 1: Python libraries and packages for NLP that were used

Library	Concretely used in the project for…	Version	Where to find it*
pdfplumber	extracting texts from PDF files	0.7.1	https://pypi.org/project/pdfplumber/
PyPDF2	extracting texts from PDF files	2.1.1	https://pypi.org/project/PyPDF2/
pikepdf	decrypting PDF files	5.1.5	https://pypi.org/project/pikepdf/
NLTK	tokenizing and lemmatize	3.7	https://pypi.org/project/nltk/
spaCy	tokenizing and lemmatize	3.2.0	https://pypi.org/project/spacy/
pandas	data handling and statistics	1.4.2	https://pypi.org/project/pandas/
pytrends	send search volume data request for each keyword	0.7.1	https://github.com/GeneralMills/pytrends
NumPy	insert null value when no data to be extracted	3.5.2	https://pypi.org/project/numpy/
Matplotlib	plotting	3.5.2	https://pypi.org/project/matplotlib/

*: Last accessed: 30th June 2022.

3.1 The AI vocabulary

The concrete terminology to search for was selected and refined in several rounds. First, we derived AI terms and concepts (terminology, in general) from (Monett et al., 2020) that form part of a corpus of definitions of artificial and human intelligence. The words used by experts when defining those concepts are of importance because many of them also correspond to "intelligent" capabilities, processes, and technologies well known or widely used in the AI field and, as a consequence, in enterprise AI. A semantic distinction is needed, though: sometimes there are verbs or actions that denote intelligent behaviours and that are more used in specific narratives, instead of adjectives or substantives, which may find a broader use in other contexts. Furthermore, many technical concepts are used in both the German and the English language indistinctly; other times, translations might be more common depending on the products that are developed and by whom. We also considered specific AI-related terminology already used in previous work, as well as undertook a general search for terms in the AI-related literature and media articles. A first refinement considered the following broad categories: *Processing*, *AI*, *Technology*, *Data*, and *Adjectives*, each comprising several terms that may be grouped together depending on both their meaning and their use in written language. A second round of refinements deleted duplicates and added concrete AI-based products or terms that have been included in recent AI narratives because of their popularity, for instance.

After a careful "human" analysis, we ended up with the following broad categories of AI-related terminology for further consideration:

- *Processing*, which contains 39 terms like "prediction," "monitoring," and "decision support," as well as their corresponding German translations.
- *AI synonyms*, which contains 41 terms like "artificial intelligence," "machine learning," "cognitive systems," and "neural networks," to name a few, also in both languages (like for the rest of the categories).
- *Product AI*, with 63 terms including "chatbot," "internet of things," "robots," and the like.
- *Business Process AI*, comprising 96 terms like "robotic process automation," "cybersecurity," and "pattern recognition," among others.
- *Data*, which contains 35 terms like "text mining," "image processing," "big data," and "clustering," for instance.
- *Adjectives*, comprising 56 other words also extensively used in AI narratives, like "autonomous," "augmented," "deterministic," "semantic," among others.

Arguably, there are terms that could belong to more than one category. We suggest here a concrete but flexible categorization. Overall, we deal with an AI vocabulary that contains 330 words.

3.2 Detecting AI terminology automatically

Which AI terminology to look for certainly depends on the goals of the researchers that analyse the data. Not all DAX 30-listed companies belong to the same branch, though. Furthermore, it is expected that the AI terminology that is used by technology-oriented companies significantly differs from the one that is used in other branches. Despite the variety that its composition might have, we think that generalizing such an analysis is a legit decision for an initial exploratory approach. We thus make no distinction depending on the branch in this paper, nor assume that the results should apply to other stock exchanges. We suggest it could be done similarly, though. The analysis of the 312 collected annual reports was systematically carried out as follows:

