ITEA 3

INNOSALE

Innovating Sales and Planning of Complex Industrial Products
Exploiting Artificial Intelligence

Deliverable 1.3
State of the Art description

Deliverable type:	Document
Deliverable reference number:	ITEA 20054 | D1.3
Related Work Package:	WP 1
Due date:	2022-09-30
Actual submission date:	2022-09-29
Responsible organisation:	VTT
Editor:	Sari Järvinen
Dissemination level:	Public
Revision:	FINAL Version 1.0

Abstract:	The deliverable presents the technology state-of-the-art from industrial and scientific perspectives. The literature review provides a view on current level of adoption of AI methods in B2B sales process as well as studies the state-of-the-art of different InnoSale related technologies.
Keywords:	literature review, state-of-the-art technologies, B2B sales processes

Table_head	Name 1 (partner)	Name 2 (partner)	Approval date (1 / 2)
Approval at WP level	Bilge Özdemir (ERSTE)	Yazmin Andrea Pabon (UC3M)	14 September 2022
Veto Review by All Partners			29 September 2022

Editor

Sari Järvinen (VTT)

Contributors

Marko Jurvansuu (VTT)

Satu-Marja Mäkelä (VTT)

Tuomas Sormunen (VTT)

Erkki Siira (VTT)

Pablo Clemente Maseda (IOTIQ)

Mario Thron (IFAK)

Alex Ivliev (TUD)

Lukas Gerlach (TUD)

Afnan Sultan (NATIF)

Nico Herbig (NATIF)

Klaus Hanisch (TARAKOS)

Executive Summary

In this deliverable, we present the technology state-of-the-art regarding InnoSale from industrial and scientific perspectives. First, we study the Business-to-Business (B2B) sales process along the sales funnel and focus on the adoption and opportunities related to Artificial Intelligence (AI) -driven solutions. Next, we consider different issues regarding system design and data acquisition, both of which are of utmost importance to the design of the InnoSale platform. The platform should be modular and able to interface with internal and external systems, providing the data to be used in the core of the InnoSale solutions. AI methods for semantic search, named entity recognition, customer segmentation and pricing are studied to identify solutions that enable knowledge construction, which could either support the sales experts in their work or be used to guide the functioning of a product configurator. User interfaces have an important role as the aim is to ease the work of the sales experts. The deliverable also reviews different ways to interact with the users to acquire and deliver information. Finally, the progress beyond the current state-of-the-art is discussed and concluding remarks are made.

Table of Content

1 Introduction ... 1

2 AI-driven tools for B2B sales processes along the sales funnel ... 2

2.1 Future of the AI tools in sales ... 6

2.2 Tools and solutions for customer segmentation ... 7

2.2.1 Customer Segmentation in major CRM systems ... 9

3 System design and data acquisition ... 12

3.1 Internal organizational systems and data interoperability ... 12

3.2 External data sources and data acquisition for customer segmentation ... 14

3.3 Solutions for data integration ... 15

4 AI methods for knowledge modelling and data analysis ... 18

4.1 Solutions for knowledge base design and implementation ... 18

4.2 Practical approaches to Existential Rule-Reasoning and possible integration of answer set programming (ASP) for optimization problems ... 21

4.3 Named entity recognition approaches ... 22

4.4 Semantic search ... 24

4.4.1 Semantic search with modern deep learning -based NLP techniques ... 24

4.4.2 Market analysis ... 26

4.4.3 Industrial state-of-the-art ... 27

4.5 Customer segmentation approaches ... 28

4.6 Fuzzy Logic based pricing methods ... 30

5 User knowledge acquisition and interaction ... 32

5.1 UI development approaches for user knowledge acquisition and user dialog ... 32

5.1.1 Frontend frameworks ... 32

5.1.2 E-commerce frontend component libraries ... 33

5.1.3 Progressive Web Apps (PWAs) ... 33

5.1.4 Micro Frontends ... 33

5.1.5 Configure, Price, Quote ... 34

5.2 Solutions for capturing user needs and providing recommendations ... 35

5.3 Tools for planning and configuring products ... 36

6 InnoSale contribution beyond state-of-the-art ... 39

7 Conclusion ... 40

8 Abbreviations ... 41

9 References ... 43

Figures

Figure 1: Tactical B2B customer journey map. 8
Figure 2: Examples of B2B customer segmentation methods. 9
Figure 3 The AI features of Power BI. 10
Figure 4. Reference Architecture Model Industry 4.0 (RAMI4.0) architecture. 13
Figure 5. Magic Quadrant for Enterprise Integration Platforms as of September 2021. 17
Figure 6. The Forrester Wave: Enterprise Integration Platforms as of Q4 2021. 17
Figure 7. Magic Quadrant for CPQ Applications Suites as of June 2021. 35
Figure 8. DEMAG KBK-configurator 37
Figure 9. E-catalog from 3Dfind.it. 37
Figure 10. Result of Interrolls Layouter (AutoCAD addOn). 38

Tables

Table 1. B2B customer segmentation research articles and their details. In the strategy column, the different segmentation frameworks are abbreviated: F: Firmographics, N: Needs-based, B: Behaviour-based, P: Profitability-based, and O: Other. 30

1 Introduction

The purpose of this deliverable is to provide a review on the adoption of AI-driven solutions to improve the performance and customer experience of B2B sales processes. The state-of-the-art literature review is based on peer-reviewed international journal and conference articles, magazines, as well as trends monitored on the market. The goal of the deliverable is to support the development of the overall InnoSale platform, guaranteeing the innovation and beyond state-of-the-art nature of the project. With this aim in mind, we have included reviews on the essential technology building blocks to this work.

First, we have studied the status of AI technology adoption along the sales funnel from prospecting to pre-approach and approach, presentation, overcoming objections, closing and finally follow-up. This definition of the sales funnel follows the approach presented by Dubinsky [8]. Special attention is given to solutions that increase customer understanding via customer segmentation. We aim at identifying the potential related to AI usage from different perspectives.

Second, as the adoption of AI technologies relies on access to relevant data sources and successful deployment of new technologies and services requires integration into the existing operational environments, we have studied and discussed different ways to interface and share data with the existing internal and external services and systems.

As we first have identified the need for AI methods and second assured access to data, we then focus on different AI methods that support the development of novel services for B2B sales processes. We have investigated solutions for e.g. knowledge base design and implementation, Natural Language Processing (NLP) methods related Named Entity Recognition (NER) and semantic search as well as solutions for customer segmentation.

Finally, the review focuses on interaction with the user from perspectives of knowledge acquisition and information delivery including visualization solutions. Interaction with the service user transforms the data into knowledge and actionable insight.

Based on the findings in the review sections we define the InnoSale contribution beyond the state-of-the-art regarding the usage of AI methods in the B2B sales processes.

2 AI-driven tools for B2B sales processes along the sales funnel

Automation and increasing interconnectivity of systems related to business operations are transforming sales activities, but affecting also other parts of operations. The current transformation is referred to as a fourth industrial revolution, and for sales, it means AI-powered tools [1]. Cuevas [2] presents three main drivers for this: 1) new buyer behaviour and rising customer requirements; 2) new information and communication technologies; and 3) globalization, concentration and competition of the business. Technology development has enabled new ways of working as well as data-driven tools for sales personnel. Earlier customer insight was often based only on salespersons' knowledge, gained with work experience; nowadays there are different internal and external data sources for companies to utilise in order to achieve customer insight.

AI-driven tools play a crucial role in analysing large amounts of heterogeneous data to provide valuable insight and support in decision-making for the sales process. The purpose of these tools is to serve the customer better and to enhance the customer experience, but also to increase the efficiency and reduce the costs of the process itself. In addition to providing new tools for optimizing the sales funnel, AI can be used to provide feedback to the salesperson for enhancing their adaptability and performance [3].

To gain benefit from the new digital tools in B2B sales organisations there is a need to change the sales culture and unlearn existing routines in order to avoid their becoming barriers for change [4]. It is also very important to educate the sales force to understand the functionalities, benefits and drawbacks of AI technologies [7] [5]. To enable the change, implementing AI technologies is not enough; the personnel need education, and management must support the change. Currently, the major technological change concerns automation of tedious and repetitive sales tasks. The future value of AI is considered to be in its capability to learn from sales activities and even perform actions autonomously [14]. In an interview with Danfoss managers, AI was seen to provide tools for process efficiency, forecast and pipeline, as well as for better pricing and temporal optimization of sales processes [5].

Customer Relation Management (CRM) systems are widely used in companies to track interactions and collect information from customers. CRMs contain a lot of data on customer relations and their behaviour. Therefore, CRM systems can be extended to support data-driven decision-making and marketing strategies with the utilisation of AI and to increase the company's performance concerning customers. In [6], an extensive review of AI in conjunction with CRM in digital B2B marketing strategies can be found. For developing an analytical CRM, the study identified the following areas where AI has a future impact: decision-making strategies, user behavioural response, innovation strategies, sales forecasting, usage of social media and customer orientation [6].

Presented below are examples and visions from the current research on the usage of AI along the sales funnel. In this review, we have adopted the sales funnel described by Dubinsky [8], which is widely used in the literature.

Prospecting
In this phase searching of potential customers and evaluating their buying tendencies takes place. New leads are generated and qualified as a prospect, i.e. a qualified contact who is selected for further sales process.

One of the greatest potentials of AI tools is seen in utilising social media and other internet sources with structured and unstructured (big) data for harvesting possible new customers. In [1], it is viewed that the prospecting phase will get help from an STP (segmentation, targeting, and positioning of customers) framework, specifically as it relates to the sales function and big data sources.

In lead generation, AI tools can help find potential customers automatically [5],[1]. Another important phase is lead qualification to select the best leads for B2B sales. The benefits of AI lies in building rich prospect profiles utilising content from social media, emails, blogs and websites [11]. In the future, *predictive* lead qualification that results in high-profit sales with high probability and *machine learning (ML)-based lead qualification models* that automatically search for new prospects will be utilised [11]. Demand estimation and sales forecasting (estimation of customer demand and profitability) can be done with NLP methods. The keywords identified in speech and emails of potential customers can be used to predict the probability of these consumers purchasing the product, and in lead qualification to predict the company's buying behaviour [1].

Regarding B2B sales, small-to-medium enterprises (SMEs) in the service sector prefer social media for finding new business opportunities. Social media tools are of most benefit when used along both digital and traditional sales communications [9], [1]. Social media usage is also found to increase sales performance when it is used for value-oriented prospecting and proactive servicing [10].

AI can be used in analysing the needs of prospects [14], demand estimation, and sales forecasting within customer validation (i.e., estimation of customer demand and profitability). On the strategic level, AI can support marketing strategies and long-lasting customer relationships along CRM [6].

Pre-approach and approach
In these phases, the sales organisation will deepen the understanding of the prospect and contact the prospect to initiate and build a relationship.

As the prospect is contacted possibly for the first time, AI can automate the process with digital agents e.g. chat bots, automated emails or targeted ads [5], [11]. AI-assisted tools will provide more value for targeting and retargeting, personalized and customized communication messages and channels, but also aid in content curation. The role of the sales professional is to guide and monitor the output of AI tools as well as look for appropriate timing for contacting [11]. In the future, this can be also automated by using chat bots with well-developed NLP features as well as with tools for timing estimation and predictive analytics [1].

Historical information of a customer's past behaviour can be applied for customer behaviour understanding [5]. This is also an important case for the AI tools to provide support, as the sales personnel can change, and their tacit knowledge rarely transfers within organisations.

Presentation
In this phase, the product is presented and problem-solving characteristics of the product are showcased. In the InnoSale context, configurators belong in this phase.

Configurators are considered as sales force automation (SFA) tools for the salesperson to fulfil the customer needs better [12]. Sales configurators are often used with complex products for fulfilling customization needs. Configurators facilitate the salespersons' communication with the customer and help understand the needs of the customer better.

There are many known benefits to configurators. They enable shorter lead times and improved product specifications. Product knowledge is also systematically saved when configurators are used. They can improve pricing accuracy and profitability with better specification, reduce routine work in the design and documentation phase, and generate better customer experience and satisfaction [13].

Configurator usage is moving toward co-design with the customer and more emphasis is given for the role of the customer as a user. In that case, it has been shown that the ease of use and adaptability are important factors in perceived effectiveness, and eventually impact the perceived usefulness of sales configurators. The user experience of the configurator cannot be underestimated as the perceived enjoyment has the most significant effect on perceived usefulness.[15]

Configurators are also seen as a bigger part of the business models. The success of the configurators depends on many things, including "*how tightly the configurator is integrated into the broader business model, user experience and whether companies leverage behavioral data to further innovate and improve their value propositions" [16].*

The more integrated use of product configurators and utilisation of the data they produce in business operations and in sales funnel can provide new opportunities to improve operations. Providing better functionalities even for the customer to use will leverage the opportunity for an improved (and even automatic) customer understanding. This can lead to improved customer service and may lead to a higher and faster pace of closing deals.

There is also potential with new AI-enabled (rapid) prototyping tools for creating the satisfying product with the customer [1] [11]. In the presentation phase, it is very important to show the customer as realistic a view of the end product as possible. Technologies such as augmented reality (AR) [1] [14], virtual reality (VR) and 360-degree video provide the salesperson tools to present on-site demos that are close to reality.

In the presentation phase, AI bots are seen as tools to explain complex products as they have endless patience [1] . They can in the future facilitate the on-the-go sales simulations and give feedback from the effectiveness of the sales presentation for the salesperson. [14]

Overcoming objections
In this phase, the negotiations of the terms of the purchase will continue. The important part is to answer to the questions and objections that follow from the presentation phase. The goal is to do the presentation so well that it would minimize these. The new AI-driven tools for configurators, co-design and rapid prototyping can provide solutions for shortening this phase.

Many of the objections at this phase relate to pricing. AI can provide new tools for dynamic pricing [11], value-based pricing and estimation and calculation of reservation price [14] (minimum price for the seller, the maximum that the buyer would pay).

At a strategic level, AI can provide value by curating competitive intelligence to support a company's value proposition and beat competitors [11]. In interaction with the customer, AI emotion recognition of non-verbal cues is envisioned to interpret potential customer reactions to avoid objections [11], [1].

Closing
In this phase, an agreeable contract is achieved.

Sales representative connection with customer at this stage is very important and it is not seen that this could very easily be replaced, e.g. by a chat bot. The AI tools may have an advisor role to support overcoming the objections, design strategies and notice abnormalities or anything that may cause problems in the sales process using e.g. data from past negotiations [1].

It is interesting to understand, that there is a vision to use AI tools by counterparts (customers) to gain better insight from suppliers and sourcing opportunities. AI systems could facilitate difficult negotiations for them with multiple decision criteria (e.g. delivery times, guarantees, prices, quantities) as well as quality and budget constraints [17]. In the future, this can lead to interesting situations if two AI systems are trying to optimise the process and achieve an equilibrium.

Follow-up
In this phase, the agreed contract is fulfilled and possible prospecting future needs, up- and cross-selling is performed and ultimately a long-term partnership is formed.

The AI can help by automating workflows of the order, introducing post-orders and follow-up services with, e.g., chatbots, [5], [11]. The follow-up process provides an opportunity to B2B sales for predicting spare part sales [20]. There is also the potential to evaluate old customers and analyse future needs of products and services from recent purchase behaviour, as well as estimate churn probability [14]. The AI tools can also create models for up-sale, cross-sale and aftersales opportunities. When the business moves from plain product sales towards servitisation, AI could also be part of providing and predicting the needs of services.

For creating long lasting customer relations the AI tools can provide value for building rich customer profiles (structured and unstructured data) to uncover new needs and utilise past purchase behaviour data [11] by using partly similar tools as in the prospect phase, but also use internal data from existing sales processes.

2.1 Future of the AI tools in sales

AI development might result in some of the tasks of today's salespersons to disappear. The sales profession can shift from sales actions to consulting, and this could happen due to the increased automation level of sales tasks [14]. In any case, it is quite clear that interaction is important in sales; in [5] it was discussed that the use of AI can diminish rational tasks, but does not replace human interaction at least in the short term. The role of AI tools is seen as assisting with the process but not as a decision maker.

In [14] three main questions are raised to be still resolved in the use of AI in the sales process: 1. Empirical studies on what areas AI technology is effective; 2. How these technologies can be integrated into the sales profession; and 3. Ethical implications of using AI.

The topics of ethical implications and bias in AI have recently gained a lot of attention. Machine Learning or AI Bias can be defined as "*systematic and repeatable errors in computer systems that create unfair outcomes*[1]. There already exist real-life examples of embedded bias that are quite famous in their erroneous behaviour. Arguably, the most notable example of AI bias is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) that is a tool used in the United States for many jurisdictions to predict recidivism. It was found that even if the tool did not feature the race of a person as an input for AI, the bias nonetheless existed[2,3]. Another interesting example is Amazon's hiring tool experiment which was designed to automate a recruiting process. By mistake, the system had a built-in bias for discriminating against women.[4]

As the systems are designed by humans, there are many reasons why and when bias is implemented into the algorithms. One of the most discussed reasons is the low quality of the training data for the AI models, where the training data itself already contains bias. Schwartz et al. [18] categorised bias into three main classes based on how biases are present in AI:

[1] Wikipedia Contributors. (2019, September 6). *Algorithmic bias*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Algorithmic_bias

[2] Rahman, F. (2020, September 8). *COMPAS case study: Fairness of a Machine Learning Model. Medium.* https://towardsdatascience.com/compas-case-study-fairness-of-a-machine-learning-model-f0f804108751

[3] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias.* ProPublica; ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[4] Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

1. Systemic Bias (the terms institutional and historic bias are also used). Systemic biases are formed from "*procedures and practices of particular institutions that operate in ways which result in certain social groups being advantaged or favoured and others being disadvantaged or devalued*" [18]. The universal design principles can also cause systemic bias (e.g., lack of accessibility for people with disabilities)

2. Human Bias. This bias is formed from systematic errors in human thinking. They occur throughout the AI life cycle and relate to how individuals or groups of people understand information.

3. Statistical / Computational Bias. This originates from errors that are caused by the (data) sample not representing the whole population.

All the people working with AI systems are responsible to avoid bias within them. With the fast development of AI algorithms and solutions, the confidence that these solutions can tackle complex problems with social, political, ecological, economic, and/or ethical dimensions has increased [18], but at the same time, the use of AI in complex problems outlines the bias and importance of finding solutions to avoid it. In addition, the developers and teams working on the AI problems often do not have an accurate understanding of how the technology will be used, which can also cause problems in the design and implementation of the solution.

Understanding the different mechanisms of bias helps the AI community build tools to monitor and assess the designed systems for responsible AI. Google AI has listed general recommendations for practices to aid in designing responsible AI. [5]

AI analytics is the only way to analyse big data. The bias in systems is partly due to it existing in our society. There is a lot of interest in finding solutions to avoid bias, but there is no conclusive solution to avoid all the different biases. [19]

The development of AI-supported tools requires data from organisations. Some technologies (e.g., chatbots) can be developed to some extent without specific data from the usage context. But when the data is needed from the end-user company, resourcing of the data availability and access to it is often overlooked. In order to implement new AI solutions for sales organisation operations, there is a need for organising and accessing data, securing the data and training the personnel to understand the limitation and possibilities of the AI systems.

AI systems in sales have the potential to increase not only the impact of the sales process but also of the entire B2B ecosystem [6].

2.2 Tools and solutions for customer segmentation

In today's extremely competitive business environment, the ability of a supplier to deliver positive customer experiences throughout the customer journey from first contact to retention can make a significant difference in the sales outcome. In B2B, customers are looking for solutions more seriously than opportunity-driven consumers in B2C. B2B products are often complicated and require technical knowledge both from the customer and supplier side

[5] *Responsible AI practices*. (2019). Google AI. https://ai.google/responsibilities/responsible-ai-practices/

during the negotiations that can take even years and involve large amount of support from supplier personnel other than sales.

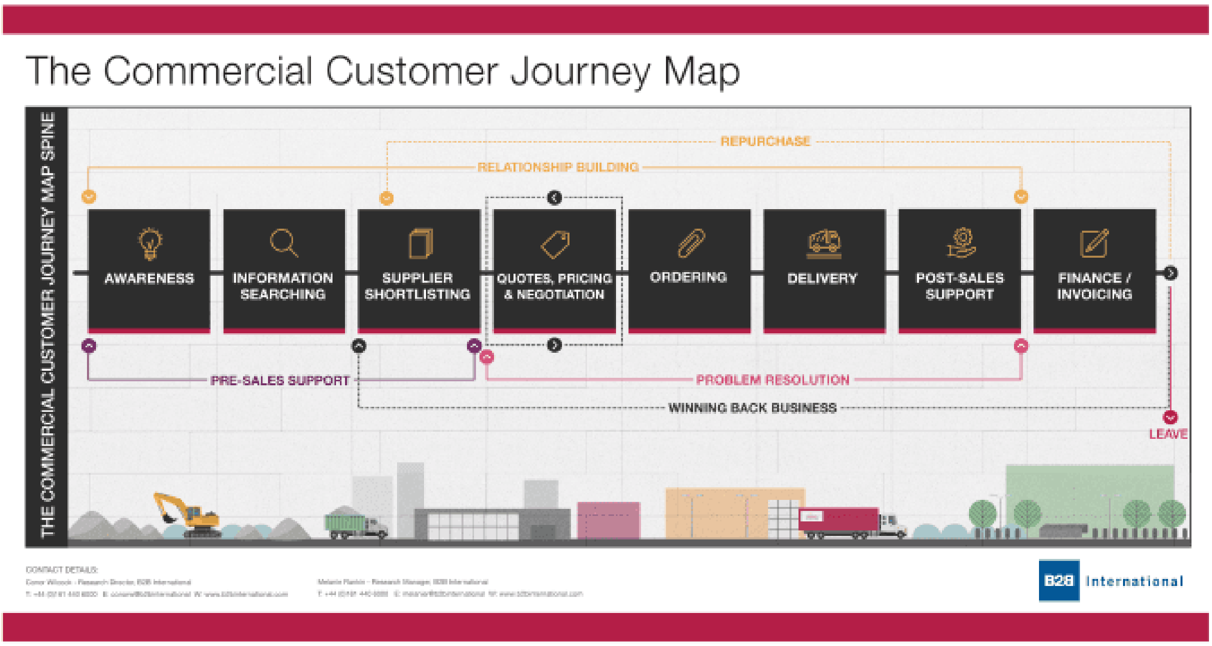

Figure 1: Tactical B2B customer journey map[6].