- Parsing the PDF documents when searching for the 330 AI-related terms, extracting relevant information, and exporting the results to corresponding TXT files (by using `pikepdf`, `pdfplumber` and `PyPDF2`). This analysis took almost 8 hours to complete on a Dell Latitude 2018 laptop running a Windows 11 operating system with 16 GB of RAM.
- Lemmatization and tokenization of terms with NLP and export of results to a CSV file (by using `spaCy` and `NLTK`, depending on the concrete comparison). This pre-processing took almost 22 hours to complete (on the same hardware as in a.).
- Frequency analysis according to given keywords and previously created tokens.
- Calculation of statistics and metrics (e.g. mean, standard deviation, etc.) in order to compare different technologies for text extraction and NLP (by using `pandas`).
- Visualization of results (with `Matplotlib`).
- Frequency analysis with Google Trends and Keywords Everywhere for the case of exemplar terms.

The reader can find source code in Python, data, and several visualizations available on a repository called *DAX30-Analyser* on GitHub (see https://github.com/tombo92/DAX30-Analyser.git) which comprises most of the work done in the first phases of the analysis. For the last phase involving a comparison with trending topics and terms, there is also a repository available named *DAX30-GoogleTrends* on GitHub (see https://github.com/liadanxa/DAX30-GoogleTrends.git). The pseudocode from Figure 1 summarizes the whole process. Four popular NLP python libraries (gensim, NLTK, spaCy and textblob), as well as a further four PDF extractor python libraries (pdfplumber, pdfrw, pdftotext and PyPDF2) were initially considered to facilitate the DAX 30 analysis. Upon further inspection of the capabilities, speed, memory usage, and general suitability with regard to the required functionality of the analysis program, the libraries were however reduced to NLTK, spaCy, pdfplumber and PyPDF2. This decision was made according to the information presented in official library documentations or specific comparison articles published by other developers (see Table 1 and related comments below).

```
GET keywords FROM previous research papers
SET categories                                        // to encompass and organize keywords
FOR EACH keyword IN list DO
    IF keyword translation or variations NOT IN list THEN // for nouns, verbs, etc.
       PUSH new keyword TO list
    IF keyword IN incorrect category THEN
       SET keyword TO correct category
READ pdf-data FROM accumulated DAX 30 annual reports
SET pdf-data IN txt-files                             // pikepdf + pdfplumber or PyPDF2
results:= lemmatized AND tokenized txt-data           // spaCy or NLTK
SET results IN csv. files
FOR results IN EACH csv. file DO
    FOR EACH keyword IN list DO
        COUNT keyword IN results
    PLOT keyword totals                               // Matplotlib
CALCULATE category means AND standard deviations
PLOT category averages                                // Matplotlib
GET bulk trend data of keywords with Keywords Everywhere
FOR EACH category or year in bulk trend data DO
    CALCULATE mean totals
FOR EACH keyword IN list DO
    REQUEST Google Trends general data
    REQUEST Google Trends AI-specific data
    IF AI-specific data better corresponds to Keywords Everywhere data THEN
       DELETE general data
    FOR EACH category AND year in Google Trends data DO
        CALCULATE mean normalisations
PLOT Keywords Everywhere AND Google Trends averages comparisons
```

Figure 1: Process of analysing the DAX 30 annual reports

Some comments regarding Python libraries for extracting information from PDF documents follow: we made a first evaluation and took a closer look at `pdfplumber` and `PyPDF2`, two of the most frequently used libraries by practitioners for these purposes. The former, `pdfplumber`, allows for searching detailed information about each character, even from graphics, as well table extraction.[1] The latter, `PyPDF2`, can retrieve text and metadata from PDFs and even merge entire files;[2] however, and depending on the PDF structure, extracted data could be lost, as discussed in (PDFExtractor, 2022). We compare the results of using both libraries in the next section, as well as compare other libraries for NLP tasks.

4. Results and discussion

Figure 2 shows the values of the averaged occurrences including the calculated standard deviation of each category (see Section 3.2) for all analysed companies and their corresponding annual reports over time.