In B2B sales, the amount of effort to convince each potential client is considerably higher than in B2C. Thus, one of the key challenges in sales is to allocate efforts to minimize the cost and increase the return on investment (ROI). To achieve these goals, one method is to use **customer segmentation,** where customers are segmented into groups based on certain criteria[7]. This allows reaching similar customers with a similar offering, thus decreasing the amount of work needed. Segmentation can be used in many company functions such as in marketing when one wants to advertise to a certain group of companies with a specific message, in sales to propose a customer with similar offering as what other relevant companies have purchased before, or in after-sales to offer spare parts based on typical service intervals of certain machinery.

However, in order to make customer segmentation work, one needs accurate data that fits the segmentation purposes. There is a wealth of customer-related data that can be used, for example, data from web pages, contacts to customer support, previous offers, and purchase documents. The amount of customer data is expected to grow. By 2025 Gartner expects 80% of B2B sales interactions between suppliers and buyers to occur in digital channels. B2B buying behaviours have been shifting toward a buyer-centric digital model.[8]

When customers are segmented into categories, the supplier can prepare their actions to create better customer experiences, which guides the customer to favour the supplier and then return in upcoming new needs.

[6] https://www.b2binternational.com/publications/customer-journey-mapping/

[7] https://peertopeermarketing.co/b2b-segmentation/

[8] https://www.gartner.co.uk/en/sales

There are multiple segmentation criteria available; some of the commonly used ones are shown in Figure 2. Criteria that require large sets of data and multiple parameters, and are susceptive to temporal changes benefit the most from the use of AI, e.g. ones that are based on customer needs, sophistication and behaviour.

Figure 2: Examples of B2B customer segmentation methods.

Customer segmentation is most often used in marketing purposes and a segmentation feature can be found in many commercial marketing tools and platforms such as CRM systems.

2.2.1 Customer Segmentation in major CRM systems

SAP

SAP[9] is one of the leading CRM system providers in the world. SAP CRM has many features related to segmentation and originally one could **configure segmentation in the SAP CRM system manually** with mainly three types of data sources: 1. Marketing attributes, 2. Infosets and 3. BI queries.

More advanced **customer segmentation is done in cloud services** with SAP marketing cloud Segmentation Model[10], which enables identifying the right target group of audience by segmenting high volumes of customer data. **SAP BW also supports recency, frequency and monetary value (RFM) analysis,** and is typically used to divide business partners into RFM segments[11]. SAP also supports **rudimentary machine learning with SAP HANA[12]** Predictive Analysis Library (PAL) that has functions to analyse regression, time series, and social networks. An example is to use the SAP HANA PAL clustering function and k-means to find sales market segments to improve the revenue strategy.

Microsoft Power BI

Power BI[13] is a cloud-based business intelligence tool by Microsoft. Power BI consolidates data across different sources and software programs into one platform, even if the programs are not integrated. Power BI enables manual customer segmentation analysis and comparing customer segments in the current period and prior period. This technique is a combination of

[9] https://blogs.sap.com/2012/03/14/configuration-of-segmentation-in-sap-crm/

[10] https://blogs.sap.com/2018/10/09/introduction-to-sap-marketing-cloud-segmentation-model/

[11] https://help.sap.com/saphelp_nw73/helpdata/en/09/f0895360b93d58e10000000a174cb4/content.htm?no_cache=true

[12] https://blog.sap-press.com/market-segmentation-with-sap-hana-pal-machine-learning-and-sap-s4hana

[13] https://cargas.com/blog/the-benefits-of-using-power-bi-with-dynamics-365-crm/

DAX formulas and correct data modelling.[14] Customer segmentation with Power BI is also possible to implement with machine learning, e.g., clustering[15], and Power BI supports bivariate and multivariate clustering methods[16]. In fact, Power BI has quite extensive AI capabilities as seen in the figure below[17].

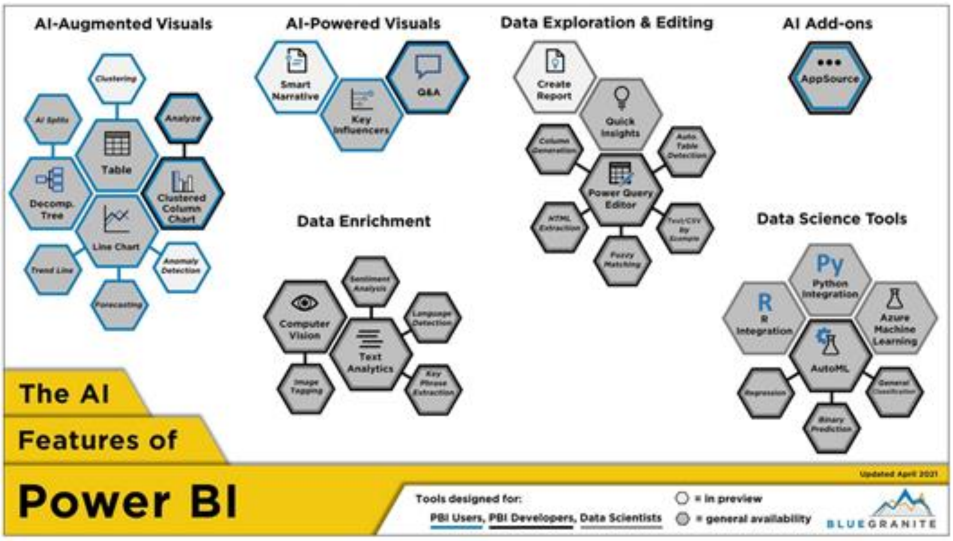

Figure 3 The AI features of Power BI.

Power BI also enables integration to popular data science tools such as Python and R for cases like DIY customer segmentation RFM analysis with Python[18] and Bayesian regression[19] with R.

Salesforce

Salesforce enables manual marketing segmentation by using the insights and data available through Salesforce Marketing Cloud (SFMC). It is possible to apply metrics to entire companies based on their size and industry, and to add geographic and behavioural data. There are two chief methods of segmenting the market natively in Salesforce Marketing Cloud: data filters, and SQL queries.[20]

Salesforce offers machine learning features via Einstein AI (built on a proprietary AutoML stack known as TransmogrifAI), which enables building custom predictions (predict business outcomes, such as churn or lifetime value) and recommendations, predictive insights, and to

[14] https://blog.enterprisedna.co/power-bi-customer-segmentation-showcasing-group-movement-through-time/

[15] https://www.linkedin.com/learning/customer-insights-and-consumer-analytics-for-organizations-tools-and-analysis/customer-segmentation-with-power-bi

[16] https://www.pluralsight.com/guides/implement-clustering-in-powerbi

[17] https://www.bluegranite.com/blog/the-ultimate-list-of-ai-features-in-power-bi

[18] https://medium.com/analytics-vidhya/customer-segmentation-analysis-with-rfm-using-python-and-power-bi-1a93e7938053

[19] https://towardsdatascience.com/how-to-use-machine-learning-in-power-bi-with-r-6b6a930f0310

[20] https://sfdcfanboy.com/2020/04/05/how-to-segment-your-audience-on-salesforce-marketing-cloud/

operationalize AI by adding it to workflows or business processes. [21] It is possible to integrate Salesforce with custom AI applications developed in Python via SQL data queries with the use of, e.g., Postgres databases.[22]

Oracle

Oracle supports Customer Segmentation Analysis with the usage of Data Flow in Oracle Analytics. Oracle provides prebuilt Machine Learning algorithms to train datasets. For Customer Segmentation, the provided algorithms for building models are K-Means Clustering and Hierarchical Clustering.[23]

Oracle also has Oracle Audience Segmentation, which enables processing of large volumes of often-siloed and cross-channel customer data. The default data model within Oracle Audience Segmentation comes with a pre-packaged set of data objects — organized into profile, behavioural, and transactional object groups as well as other attributes — that help manage data more effectively[24]. Like the other CRM/BI systems, Oracle Machine Learning in Oracle Database supports data exploration, preparation, and machine learning modelling at scale using SQL, R, Python, REST, AutoML, and no-code interfaces.[25]

[21] https://www.salesforce.com/products/einstein/features/

[22] https://atrium.ai/resources/integrating-custom-machine-learning-models-with-salesforce/

[23] https://codingsight.com/customer-segmentation-with-data-flow-in-oracle-analytics/

[24] https://www.oracle.com/cx/marketing/audience-segmentation/

[25] https://www.oracle.com/data-science/machine-learning/

3 System design and data acquisition

Successful implementation of AI-driven solutions requires efficient access to reliable data sources. Similarly, efficient adoption of these new solutions necessitates seamless integration into the existing operational systems and services. In this section, we review the interoperability with existing internal organisational and external systems. We also discuss data sharing solutions and challenges between systems and organisations.

3.1 Internal organizational systems and data interoperability

Industrial systems such as ERP and MES are traditionally based on ANSI/ISA95 model, which describes the hierarchy and roles of enterprise systems and automation systems. ISA95 dates back to 1980s when proprietary systems were often insulated and had very controlled interfaces to other systems [52]. The data-centric Industry 4.0 approach has challenged this model by introducing concepts such as digital twins, AI and/or Cloud. Industrial IoT (IIoT) platforms such as IBM Watson or Microsoft Azure IoT are flattening the architecture and providing functionalities that were absent from the traditional model. IIoT platforms are plentiful [52] and there is a lot of competition regarding which platforms would survive [53].

On the other hand, the industry is filled with legacy systems, especially with machinery that have very little ability to communicate with other systems or be controlled with modern systems. The legacy systems are slow and often expensive to be replaced, and often the solution to this problem has been to build middleware to connect them. The middleware approach makes the whole IT system more complex, more prone to errors and even harder to replace as organizations have increased the number of layers during years of operation [54].

In 2021 the ISA95 model has been extended with an activity model to be more in line with Industry 4.0 thinking. From the conceptual point of view, it has replaced the hierarchical (the pyramid) approach with a networked node model. To fulfil the requirements for the Industry 4.0 thinking imposed on smart factories, interoperable communication between machines, processes and humans must be realised. A variety of standards and even bigger standardization frameworks (such as RAMI4.0, Figure 4 below) have been built to foster interoperability [55].

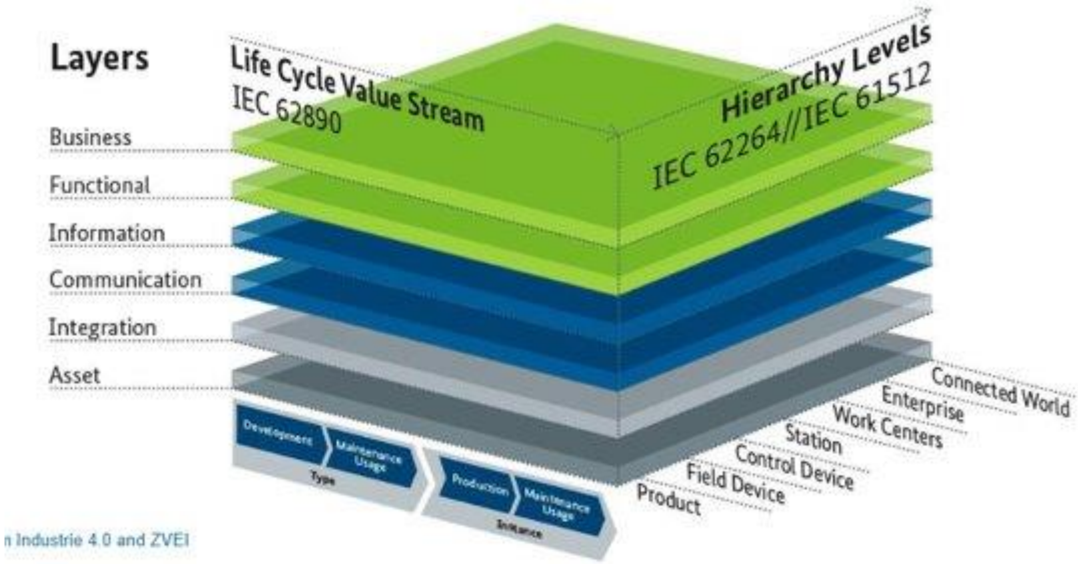

Figure 4. Reference Architecture Model Industry 4.0 (RAMI4.0) architecture[26].