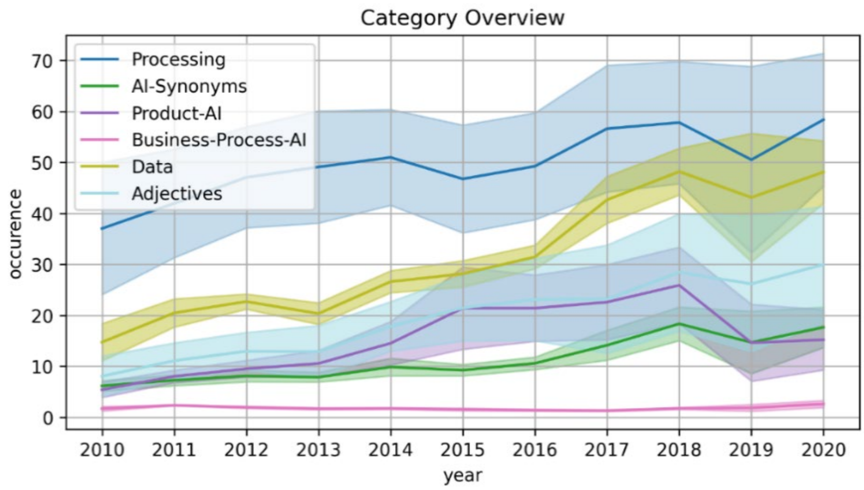

Figure 2: Overview of the words occurrences in each category for all the DAX 30 annual reports together with their range, over time

Terms belonging to the category *Processing* are the most used in the documents, followed by the category *Data*. All categories experiment an increase of the AI-related terminology over time (which supports the first group of research questions introduced in Section 2.2). The coloured areas show the ranges of the averaged values for each category when considering the standard deviations.

By analysing the data in detail, we observed that, as expected, companies from technology branches use much more AI-related terms, in general. For example, from 2015 onwards, SAP used almost 100 times more AI-related words than all other listed companies did together. The maximum values for most of the categories remain stable compared to the ones from previous years. However, the minimum occurrences decreased notably in 2019. This might be explained by a change in format for the annual reports published by certain companies. We see a sizable dip in terms of minimum values particularly for keywords pertaining to the category *Processing*. The smallest returned mean value of occurrences for this category and year is 1 from the company Linde (the highest being 120 from Daimler). Linde's mean occurrences from 2019 compared to its previous year (20.5) and consequent year (10.5) exhibits a major jump. After referring to Linde's annual reports for 2018, 2019 and 2020 to investigate what may be causing this, we found a potential reason: the annual reports for 2018 and 2020 amassed a total of 147 and 134 pages, respectively. These reports were also published in the form of a text document. However, the annual report for 2019, whilst still in PDF form, depicted its information in more of a slide-show/PowerPoint format with a higher graphic-to-text ratio in the scope of only 28 pages. This may explain the greater lower-bound variation; simply a lack of data for this year for our NLP algorithm to work with.

We observe the lowest percentage of mentions for *Business Process AI*. The purpose of this category is to record AI-relevant terms that are directly related to the possibilities of applying such systems in operational automation, i.e. which ultimately lead to process optimisation or renewal. Different explanations for this low rate are possible. On the one hand, it could be due to a low integrative operational use of concrete AI solutions.

[1] See pdfplumber's project description at https://pypi.org/project/pdfplumber/ for more.
[2] See PyPDF2's project description at https://pypi.org/project/PyPDF2/ for more.

On the other hand, AI solutions might be part of larger business information systems that contribute to automation but that do not necessarily run under the name "AI." This bears the question of whether such an AI lingua is adequate in those official contexts. We pose that writing about AI in the media and other unofficial contexts is perhaps less caring about the actual vocabulary that reflects reality. Or is the AI maturity level of the analysed companies the one that needs improvement? For giving an informed answer, it would be interesting to consider similar data of other stock exchanges and/or world regions, something we do not refer to in this paper for time and scope reasons.