OPC-UA (Open Platform Communications – Unified Architecture) made by the OPC Foundation has become a dominant standard of communication for automation systems which has evolved from its predecessor OPC. It is platform-independent to cover as much of the automation field as possible. As use cases differ from each other, it needs to be very flexible and resource-rich [56]. MQTT (Message Queuing Telemetry Transport) could be said to be a rival for OPC-UA and it excels by being so lightweight that devices with less resources can utilize it. It can work with less reliable networks and has been adopted in the IoT world. [57]

Previously separated systems of ERP and CRM have been in flux and lots of blurring of lines has occurred. Most ERP systems are modular and have been expanded to include the basic functionality of sales and CRM systems. On the other hand, traditional CRMs have included ERP functionalities. Integration of CRM and ERP has lots of benefits for an organization as implementing point-to-point integration between CRM and ERP can be costly, complex and carries a risk that something breaks when either of the system receives an update. [58], [59]

One important shared functionality between ERP and CRM is sales forecasting. Many different things affect forecasting. There are internal factors such as changes in product lines or personnel and there are external factors like competitive, legislative or seasonal changes. To accomplish forecasting, the flow of information is needed from both internal and external sources. [60]

[26] https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf

3.2 External data sources and data acquisition for customer segmentation

External systems in the InnoSale context are the systems outside of the organisation's own control which need to be accessed and with which data needs to be exchanged. The InnoSale systems need to conform to the rules of external systems including the technical interfaces and contractual limitations regarding data usage. The use cases in InnoSale require obtaining data from external systems; sharing InnoSale data with external systems might not be important. Data sharing as an umbrella term covers both cases.

Cross-organisational data sharing has several challenges. Some are technical such as syntactic and semantic interoperability [61] and the inability to ensure data quality. There are organizational issues where the need for data sharing is challenged due to the business issues or the risk of losing control of the data [62]. Especially, convincing multiple organisations to work within the same ruleset of data sharing has proved to be a substantial barrier to overcome.

To use third-party data, one may need to accept a license for the data usage. There are several existing licenses, and they may allow for commercial or non-commercial use. The terms of license can be complex and ambiguous; an example case is how a model trained with a licensed data set can be utilized in other contexts where the data itself is not present [63].

In the web context, REST (Representational State Transfer) has become the main way to realize web services [64][65]. It is based on a service-oriented architecture where client/server roles are clearly defined, and requests and responses are built on a transferring process of resources. The REST is stateless and all requests need to carry all the data they need. The combination with JSON has made RESTful web services the dominant way to cloud providers to deploy their services thanks to its easy scalability [66].

Data spaces present the possible future way of sharing data in domains such as agriculture or energy. Big Data Value Association (BDVA) defines data spaces as an ecosystem with shared data models, data sets, ontologies and data sharing contracts [67]. International Data Spaces (IDS) is one such implementation, and its focus is on data sovereignty and data governance. Data spaces have yet to prove themselves as a mainstream way of data sharing, but especially in the EU large investments have been made. Data spaces can also be seen as a way to introduce a data economy in a larger scale [68].

APIs have a crucial role in data sharing on the web. To create good APIs there are several key issues, which should be considered. First is to use standard web technologies such as previously mentioned JSON and REST. Second is to use conventions of error handling and endpoint management. Third is to ensure the security and provide access control whenever needed. There are software tools that enable communication from machine to machine. They come with a contract that has a terms and conditions license, guarantees and liabilities in case of infringement, and outline how the interface can be used by developers. [69]

Cross-organisational data sharing entails lots of small details to think about case-by-case. For example, if you would like to buy weather data for your organization you may be limited by an API request per month, Queries per Second and get different SLA (Service Level Agreements) for ensured uptime. There is also a requirement that you have to cite the data source in your own service. This is an example of how the use of external data could occur; other data sources have different approaches how they share their data. From a protocol point-of-view, the API works with HTTP requests and returns JSON formatted data. A data marketplace would be a place to find and purchase data; the marketplace would be working as an intermediary of the transaction.[27]

Privacy and protecting trade secrets in data sharing is one of the most important topics: How to share cross-organisational data and ensure that the data is anonymized and treated in a way that it does not reveal trade secrets? If these cannot be ensured, the resulting situation is one where no data is shared. To resolve the issues there have been several methods proposed, e.g., time-slicing shared data to hide the how the data is being obtained and just sharing a small part of it. This holds mostly for sensor data [2]. There are several data anonymization tools available, free tools like ARX or commercial ones matching several different needs like BizDataX and Clover DX. Another way to use data is to generate synthetic datasets with similar statistical properties to the input data and with the same format [70]. There are several tools to generate data sets, but usually these are domain-focused. For example, commercial synthetic data generators like Mostly.ai and Datagen could be used.

Within the EU and especially globally regulatory inconsistencies may cause trouble when using data from a third party. Data flows easily through borders but handling disputes and contractual breaches are dealt nationally. There are several legislations covering the use of data inside the EU, GDPR being among the major ones [69].

3.3 Solutions for data integration

New business models, changes in application development trends and more importantly urgency of modernization fuelled by customers' demand for seamless user experience direct software engineering leaders to diversify and increase demand for integration platforms that stitch together the ever-growing number of distributed apps and data. In addition to real-time data needs for user experience requirements and digital business capabilities, the COVID pandemic and quarantines accelerated business process automation needs. In order to deliver these expectations, IT business units embrace data integration tools more than ever.

When evaluating a data integration tool companies look for providers that: 1) align with their unique strategic business usages for integration; and 2) enable modern application architecture that enables agile and scalable deployment. On top of that, full life cycle API management offering capabilities are a necessity in functional areas such as developer portals (ecosystems of developers who produce/use APIs), API gateways (including runtime monitoring for APIs), policy management and analytics (security configuration and usage analytics), API design and development, and API testing (including performance and security).

[27] https://corporate.foreca.com/en/foreca-weather-api?hsLang=en

Gartner Magic Quadrant[28] and Forrester Wave[29] are industry-standard guides when it comes to market research and qualitative data analysis regarding different vendors offering enterprise software solutions. Integration platforms/tools are no exception, and an extensive analysis is done to assess top vendors in terms of their current offering (integration scenarios supported, capabilities to ease integration development, deployment characteristics etc.), strategy (product vision and roadmap, partner ecosystem, supporting products and services etc.) and market presence. Figure 5 and Figure 6 below depict the vendor landscape for integration tools. Leaders in that landscape are:

- **Boomi:** Boomi is an integration platform as a service (iPaaS) provider with a focus on data quality and customer experience. It provides application, data and B2B/electronic data interchange (EDI) integration, API management, low-code application development and master data management (MDM). Market understanding, strategy and traction alongside vertical industry strategy are among the strengths of the tool.
- **Mulesoft:** MuleSoft's AnypointPlatform combines iPaaS with API management and classic enterprise service bus (ESB). It provides advanced tools for API design and management and excels at developer productivity with a full template marketplace and debugging tools. Product capabilities are among the tool's strengths: combining application, data, B2B and event-driven integration, as well as API management and microservices development.
- **TIBCO:** TIBCO Cloud Integration provides functionality for integrating applications, data, APIs, B2B and IoT as well as automating processes. It is supported with additional products like data virtualization, internal service mesh and data governance. It also has enhanced augmented data mapping and suggestions on source data connection to tap into ML-powered data sharing scenarios.
- **Workato:** Workato is a cloud-focused serverless tool that excels in security with multi-layered fine-grained access grants and can trigger integration flow in response to internal DevOps events. It lacks key B2B features such as EDI and partner management but is praised for ease of use and breadth of features.
- **Jitterbit:** Jitterbit provides a wide array of connectors and integrates diverse cloud, inter-enterprise and on-premises environments and provides API creation and management as well as B2B support. Jitterbit targets mostly small and mid-size organizations in the manufacturing and technology sectors.
- **Software AG:** Software AG offers webMethods.io, which is a tool that includes B2B and API capabilities and supports a broad set of use cases, including application integration, data integration, B2B integration, managed file transfer and API-based integration. It offers a headless version for OEMs that can be embedded within third-party software components.

[28] Gartner Magic Quadrant for Full Life Cycle API Management, Gartner, September 28, 2021

[29] The Forrester Wave™: Enterprise iPaaS, Q4 2021, Forrester Research, October 5, 2021.

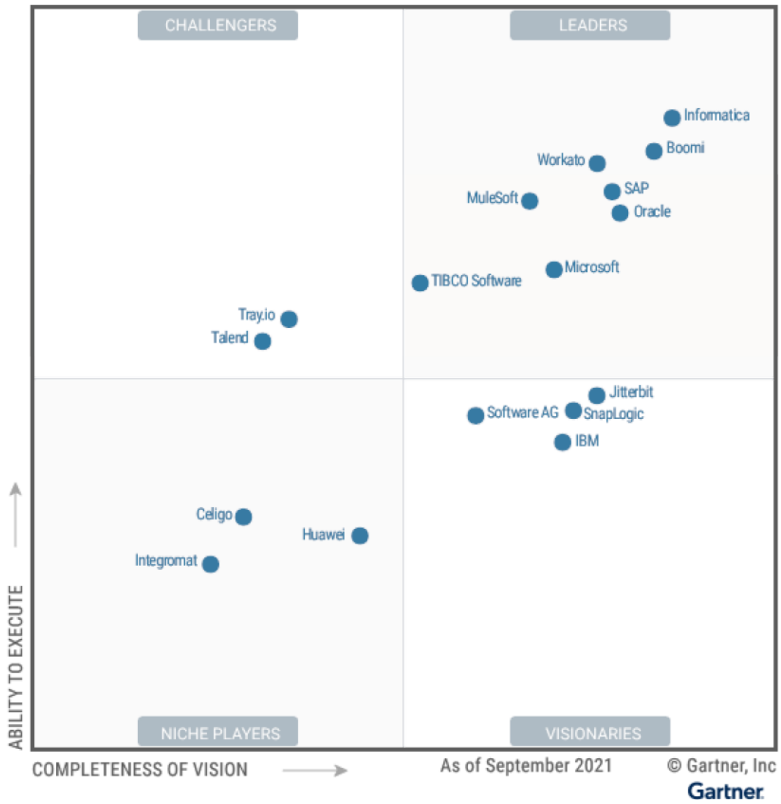

Figure 5. Magic Quadrant for Enterprise Integration Platforms as of September 2021[30].