How good are the NLP tools that are used for searching and processing the content of the annual reports? Are there significant differences in their results? We compared the processing using the libraries `NLTK` and `spaCy` for NLP and depending on the libraries for working with PDF files. Especially, we analysed the results of using `PyPDF2` and `pdfplumber`. The results for the particular case of one company (Allianz) are depicted in Figure 3.

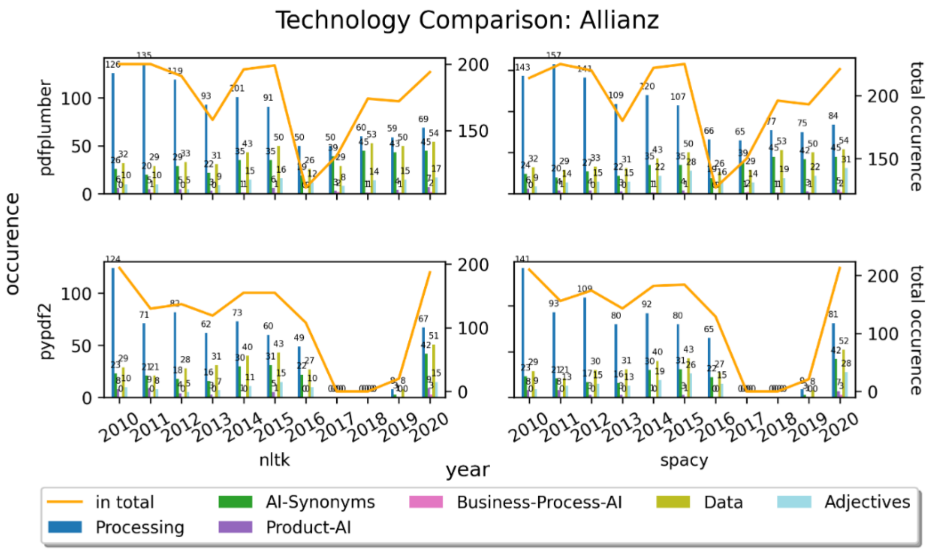

Figure 3: Results when using different NLP and PDF-file-processing libraries when analysing the Allianz's annual reports

Notice how all depicted values are fundamentally different. Which library should we trust then? The library `PyPDF2` is less robust than `pdfplumber`. There are even two years (2016 and 2017) where no results (i.e. occurrences of any of the 330 AI-related terms) are found. This is suspicious and might be explained by a problem of the library (or the actual functions that are used) when dealing with certain types of formats of the PDF files. We did not explore this possible cause in detail. Notice, however, how it is `pdfplumber` the one that might have troubles with documents from 2015 instead!

Both `NLTK` and `spaCy` seem to be very robust in general (compare the orange curves from the left-hand side with those from the right-hand side: they are almost the same). Processing the PDF files is what makes the difference, and not applying the NLP techniques like the lemmatization and tokenization of terms. Nonetheless, we suggest considering the pros and cons of both NLP libraries, especially "the fastest and most accurate syntactic analysis of any NLP library released to date" in the case of `spaCy` despite its inefficient memory usage (see (ActiveState, 2021) for a detailed comparison). In our experiments, `NLTK` was much slower.

Finally, we also analyse the general frequency of some terms, which have been used over the years by accessing both Google Trends' and Keywords Everywhere's APIs. Figure 4 shows a concrete comparison for the case of the category *AI Synonyms*.

Keywords Everywhere totals were calculated by adding the values from January to December for each year for every keyword (i.e. the 330 words) in the list. The totals of each keyword in every category were then added together per year, where data for a keyword was available (some keywords returned null values). Google Trends normalisations were calculated by finding the average value from January to December for each keyword per year. Each of these averages were used to create an average normalisation for each keyword value per year, where data for a keyword was available (some keywords returned here null values, too).