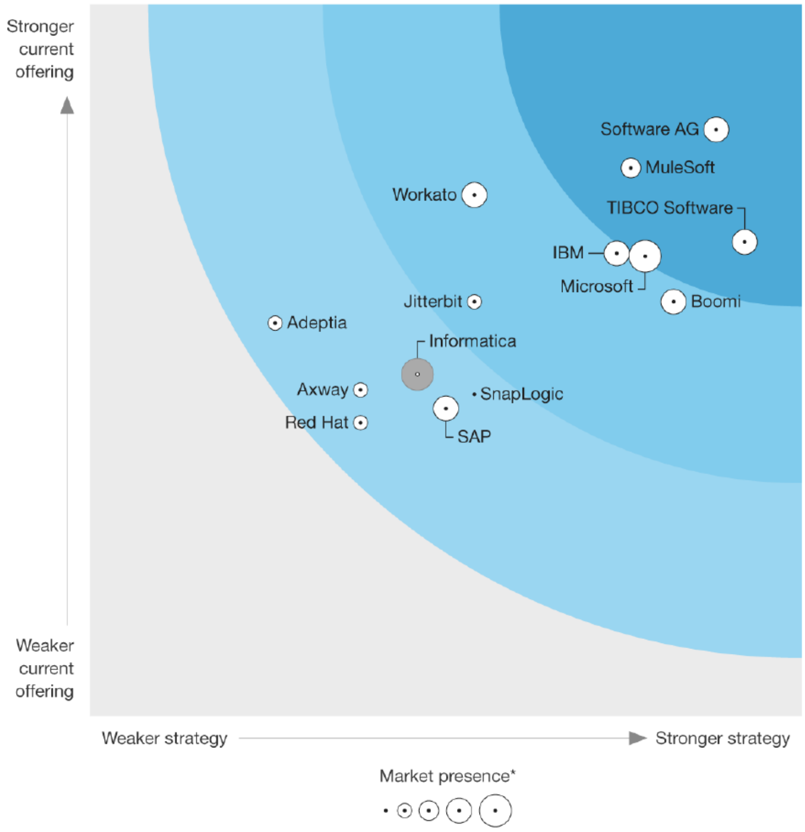

Figure 6. The Forrester Wave: Enterprise Integration Platforms as of Q4 2021[31].

4 AI methods for knowledge modelling and data analysis

4.1 Solutions for knowledge base design and implementation

Knowledge base (KB) is a structure that mimics the human knowledge portrayed in the brain so that it can grasp and answer almost all the questions customers may ask. Knowledge base systems have three sub-categories, namely linguistic knowledge bases, expert knowledge bases, and ontology, the last of which will be used in this project to strengthen the data encoded and provide better decision-making. Machine learning (ML) is an application of artificial intelligence (AI) and in order to coalesce new information into the knowledge base, knowledge fusion is necessary to extend ML methodologies from external knowledge acquisition to internal knowledge generation. [143]

Since knowledge bases contain a vast amount of information about the world, it is a smooth process for the architecture of KBs to comprise a considerable amount of information and inquire about them. There are not only text sources, documents, websites, and emails, but also non-text sources, videos and images; the texts can also be either plain or formatted. As for knowledge to be applicable, it should be detached from these sources and appear in KB.

Given the points above, it is easy to understand that the process of designing and implementing KB is extremely important. The terms “design and implement” both address knowledge base construction. The process of comprehending and subtracting pertinent facts from huge data sources is called Knowledge Base Construction (KBC), and by taking into consideration the above points, it can be said that filling the knowledge base with information is significant. One of the recent interests in knowledge base construction due to question-answering (QA) agents, called chatbots, has been described in [148]. Knowledge bases have high importance in QA systems like chatbots. The knowledge base should be relevant and complete so that it includes all the potential facts and has the ability to answer queries from users correctly. Therefore, academics and companies are working on this topic.

KBC is mostly done manually, and for this reason, sometimes several challenges may occur. KBC, by nature, includes the transformation of unstructured data into structured knowledge. Text, scanned papers, web pages, photos, and other unstructured sources are examples[32]. Because of the following qualities, for computer programs, these materials are difficult to understand [146], [145], [144].

In order to constitute structured knowledge, KBC passes through the process of unstructured sources which is not simple for computer programs to read because of the following characteristic traits:

- Uncertainty: The incorporation of features such as graphs and figures, as well as enriched formatting, introduces uncertainty into the explication of most disorganized sources. Considering the nature of Natural Language, even when the source is pure text, the meaning of the text is uncertain. The scans and photos include multiple meanings, increasing uncertainty [144].

[32] https://cbi.gwu.edu/workshop-rapid-biomedical-knowledge-base-construction-text

- Complexity: Means, connections, aims, goals etc. are integrated for making up knowledge. It is the enormous number of possible combinations of these factors that make it difficult for systems to recognize only those features that are both relevant and beneficial [144].
- Configuration efforts: The KBC system belongs to people who are experts in the domain. However, the complexity of the system and the need for technical activities push field experts to work with technical experts. This makes it difficult to organize the configuration in the first place. In addition, the system must be designed to be robust to require at least manual operation. This can increase the cost of the configuration. Balancing the ease of setup with the strength of information retrieval automation is a challenging task [144].
- Quality requirements: The KB serves a certain goal such as training a machine learning model or responding to inquiries. To achieve this, the created KB must fulfil the set of exact quality standards. The main requirements are:
  - Completeness: It is important to have the total knowledge required for the task, but apart from this, it is essential to harvest that knowledge [147], [144].
  - Accuracy: It is important to harvest knowledge, but one must be assured whether it is accurate or not [148], [147], [144].
  - Two papers [148], [147] give us recent research on how to create complete and accurate knowledge bases. According to these articles, there is no exact way to automatically determine if the knowledge base is accurate and complete [144].

For instance, the expertise of a chatbot should be sufficient so that it can offer appropriate responses to all expected responses. The KBC system faces a tough task in meeting these quality objectives [144].

There is some scientific research for the architecture of KBC and state-of-the-art samples for KBC. The remaining part will focus on these topics.

There are two different elements in the knowledge base: first, schema elements which are used for rationing and picking up the exact data elements and appear in natural language format; and second, data elements which symbolize the conclusive use of the knowledge base and appear both in natural language format and formatted layout [144]. Natural Language Processing (NLP) is used to extract information from text. Grammatical analysis implemented using regular expressions, and FSA-based algorithms can be used by NLP-based knowledge extractors. The grammatical technique is quick and efficient, but it requires extensive programming and has limited scalability.

Multi-layer Neural Networks (MLNN), often known as Deep Learning (DL) models, have become popular in recent years for text processing. These models require a lot of data to initialize, but they 'equalize' the KBC process by allowing non-programmers to train and then use DL-based extraction pipelines. Pipelines to extract instances from these sources must use spatial relationships and include modules to transform images to text (commonly known as OCR), analyse layouts, and tag values with appropriate schema elements [150], [149]. In addition, [161] informs us about more about recent surveys on Knowledge Graph Construction, and two articles [157], [159] provide a variety of KBC architectures.

There are several different existing KBs. The following parts will focus on three different KB systems.

- Publicly Available KBs: DBPedia is the most common publicly available KB. It extracts data from Wikipedia and makes it public as an RDF dataset, which depicts web resources and data interchanging over the World Wide Web. Info boxes, which is a component of Wikipedia, is used by DBPedia and so its attention is on relationship extracting and linking. The NELL project, which is a KBC system that detaches necessary information from the web, is another public KB [144]. If we utilize this form of KB, [163] can assist us in deciding which one to employ. In the article, five different knowledge graphs, which are a type of knowledge base, are mentioned. These are DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. It lists simple guidelines on when and how to use which knowledge graph in a specific situation, as well as how to choose the proper graph for the purpose.
- Open-Source Platforms: Several KBC systems were created in academic institutions and then released as open-source software to the general public. Deep Dive is a Stanford University -developed KBC platform [151]. Apple Inc. bought the commercial version in 2017. Snorkel is another Stanford-developed method that focuses on providing weakly supervised data for deep learning model training [152]. Founder, another Stanford project, tries to extract knowledge from elaborately formatted text [144], [153].
- Commercial Products: Software products capable of KBC are developed by several technology companies and the pioneer in this market is IBM with its product called Watson. IBM has produced a product for empowering semantic search via the KB detached from the text called Socrates[33] and produced SystemT which is used for commercial KBC projects [154]. Knowledge Vault [160] which is developed by Google is used to strengthen its 'information boxes' and provide answers to Google Assistant [144].

Since InnoSale focuses on e-commerce we especially searched this subject and discovered an article produced by AliBaba Group in this field from which InnoSale may benefit. The stages of constructing the domain knowledge graph created and used by them in the light of deep learning technology are explained in detail: how they do knowledge mining, named entity recognition and relation extraction. It also explains how this knowledge graph is used for QA and recommendations in the pre-sales phase and what results it provides from it. Alibaba Group's work on the subject is ongoing and Multimodal KG is a key research topic to meet the need for multi-modality item content (text, image, video, etc.) as the next step [164].

To sum up, one of the main challenges in InnoSale is understanding and answering the query coming from customers automatically. There are recent articles that we can benefit from regarding solutions for complex questions answering over the knowledge bases [155],[156]. The complexity and uncertainty of the source material make KBC a challenging task. In this section, academic articles published in recent years for KBC were examined and some results were presented.

[33] https://www.ibm.com/blogs/research/2017/11/knowledge-base-construction-iswc-2017/

4.2 Practical approaches to Existential Rule-Reasoning and possible integration of answer set programming (ASP) for optimization problems

In knowledge representation and reasoning, information is often encoded in knowledge bases, which contain a set of facts enriched with an ontology that captures general knowledge of the application domain by describing a set of concepts and the relations between them. Such a knowledge base can be queried, providing answers that are not only based on the explicit information stored in the database, but also based on the implicit information that can be derived from the ontology [71], [72]. Implementations of algorithms for rule-based query answering often use an iterative bottom-up reasoning algorithm, known as the chase, that results in a universal model on which the query is evaluated [81].

Existential rules, also known as Datalog$^{\pm}$ or tuple generating dependencies, is a powerful formalism able to express this kind of ontological information [72], [82]. Syntactically, they extend Datalog by allowing existential variables in the rule's consequent. This allows to assert the existence of unknown entities, which is a fundamental feature for reasoning on incomplete data. Many description logic (DL) ontologies can be captured in this way, allowing rule engines to be used as ontology reasoners [77].

However, the great expressive power of existential rules comes with many theoretical and practical challenges. For a knowledge base with existential rules, conjunctive query answering is already undecidable and it is also undecidable if the chase terminates on a given knowledge base [73], [74]. One common approach to solve this problem is to restrict it to fragments of existential rules for which conjunctive query answering is decidable, e.g. guarded existential rules [75], [76], [80]. There are also sufficient conditions for chase (non-)termination that allow to give a definite answer to whether or not the chase will terminate in certain cases [77], [78], [79].

Implementations for the chase and sophisticated tools for conjunctive query answering have been released and are still under active development throughout the recent years [83], [84], [85], [86], [87], [88]. These systems are generally based on Datalog, with various restrictions and extensions; several systems now also support existential quantifiers in rules.

Recently, a new type of chase-based reasoning algorithm that stores inferences in tables organized column-by-column rather than row-by-row was suggested [89]. Columnar databases can lead to faster joins and are very memory efficient, but suffer from poor update performance. To solve this, the new approach never inserts new inferences into an existing table but rather creates a new table instead. The inferred facts for each predicate, therefore, are distributed over several, read-only tables of disjoint content. This approach has been implemented in the rule engine Vlog. A Java API for this tool is provided in the Rulewerk library for easier integration into other systems.

While existential rules focus on expressing ontological knowledge in an open-world domain, Answer Set Programming (ASP) is a rule formalism, which is particularly suited to solving combinatorial problems in a closed-world environment. One important feature of ASP is its ability to express negation as failure even for non-stratified programs. This allows for non-monotonic reasoning, where new information can invalidate prior assumptions. Furthermore, the stable model semantics employed by ASP gives rise to a unique modelling strategy, by which each stable model represents a solution to a problem encoded as an ASP program [91]. Since its conception, many other useful features have been added to the ASP language. This includes disjunctions in rule heads, which provide a direct way to express non-determinism, classical negation, aggregates, function symbols and even optimization constructs [90].

Programming in ASP often follows the simple pattern of guess, check, and optionally optimize [92]. In the guessing step, the ability of ASP to express non-determinism is used to describe the set of candidate solutions to a given problem. By formulating constraints, the set of candidates is narrowed down to the actual solutions to the problem. Each answer set of the program then corresponds to a valid solution. Optionally, each constraint may be assigned weights and priorities, which define a set of optimal solutions. ASP, therefore, provides a declarative programming approach that is able to easily express and efficiently solve NP-complete problems.

ASP programs are typically evaluated in a two-step process. In the first step, called grounding, the given program is transformed into an equivalent variable-free (ground) program. The second step, called solve, takes the ground program as input and computes its stable models [93]. Today, there exist a wide variety of systems that implement this general approach [92], [94], [95], [96], [97], [98], [99], [100], [101]. However, this strategy is unsuitable if the ground program becomes prohibitively large. To alleviate this problem, some systems opt for a groundless approach that grounds rules on-demand during solving – a technique which is called lazy grounding [102], [103], [104].

While useful for modelling ontologies, existential rules lack the crucial feature of expressing negation. ASP on the other hand is unable to deal with existential assertions. Combining both formalisms, therefore, holds great potential. A seamless integration of the two, however, is difficult since existential rules operate under the open-world assumption while ASP uses closed-world semantics. Earlier attempts focused on combining Description Logic with ASP rules by splitting the knowledge base into an ontological part and the ASP program [107], [108], [109]. More recently, core-models have been suggested as a way to define the semantics of non-monotonic negation when used in existential rules [105]. Another line of work generalizes ASP and existential into so called existential non-monotonic rules (ENM-rules) by using skolemization [106].

4.3 Named entity recognition approaches

One of the early definitions for Named Entity Recognition (NER), is that a named entity (NE) is any proper noun that specifically refers to something or someone [110]. It is a word or a phrase that distinguishes an item from a list of other similar items and can differ in nature according to the knowledge domain [111]. In generic knowledge, NE can be a person, location, etc. In a specialized field like cranes, it can be a chain hoist, hook path, etc. NER is then the task of categorizing words/phrases from an input sequence into their corresponding named entities. NER is an essential step in Information Extraction (IE) and serves as an important factor in many tasks such as text processing and understanding, information retrieval, and knowledge base construction [112].

Early approaches for solving the NER problem were rule-based, which relied on hand-crafted rules using semantical and lexical patterns. These approaches perform well when there are highly trained experts to create rules, an exhaustive lexicon, and a clear formalisation of the rules [113]. This basically precludes generalizability and disables adapting to future changes in the same domain. Unsupervised approaches helped in generalizability by leveraging large corpora but did not dispense the need for rules seeds [113]. Feature-based supervised learning also played a role when a large amount of annotated data is available, besides well-curated features (e.g., case, morphology, and part-of-speech tag) that represent the data for training these models [114]. Finally, deep learning models were used to achieve a new state-of-the-art performance and to overcome the limitations of the three pre-mentioned techniques, including fine-grained rules crafting, generalizability, and feature engineering [112].

The first essential component for the NER task is the representation of the input sequence. Given a sentence, multiple strategies can be adopted to convert the raw language of that sequence into a lower dimension of numerical vectors (embeddings), where each dimension captures a latent feature of the sequence. Three approaches have been studied, which worked on the word level, character level, or hybrid levels. Word-level embeddings are obtained by utilizing unsupervised models and a large corpus of data, that can then learn to assign similar vectors to the words of similar context or relationships [115]. Character-level embeddings are obtained using deep learning models, mostly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). They are better at capturing sub-word structures like prefixes, suffixes, morphemes, etc. [112]. A hybrid embedding would incorporate multiple features such as gazetteers, word embeddings, spelling, and context features [116] or a balance between the word-level and character-level embedding to produce sub-word embeddings [117].

The second essential component is an encoder-decoder model. The encoder is a context-aware model that creates a representation of the input sequence, enabling the decoder to perform its task. The decoder is responsible for outputting a label (location, person, etc.) per token after utilizing the embeddings and the context encoding. RNNs are considered the de facto approach for a NER encoder [112], especially bidirectional Long-Short Term Memories (LSTMs) capturing information when encoding the sequence from left-to-right and right-to-left [116]. The most commonly used model as a decoder is a Conditional Random Field (CRF) that has been coupled with most of the state-of-the-art performing encoder models [112].

With the advent of transformer models, many studies were able to utilize the Bi-directional Encoder Representation from Transformer (BERT) as a standalone model for NER, utilizing its intrinsic hybrid embedding and encoding, and applying a simple softmax [118] or a softmax + dice loss [119] layer on the top of the model for decoding. The performance of these two models is very close to the current state-of-the-art models.

Even though the BERT model has shown significant improvement over the state-of-the-art in multiple Natural Language Processing (NLP) tasks [120], it falls short in the NER task against other established techniques, because it was not tailored properly for NER during pre-training [121]. However, the achievements of transformer models on other NLP tasks are pushing NER researchers to explore different ways to utilize them for the task. For example, adjusting the pre-training step can generally improve the performance on NER tasks [122], and it can be made more transferable to other domains [123], or low resource domains [124].

4.4 Semantic search

Semantic search has been comprehensively summarized in a previous study [49]: "In a nutshell, semantic search is 'search with a meaning'. This 'meaning' can refer to various parts of the search process: understanding the query (instead of just finding matches of its components in the data), understanding the data (instead of just searching it for such matches), or representing knowledge in a way suitable for meaningful retrieval". In this section, we look at the scientific state-of-the-art of semantic search, its markets, and how it has been applied in the industry.

4.4.1 Semantic search with modern deep learning -based NLP techniques

Natural language processing (NLP) is the area of automatically representing and analysing human language through computational techniques [125]. NLP was limited to tasks that utilized intrinsic statistics about the language, without much understanding of its actual meaning. This was shown in famous algorithms such as Bag Of Words (BOW) [126], and Latent Dirichlet Allocation (LDA) [127]. Later on, deep learning techniques started to control the field with context encoding using algorithms such as Recurrent Neural Networks (RNN) applied in Long Short-Term Memory (LSTM) way [128]. That was the de facto stage until 2017, when the paper, "attention is all you need" [129] introduced the transformer model; a model that is dropping RNNs altogether and relies mainly on the self- and cross-attention mechanisms. This proposal significantly boosted the NLP field by allowing pre-trained models to achieve new state-of-the-art performance, unprecedented reduction in training time, and the processing of much longer sequence lengths.

The transformer model is an encoder-decoder model where the encoder is used for extracting a semantically informative representation of the input, while the decoder generates appropriate output. The encoder takes a sequence and starts by embedding it into a high-dimensional numerical representation that captures the word co-occurrences. Since the model does not use recurrence, the positions of the words within the input sequence need to be provided as additional information, for which the authors propose so-called positional encodings. The sequence then goes through self-attention, which is a matrix multiplication for the numerical vectors of each word. The aim of such multiplication is to capture dependencies between the words and identify which words are more important and relevant to each other. Attention can be performed multiple times to capture different aspects of the sequence with each round. The new "attended" representation of the input sequence can then go through a simple feed-forward neural network for training and outputting a contextualized representation of the input. The decoder then tries to generate sequences based on the contextualized representation of the input sequence. It has a similar architecture to the self-attention-based encoder, except that it additionally applies cross-attention to model dependencies between input and target sequence. An example of an encoder-only model is the famous BERT [120], which is now serving as the building unit for several state-of-the-art

NLP models. An example of a decoder-only model is the famous GPT [130] model and its subsequent generations.

BERT and GPT models have been trained on a huge amount of generic data and became very powerful for a variety of tasks such as text classification, machine translation, question answering, text summarization, text generation, etc. The reason for such performance, besides the amount of data, is the pre-training objectives. For example, BERT uses two tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP), both of which require no labelled data but work on any text in an unsupervised fashion.

Taken out of the box, these models were able to achieve state-of-the-art performance on general knowledge tasks. However, when it comes to domain-specific knowledge without enough data for the models to train on, an option of fine-tuning the model on a subset of the specialized data is available to enhance performance. For more improvement, pre-training from scratch [131] is also possible. However, in practice, a huge amount of data would be needed for training. Another possibility in between fine-tuning and pre-training is to build on the original model's learned weights and parameters, and further pre-train with a corpus from the target domain [132].

When a limited amount of data is available in a domain-specific corpus, a larger model can be leveraged to perform meta-learning, as introduced in the famous paper on the GPT-3 [133] model. The meta-learning is formulated as giving a pre-trained model a task, that is expressed in a natural language form, to perform it with one of three settings: zero-shot (i.e., with no training examples given), one-shot (i.e., with one example given), and few-shot (i.e., multiple examples are given, typically between 10-100). While research showed how GPT-3 was able to perform some tasks with as few as zero examples given, the downside was the limitation of maintaining a large GPT model for normal commercial services.

Fortunately, another publication [134] managed to utilize the smaller and more affordable models to perform few-shot learning in the presence of a large amount of unlabelled data (typically more than 1000 examples). The algorithm performance was reported for the text classification task and relied on semi-supervised learning by utilizing the labelled data to fine-tune a model, using the fine-tuned model to annotate the unlabelled data and go through a cycle of annotating and fine-tuning, and finally using a classifier on the average of these models. This method was further used in another publication [135] that showed comparable results despite discarding the need for unlabelled data. It achieved such results by adding more supervision by utilizing labelled data.

A well-trained transformer model can then be utilized to perform multiple semantic search tasks. For example, a query formulated in a natural language would be contextually encoded and matched to the most similar encodings from the model's trained database, hence enabling a powerful semantic question-answering task. Utilizing the word encoding was also cleverly used to obtain sentence encoding [136] that significantly reduces the time needed for semantic similarity lookup and clustering of similar documents.

4.4.2 Market analysis

Market interest for semantic search has grown due to enormous growth in various internet-based services. Search engines have moved from keyword-based retrieval to parsing and understanding the search phrase, providing capabilities for refined services such as question answering and optimization based on user behaviour data. The e-commerce sector is also a domain in which harvesting the meaning and refining the search results is important for facilitating the user experience. In the realm of conversational chat bots, semantic comprehension is a necessary tool for successful operations. In all of these cases, semantic search is the enabler.

According to a recent survey [21], semantic search is seen as one of the key AI solutions that will drive future growth potential in the retail sector, as the use of AI is steadily rising despite the viewed challenges and restraints, e.g. high investment costs and cybersecurity concerns. Semantic search engine capabilities enhance the customer experience in online retail by facilitating easier search of relevant products and finding precise product information. It is also seen to help in search engine optimization and enable personalized shopping recommendations. A Singapore-based AI retail technology company ViSenze was used as a case profile in [21]. The company provides AI- and machine-learning-based discovery suite that implements Smart Search, enabling customers to search from uploaded images, social media, chat bots, retail catalogues and in-store products[34]. Their estimated annual revenue is 12.6 and total funding is 34 million dollars in 2022[35]. In another study [22], Finnish company Klevu was seen as a contender in offering smart search for SMEs and other online enterprises with a similar profile as ViSenze. Their annual revenue is estimated to reach 8.9 million dollars in 2022[36].