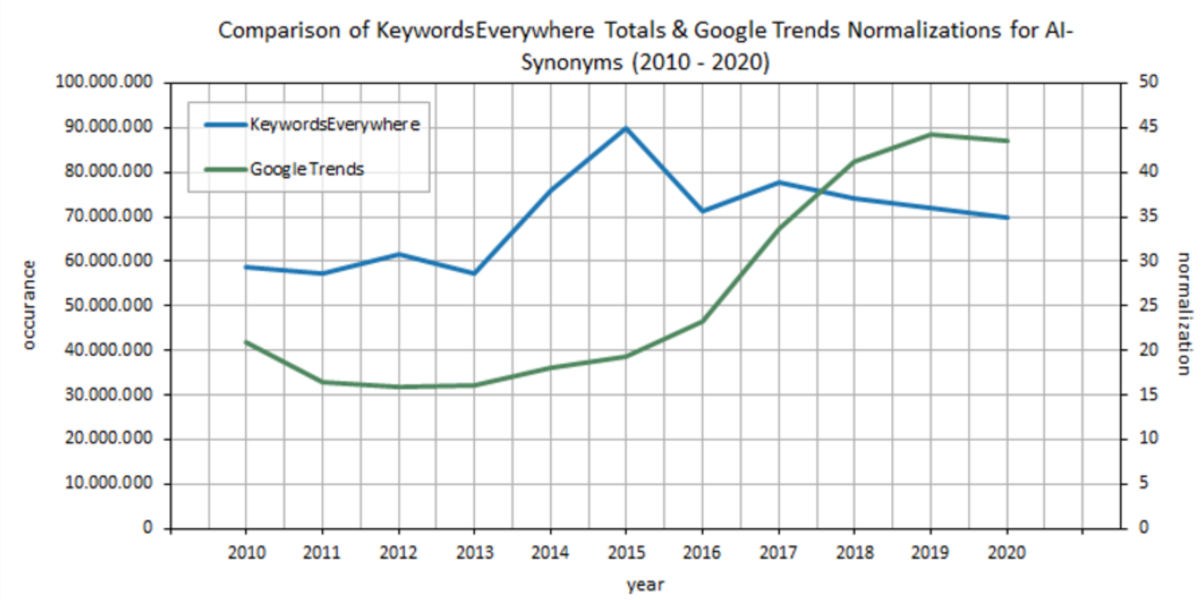

Figure 4: Keywords Everywhere and Google Trends occurrences for the category *AI Synonyms* over the years

The Google Trends normalisations, both from the general and AI-specific requests are then compared to the Keywords Everywhere totals. Although the Keywords Everywhere data is unfiltered by category, the corresponding AI-specific Google Trends data shared the most similarities pertaining to where null values were produced for certain keywords. For this reason, the AI-specific Google Trends data has been taken for the Google Trends and Keywords Everywhere comparison.

Notice how different the results are (i.e. the number of occurrences) when comparing the words commonly used to denote AI algorithms and techniques like the ones included in the category *AI Synonyms*. Both curves increase with time; however, the totals (after normalizations and filtering) differ notably. Comparing Figure 4 with the *AI Synonyms* data of Figure 2, we can see that the Google Trends normalisations shows a similar trend over time up until 2019. This could consolidate our previous claim that the algorithm simply found insufficient data for this year, rather than suggesting a general lack of interest or investment in AI during this time. This correlation may also imply that the increasing implementation of AI (or at least the reporting thereof) by the DAX 30 companies is a direct by-product of the growing general interest in AI as seen by the corresponding search trends at the time.

The DAX 30 annual reports are published by German companies for the German public and are therefore written in German. While it is true that many AI terms were originally coined in the English language, many have become a global standard needing no translation, like “Internet of Things,” for instance. Despite these terms being popularly used in English, there are still cases where a literal translation into a country’s native language has become more and more standardised in that country’s AI terminology. To ensure all versions of a term are caught by the algorithm, each English term in the keywords list was awarded its own German translation. Although this decision proved necessary in most cases, as results were found for many of the German counterparts by the algorithm, the majority of the null values returned on keywords by the Google Trends and Keywords Everywhere API (meaning there was insufficient data on these terms) were German translations like “Wissensextraktion,” meaning “knowledge extraction.”