In a study regarding innovations in contact centre services [23], the Canada-based company Enghouse Interactive was highlighted as an enabler of AI-powered solutions for NLP-related tasks, including semantic search, intent recognition, and sentiment analysis. In this domain, legacy software has been a major issue until recent years, when many start-up companies building on modern software have allowed for the use of cloud-based AI and machine learning for facilitating their services, including chat bots. Enghouse Interactive uses Microsoft's Azure infrastructure for its services and has seen a strong growth in the market. Their annual revenue is estimated to 94.8 million dollars in 2022[37].

The US-based company AlphaSense was named as a strong competitor in AI-based solutions for search automation in a recent analysis [24]. Their NLP-based technology is used to automate the classification and extraction of unstructured text, including the fusion of many data sources. AlphaSense's semantic search engine is provided as a business intelligence service for finding relevant information on any company, industry, trend, or topic from over 10 000 sources[38]. The search can be used for discovering

[34] https://www.visenze.com/discovery-suite/modules/smart-search/

[35] https://growjo.com/company/ViSenze

[36] https://growjo.com/company/Klevu_Oy

[37] https://growjo.com/company/Enghouse_Interactive

[38] https://www.alpha-sense.com/content/

regulatory filings, trade journals, disclosures, and many other types of documents. Their annual revenue is estimated to 156.7 million dollars in 2022[39].

In the domain of chat bots, the American Avaamo has been named as a fast-growing company [25]. Their integrated platform provides conversational interfaces for enterprises, utilizing speech synthesis and recognition as well as semantic search. They specialize in services for healthcare and service desks, e.g., banking, education, finance, and IT assistance[40]. They have been financially supported by Intel Capital. Their annual revenue is estimated to be 14.1 million dollars in 2022[41].

In addition to the emerging companies listed above, the market leader companies in the area of semantic (or cognitive) search include Attivio, Coveo, Sinequa, MindBreeze, LucidWorks, and IBM, followed by contenders such as Microsoft, Micro Focus, Squirro, Grazzitti Interactive, IHS Markit and Elastic[42].

4.4.3 Industrial state-of-the-art

In this section, existing and well-known tools by established companies that utilize semantic search are listed. In addition, the industry partners that utilize these semantic search components are enumerated as well.

IBM is known for its Watson, a customizable system that allows for natural language question-answering, effectively utilizing semantic search in its methodology. Watson has seen use in multiple domains, including healthcare, engineering, retail, finance and the legal sector[43,44]. IBM recently teamed up with Salesforce to realize visual search, enabling image recognition in CRM[45].

Oracle offers several database semantic technologies for enterprises, enabling graph-based search on structured (RDF and OWL) data[46]. Partners that use Oracle's Spatial and Graph services include multiple small consulting companies and open source projects that utilize the tools, e.g., ontology modelling and engineering, graph visualization, entity extraction and natural language categorization[47].

39 https://growjo.com/company/Alphasense
40 https://avaamo.ai/
41 https://growjo.com/company/avaamo
42 https://www.datanami.com/2020/11/24/think-search-is-solved-think-again/
43 https://www.techrepublic.com/article/5-companies-using-ibm-watson-to-power-their-business/
44 https://www.zdnet.com/article/ibm-watson-what-are-companies-using-it-for/
45 https://geomarketing.com/as-salesforce-and-ibm-watson-team-on-ai-visual-search-comes-into-focus
46 https://docs.oracle.com/cd/E11882_01/appdev.112/e25609/sdo_rdf_concepts.htm#RDFRM100
47 https://www.oracle.com/database/technologies/spatialandgraph/semantic-graph-partners.html

SAP offers a semantic service in its relational database management system. The HANA Enterprise Semantic Services provides semantic search capabilities, allowing for the discovery of objects such as tables and views. Keyword queries in natural language are also supported[48].

Semantic search is available in Microsoft Azure. It can be used to rank documents semantically based on the search terms, extract sentences from a document that summarize its content and also implements question-answering functionality. Under the hood, pre-trained machine learning models are used[49]. Companies using the Azure-based service are Howden and OrangeNXT[50].

Kendra is an intelligent search service provided by Amazon[51]. It enables answering natural language questions (in addition to keywords) by sifting through a document database and finding the closest matching answer to the question. Customers that use Kendra include 3M and The Wall Street Journal, and comprise such domains as news, material engineering, administration, education, tax and advisory, research, gas industry, and software[52].

Salesforce is a software provider of CRM systems, with an in-built semantic search component called Einstein. It enables personalized, natural-language-based search results for users with different goals, such as salespeople and marketers. The search component is used by companies such as iHeartMedia and MightyHive[53].

4.5 Customer segmentation approaches

Business-to-business customer segmentation varies highly from the business-to-customer sector. Instead of demographic and human behaviour information, the companies must find ways to gain information on higher, company-level. In general, B2B customer segmentation approaches are often[54,55,56,57] divided into five categories, of which the first four are the most relevant for the purposes this report:

1. Firmographics, i.e., business version of demographics,
2. Needs-based,
3. Behaviour-based,
4. Profitability-based segmentation,
5. Customer sophistication.

[48] https://help.sap.com/docs/HANA_SMART_DATA_INTEGRATION/cc7ebd3f344a4cdda20966a7617f52d8/c23380b3f421498884f7b6d8524024d7.html?version=2.0_SPS03&locale=en-US

[49] https://docs.microsoft.com/en-us/azure/search/semantic-search-overview

[50] https://customers.microsoft.com/en-us/story/1344058341075309890-howden-energy-azure-ai

[51] https://aws.amazon.com/kendra/

[52] https://aws.amazon.com/kendra/customers/

[53] https://www.salesforce.com/blog/introducing-einstein-search/

[54] https://www.leadspace.com/blog/popular-methods-of-segmentation-for-b2b/

[55] https://sopro.io/added-value/blog/b2b-market-segmentation-guide/

[56] https://peertopeermarketing.co/b2b-segmentation/

[57] https://surveysparrow.com/blog/customer-segmentation-examples/

In addition to the above classification, the *target* for why customer segmentation is done can also be subdivided. Often, the characteristics obtained via either approach can be used to

1. Predict the win/loss of the sales opportunity
2. Evaluate customer loyalty/churn
3. Rank the potentiality of customer leads
4. Predict the profitability of customer

As can be seen, customer segmentation can be done in many steps of the sales funnel. The key enabler for segmentation is data about the customers, which is often available for the company, e.g., from CRM, transaction logs or their own web page analytics. The data can also be external, e.g., from social media or public register. This type of data can be thought of as "indirect", in that the customers to be segmented do not have to explicitly generate the data themselves. Another approach is to use "direct" data, which can be surveys or questionnaires to customers. In this section, the former approach is focused on more closely.

Machine learning methods are used to build a model based on the data. The methods are usually divided into two: supervised and unsupervised methods. Supervised methods require annotations to the data, e.g. whether a customer lead opportunity is won or lost, or the price obtained for a marketed product in a certain segment. Conversely, unsupervised methods generate structure from non-labelled data, which can then be used for posterior evaluation.

In Table 1, research articles found using relevant search terms in the Web of Science portal are enumerated with their associated details. The articles are categorized according to which of the above approaches are used, what is the target for segmentation, what data is used, and whether supervised (classification, mostly) or unsupervised (clustering) method is used.

As can be seen, utilizing the behaviour-based approach (13 articles) seems to be the most common way of segmenting customers, followed by firmographics (10) and profitability (9). Needs-based classification is the least commonly used approach (6). What comes to the target, the most occurring segmentation goal is to evaluate the loyalty or churn of customers. Ranking the leads and predicting the win/loss of each opportunity are equally uncommon, while papers predicting the profitability of the customer are rare. The studies most commonly use supervised methods (classification, or in one case regression), while unsupervised methods constitute a fourth of all articles. Mostly, transaction data is used, which can be used for extracting buying behaviour-related features, such as the famous Recency, Frequency, Monetary or RFM-method for analysing customer value. RFM, or its expansion, has been used in many articles [27],[28],[31],[32],[34],[36],[38],[47].

The methods used were also varied. Classical K-means clustering or its extension was used in all but one article [41] utilizing unsupervised approach, in which Hidden-Markov-Model-based clustering was used. In addition, methods such as Fuzzy C-means, CLARA and PAM were used [34]. In classification tasks, Random Forest [33], [35], [37], [44], [48], artificial neural networks [31], [32], [33], [35], [45], decision trees [31], [35], [36], [39], [48], logistic regression [32], [38], [39], [40], [45],[48], support vector machines [32], [33], [35], XGBoost and LGBM [26], C4.5 [38], [42], multilayer perceptron [38], naïve Bayes [35], [37], and bagged tree [39] were used. More elaborate approaches were also utilized: mixed-logit model [37], finite mixture regression model [47], optimization with XGBoost [29], as well as weighted alternative least squares and Bayesian personalized ranking [46].

Table 1. B2B customer segmentation research articles and their details. In the strategy column, the different segmentation frameworks are abbreviated: F: Firmographics, N: Needs-based, B: Behaviour-based, P: Profitability-based, and O: Other.

	Approach							
Article	**F**	**N**	**B**	**P**	**O**	**Target**	**Data**	**Supervised / unsuper-vised**
[26]	x				x	Win/loss	CRM	S
[27]			x	x		Churn/loyalty	transaction	U
[28]	x		x			Churn/loyalty	transaction, firmographic data	U
[29]		x		x		Churn/loyalty	transaction, other	S
[30]	x		x			Win/loss	web page behaviour, firmo-graphic data	S & U
[31]				x		Churn/loyalty	transaction	S & U
[32]	x	x	x	x		Churn/loyalty	transaction	S
[33]	x		x	x		Win/loss	CRM	S
[34]			x	x		Churn/loyalty, lead ranking	transaction	U
[35]	x	x	x			Win/loss	CRM + other	S
[36]			x			Churn/loyalty	transaction	S
[37]					x	Win/loss	CRM + other	S
[38]			x	x		Churn/loyalty	transaction	S
[39]				x		Lead ranking, profitability	web page + other	S
[40]	x	x				Lead ranking	transaction	S
[41]			x			Other	transaction, web page behav-iour	U
[42]	x	x				Other	CRM, ERP	S
[43]			x			Other	web page behaviour	-
[44]	x				x	Lead ranking	web data, social media, com-mercial data	S
[45]					x	Churn/loyalty	unclear	S
[46]		x				Win/loss	transaction	S
[47]			x	x		Profitability	transaction	S
[48]	x		x			Churn/loyalty	CRM, web data, transaction, software use data	S

4.6 Fuzzy Logic based pricing methods

A literature research on **fuzzy logic -based pricing** methods led to the following documents of interest for the InnoSale project:

- The authors describe a system to evaluate companies by use of fuzzy logic. They process quantitative and qualitative information and consider 29 parameters. The result is a value in the interval [0, 1]. The higher value, the better is the ranking of the company. Thus a comparison of companies becomes possible. Simple IF-THEN rules have been used without rule weights [137].