5. Conclusions

When assessing or determining the opportunities and challenges of AI in enterprise, it is crucial to have an understanding of the reality of AI adoption in that context. Such a reality, when legal aspects are involved, is carefully narrated, made sense of (Coeckelbergh, 2021) in a way that might be different from the vocabulary used in other contexts. Demystifying AI in the public sphere, what experts have already been calling out extensively, may consider *actual* narratives that provide an invaluable evidence of what does not go easily hyped.

There still remains much to be investigated, e.g. regarding the research questions introduced in Section 2.2. However, and although a preliminary, exploratory study aimed to analyse the use of AI-related terminology in the annual reports of DAX 30 companies over a limited period, the methodology, the methods, and the results presented in this paper could be extended to other branches, types of documents, and analyses with ease. As such, it could provide an informed support to how the state of enterprise AI is perceived, thereby filling the gap between the narratives that are used and the actual state of AI development.

5.1 Limitations of our work

Whilst the compiled list of keywords are all relevant to the AI field, particularly in the context of their allocated categories, many words in our constructed vocabulary might have a different meaning in other business and enterprise contexts. Examples of these are "decide" in the *Processing* category, "model" in *AI Synonyms*, "agent" in *Product AI*, "cyber security" in *Business Process AI*, "data" in *Data*, or "digital" in *Adjectives*. We do not dive into details in this respect.

Nonetheless, a brief manual context analysis on the 2019 Allianz report was conducted by using the search function for each keyword, manually determining the context under which it was used and only counting it if it was referenced in an AI context. The totals indicated that the category *AI Synonyms* returned the highest number of positive results in terms of keywords found in the context of AI. Yet, the values for 2019 in Figure 3 demonstrate *AI Synonyms* to have only the third highest results out of all the categories. For example, the results created with `pdfplumber` and `spaCy` show totals of 50 for *Data* and 75 for *Processing*, but only 42 for *AI Synonyms* in comparison.

Due to the highly time-consuming nature of this kind of analysis, a full investigation of all 312 annual reports was not possible. It is also important here to note that even with this method, it is possible for keywords to have been ignored, e.g. in the case where words are displayed as images.

Acknowledgements

We acknowledge the support of Jessica Uszpelkat and Konstantin Köhler who contributed with the data collection (of the DAX 30 annual reports).

References

ActiveState (2021) Natural Language Processing: NLTK vs SpaCy. ActiveState, https://www.activestate.com/blog/natural-language-processing-nltk-vs-spacy/ (Accessed: 20.06.2022).

Beattie, V., McInnes, B., and Fearnley, S. (2004) A Methodology for Analysing and Evaluating Narratives in Annual Reports: A Comprehensive Descriptive Profile and Metrics for Disclosure Quality Attributes. *Accounting Forum* 28 (3): 205-236.

Bender, E. M. (2022) Look behind the curtain: Don't be dazzled by claims of 'artificial intelligence.' The Seattle Times. https://www.seattletimes.com/opinion/look-behind-the-curtain-dont-be-dazzled-by-claims-of-artificial-intelligence/ (Accessed: 29.06.2022).

Bonsón, E., Alejo, V., and Lavorato, D. (2021) Artificial Intelligence Disclosure in the Annual Reports of Spanish IBEX-35 Companies (2018–2019). In Kumar, V., Rezaei, J., Akberdina, V., Kuzmin, E. (eds.), *Digital Transformation in Industry*. Lecture Notes in Information Systems and Organisation 44, Springer, Cham.

Bowman, E. H. (1984) Content Analysis of Annual Reports for Corporate Strategy and Risk. *Interfaces*, 14(1): 61-71.

Brennan, N. (2001) Reporting intellectual capital in annual reports: evidence from Ireland. *Accounting, Auditing & Accountability Journal* 14(4): 423-436.

Chen, J. (2022) DAX Stock Index. Investopedia. https://www.investopedia.com/terms/d/dax.asp (Accessed: 29.06.2022).

Chui, M., Hall, B., Singla, A., and Sukharevsky, A. (2021) The state of AI in 2021. McKinsey & Company. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021 (Accessed: 25.06.2022).

Coeckelbergh, M. (2021) Narrative responsibility and artificial intelligence: How AI challenges human responsibility and sense-making. *AI & Society*. https://doi.org/10.1007/s00146-021-01375-x (Accessed: 30.06.2022).

Cready, W. M. and Mynatt, P. G. (1991) The Information Content of Annual Reports: A Price and Trading Response Analysis. *The Accounting Review* 66(2): 291-312.

David, C. (2001) Mythmaking in Annual Reports. *Journal of Business and Technical Communication* 15(2): 195-222.

Elsevier (2018) Artificial intelligence: How knowledge is created, transferred, and used. Trends in China, Europe, and the United States. https://bit.ly/2Lajwaa (Accessed: 31.03.2021).

Hao, K. and Kruppa, M. (2022) Tech Giants Pour Billions Into AI, but Hype Doesn't Always Match Reality. The Wall Street Journal. https://www.wsj.com/articles/tech-giants-pour-billions-into-ai-but-hype-doesnt-always-match-reality-11656508394 (Accessed: 29.06.2022).

Marcus, G. (2013) Hyping Artificial Intelligence yet again. The New Yorker. https://www.newyorker.com/tech/annals-of-technology/hyping-artificial-intelligence-yet-again (Accessed: 29.06.2022).

Monett, D., Lampe, N., Ehrlicher-Schmidt, M., & Bewer, N. (2020) Intelligence Catalog-guided Tracking of the Evolution of (machine) Intelligence: Preliminary results. In Basile, P., Basile, V., Cabrio, E., & Croce, D. (eds.), Proceedings of the 4th Workshop on Natural Language for Artificial Intelligence, NL4AI 2020, 2735: 118-129, CEUR-WS.

Monett, D. and Lemke, C. (2021) AI-ware: Bridging AI and Software Engineering for responsible and sustainable intelligent artefacts. In van Giffen, B., Koehler, J., Brenner, W., & Albayrak, C. A. (eds.), Managing Artificial Intelligence (pp. 66-73), Workshop Paper Series, INFORMATIK 2021, Institute of Information Management, University of St. Gallen.

PDFExtractor (2022) Using Python for Data Extraction from PDFs. PDFExtractor.org. https://pdfextractor.org/python-for-data-extraction (Accessed: 20.06.2022).

Quintana Pelayo, G. (2020) NLP approach to Annual Reports Analysis. https://guilleqp.github.io/papers/IRTM_Project_Report_GuillermoQuintana.pdf (Accessed: 25.06.2022).

Rao, A.; Greenstein, B. (2022) PwC 2022 AI Business Survey. PwC. https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-business-survey.html (Accessed: 25.06.2022).

Statista (2022) Ranking der 20 Länder mit dem größten Bruttoinlandsprodukt (BIP) im Jahr 2021. Statista. https://de.statista.com/statistik/daten/studie/157841/umfrage/ranking-der-20-laender-mit-dem-groessten-bruttoinlandsprodukt/ (Accessed: 25.06.2022).

Thomas, J. (1997) Discourse in the Marketplace: The Making and Meaning of Annual Reports. *The Journal of Business Communication*, 34(1): 47-66.

Vinsel, L. and Funk, J. (2022) Blinded by the Hype. Open Mind. https://www.openmindmag.org/articles/blinded-by-the-hype (Accessed: 29.06.2022).