- Joint venture (JV) companies combine the power of multiple companies. A constantly existing problem is how the product price can be shared among the participating companies. The author uses a combination of fuzzy logic and game theory to estimate acceptable price revenues. Simple IF-THEN rules have been used. [138]
- The authors modelled consumers of so-called consumer goods (soft drink is used as an example). The model provides estimates on whether the pricing for a new product can be made by considering the same customer habits and behaviours even if there are bigger challenges for the customers caused, e.g., by economic recessions. The model of the authors includes fuzzy logic as well as microeconomic modelling, utility factors, indifference curves and an experimental hierarchic clustering model. [139]
- The authors describe optimized pricing based on fuzzy logic regarding telecommunication service prices. Thus, the prices of the competitors are visible in the pricing process. The pricing factors here are for example the network capacity of the company or the network quality. Simple IF-THEN rules were used. [140]
- Real-time bidding is a mechanism web publishers use to sell their advertisement space. Publishers must set a reserve price to create a minimum bid. In this non-stationary environment, the authors propose a method called FL-ETC-RAP (fuzzy-logic-based explore-then-commit algorithm with risk-aware pricing) to learn the optimal reserve price. Since there is no information about bids and bid history, the method combines an adaptive search procedure and fuzzy logic pricing. Seven IF-THEN rules are used in the method. [50]
- The authors of the article express the benefits of an algorithm developed for the pricing process that reduces the cost, the regression-based linear performance pricing (LPP) algorithm. The results in the article indicate that fuzzy logic can be used for reducing the cost. In the article, the authors show that the pricing problems can be solved using intelligent approaches which can be considered as alternatives to classical mathematical models. The authors mentioned that they suppose that the benefits gained from this study include many conversations across the supply chain network including the transparency of cost factors. They also claim that this study can be extended to apply to all types of industries. [51]

Key takeaways for the InnoSale project:

- [137] and [138] show a good approach to edit fuzzy logic rules in tabular form. This could be a starting point for the optional fuzzy logic rules editor.
- The approach of [139] would make it necessary to interview many customers, which is not applicable in the InnoSale project, partly because the products which industrial customers aim for is much more customer-specific than a consumer good like a soft drink.
- All approaches use simple IF-THEN rules. The InnoSale project should investigate how to solve the pricing of complex industrial products by using those simple rules. However, the first discussions in the consortium with sales engineers seem to make priority rules and weighted rules necessary.

We made follow-up research on available **fuzzy logic technologies**, which could be used for the implementation of the fuzzy-logic-based pricing component in the InnoSale project:

- IEC 61131[58] is the standard of programming languages for industrial programmable logic controllers (PLCs). The standard describes all necessary elements of a fuzzy logic system like fuzzy term definitions, fuzzyfication and defuzzyfication algorithms.
- The Fuzzy Logic Toolbox[59] is a commercial tool by the MathWorks corporation. The toolbox includes possibilities for modelling linguistic variables, membership functions, fuzzyfication and defuzzyfication algorithms. A graphical fuzzy logic designer is also available.

Discussion and decisions:

Both, the standard (IEC 61131-7) and the commercial solution (MathWorks FLT) provide broader functionality than what is used in the literature survey for pricing. This makes it a very flexible and a good candidate solution for the pricing technologies to be used in InnoSale. However, the commercial solution is very expensive. The development of an open-sourced fuzzy logic system with the functionality of e.g. IEC 61131-7 would be an easy-to-use contribution to the European research community.

5 User knowledge acquisition and interaction

5.1 UI development approaches for user knowledge acquisition and user dialog

UI development approaches can be studied from several points of view based on different techniques, patterns or solutions applied. Although this could be an extensive list of different approaches, we summarize below some of the concepts that we find more relevant for this topic in the modern environment.

5.1.1 Frontend frameworks

UI development approaches are based on several alternatives depending on the development effort, design flexibility, and communication with other back-end services, platforms or architecture layers. In terms of technology used for UI frontend development, there are multiple frameworks like the well-known React[60], Angular[61], Vue.js[62], etc., that mainly differ in the paradigm used within the development design abstraction layer. It is known that the use of one or another has been a decision based on trends over the last years, but some studies show that frontend development may be entering a more stable phase.[63] Some of the solutions that seem to rise in popularity among the use of those frameworks are Accessibility[64], Component-driven development[65], cross-platform applications[66] and server-side rendering[67].

[58] IEC 61131-7, Part 7: Fuzzy control programming. First edition 2000.

[59] MathWorks Help Center: Fuzzy Logic Toolbox. https://de.mathworks.com/help/fuzzy/

[60] https://reactjs.org/

[61] https://angular.io/

[62] https://vuejs.org/

[63] https://tsh.io/state-of-frontend

[64] https://accessibility.digital.gov/front-end/getting-started/

[65] https://www.componentdriven.org/

[66] https://www.sam-solutions.com/blog/cross-platform-mobile-development

[67] https://prerender.io/what-is-srr-and-why-do-you-need-to-know/

5.1.2 E-commerce frontend component libraries

As discussed, the use of component-driven development has increased during the last years. It offers ready UI solutions for some specific and common designs, which facilitates reusability, provides a uniform look and feel, and serves prepared interfaces with other components and backend platforms and services. In particular, for eCommerce, multiple component libraries can be found depending on the frontend frameworks used. For example, ReactStorefront[68] for React, Storefront UI[69] for Vue.js, or Intershop[70] for Angular. One of the key points for using component libraries is not only faster UI development but also the ready-made interfaces or connectors that they provide to interact with all major eCommerce Backends systems and Customer Relationship Management (CRM) platforms.

5.1.3 Progressive Web Apps (PWAs)

Progressive web apps are one of the latest trends in UI development. We also observe within the eCommerce frontend component libraries shown in the paragraph above to be one of the most mentioned advantages. PWAs link properties from websites with many features of native mobile apps while reducing the development effort for each specific platform where it is installed. It basically allows for the application to be detached from its operating system, functioning as a website and an application at the same time, also giving the users an experience on par with native apps. [71] They provide a number of advantages such as being installable, progressive enhancement support, responsively designed, re-engageable, linkable, discoverable, network independent, and providing a secure delivery mechanism[72].

5.1.4 Micro Frontends

The term micro frontend extends the concept of microservices where the application is structured as a combined collection of independent services but applied to the user interface. This development pattern is especially useful when separate teams develop different parts of the application, or it also expands at a level that is difficult to maintain as a unit application. With this idea, the application is structured as a composition of different features handled by independent teams each of which represents a distinct area of business that they are interested and specialized in. Each of these pieces could be implemented using different technology stacks, which makes it easier to integrate different ones. This approach virtually eliminates the lock-in phenomenon to a particular technology, or at least reduces it considerably. This is because each team can always decide to opt for a new technology stack without having to translate what was developed previously. In addition, each chunk of the micro frontend's architecture consists of is surely smaller than a frontend monolith; translating it into a new technology would take less time[73].

---

[68] https://www.reactstorefront.io

[69] https://www.vuestorefront.io/storefront-ui

[70] https://www.intershop.com

[71] https://developer.mozilla.org/en-US/docs/Web/Progressive_web_apps/Introduction

[72] https://inoxoft.com/blog/benefits-of-progressive-web-apps-pwa-advantages-and-disadvantages/

[73] https://micro-frontends.org/

5.1.5 Configure, Price, Quote

Configure, Price, Quote (CPQ) refers to a software system that producers and sales teams use to craft personalized configurations, according to customer needs, of complex products and services that have a huge number of interconnected options. Consequently, the system then calculates accurate price quotes in an automated manner based on established pre-programmed rules using the help of AI algorithms and delivers generated and tailored proposals to the clients. [74,75]

This system enhances especially the sales process not only by reducing the overall quotes configuration and pricing errors increasing the forecast accuracy but also by improving the average speed of generating a quote by reducing the approval and close time.

CPQ solutions can also provide a guided selling experience. Guided selling is digital interaction based on eliciting the customer's needs in order to quickly navigate through the product catalogue, sort out options and provide relevant recommendations. It works via the principle of in-person communication between a user and a salesperson: the system asks targeted questions to guide buyers toward the right product. At this stage, a program helps clients get exactly what they need and simultaneously detects additional sales opportunities. For example, by analysing individual configurations, CPQ algorithms determine what complimentary items or related products can be offered with a particular order, thus promoting cross-sales. Furthermore, it is also common to integrate the relative business CRM system to the CPQ solution, which also improves the management of the entire workflow from lead acquisition to closing a deal from a single source. Figure 7 outlines relevant companies and products in the field of CPQ.

[74] https://www.salesforce.com/products/cpq/resources/what-is-cpq/

[75] https://www.sam-solutions.com/blog/what-is-cpq/

Figure 7. Magic Quadrant for CPQ Applications Suites as of June 2021[76].

In conclusion, we can mention different advantages and disadvantages depending on the UI development strategy to be adopted. Tailored UI development will make it easier to adapt and customize to the specific needs but will require more effort to implement. Another solution based on other straightforward and ready-made solutions will require less effort in terms of development at the cost of a more constrained design. It is also important to notice that these approaches are not mutually exclusive, and a combination of different approaches might be used too.

5.2 Solutions for capturing user needs and providing recommendations

In [141] the authors claim that recommendations based on short-term sessions can help suggest a series of items that match the user's preferences. However, the following aspects are not taken into account: First, a user's micro-behaviours, such as item browsing, clicks, item collection and item purchase provide detailed information and can help infer user intent. Second, the first item in a session usually implies the user's original interest. Therefore, they propose a model called AMAN (Adaptive Multi-Attention Network) that considers both the user's micro-behaviours and their interest in the original product for session-based recommendations. There are three key components in AMAN:

- A Directed Graph Attention Network (DGAN): determines the associations between items in the same sequence of micro-behaviours.
- A transformer-based operation-level attention network (TOAN): determines the associations between items in different sequences of micro-behaviours.

[76]https://www.tacton.com/cpq-blog/tacton-named-a-visionary-in-the-2021-gartner-magic-quadrant-for-configure-price-and-quote-application-suites/

- A micro-behavioural joint attention network (MBCAN): determines the co-dependency between different sequences of micro-behaviour from the item level, and explicitly models the first item and the last one.

Finally, based on the trilinear combination of the target behaviour sequence representation, the auxiliary behaviour sequence representation and the item representation, the probability of buying the next item is predicted. This model was evaluated with three data sets, obtaining results superior to those presented in the state-of-the-art since, unlike the others, several types of behaviour and their dependencies are considered.

Entezari et al. propose an intelligent cloud-based customer service system (Intellitag) [165], in this system, a series of clickable tags are presented to the user, thereby capturing the intent of the user's question and solving the problem of lack of personal customer information. One of the problems it solves, according to the authors, is the selection of the labels to be presented to the user. In this system, a multi-task learning model based on BERT is used to extract the most relevant tags for the user. This paper presents data construction, model design, implementation, and experiment evaluation, so it can be used as a reference in similar works.

In [142] the authors claim that in order to satisfy customer needs, it is vital to provide personalized recommendations to customers and thus facilitate the shopping experience. Therefore, they propose a tensor-based method to provide product recommendations to customers. Using three-mode tensors, they can model the product-product interaction within each basket for different users. The decomposition of this tensor allows them to find latent components that reveal the product-product and user-product interaction to infer personalized complementary recommendations. In the evaluation of the proposal, the Instacart public dataset was used, showing good results under conditions with no prior information about the purchases made by the customers. Under these conditions, non-personalized recommendations were generated, rating the products only according to their insertion in the basket.

5.3 Tools for planning and configuring products

Layouts for material flow and manufacturing systems are created as 2D drawings in most companies. The most widely used tool for this purpose, especially in medium-sized companies, but also in corporate groups, is AutoCAD. The pioneer for 3D factory planning is the automotive industry. Here, Microstation software with special libraries tailored to the automotive industry has been used for years.

Various manufacturers of conveyor modules provide their customers with product configurators free of charge. The configurators are often part of the web frontend of the suppliers, e.g., DEMAG (hoists) and Interroll (roller conveyors, etc.). See Figure 8 and Figure 9 for visual examples of DEMAG and Interroll configurators, respectively.

Figure 8. DEMAG KBK-configurator[77].

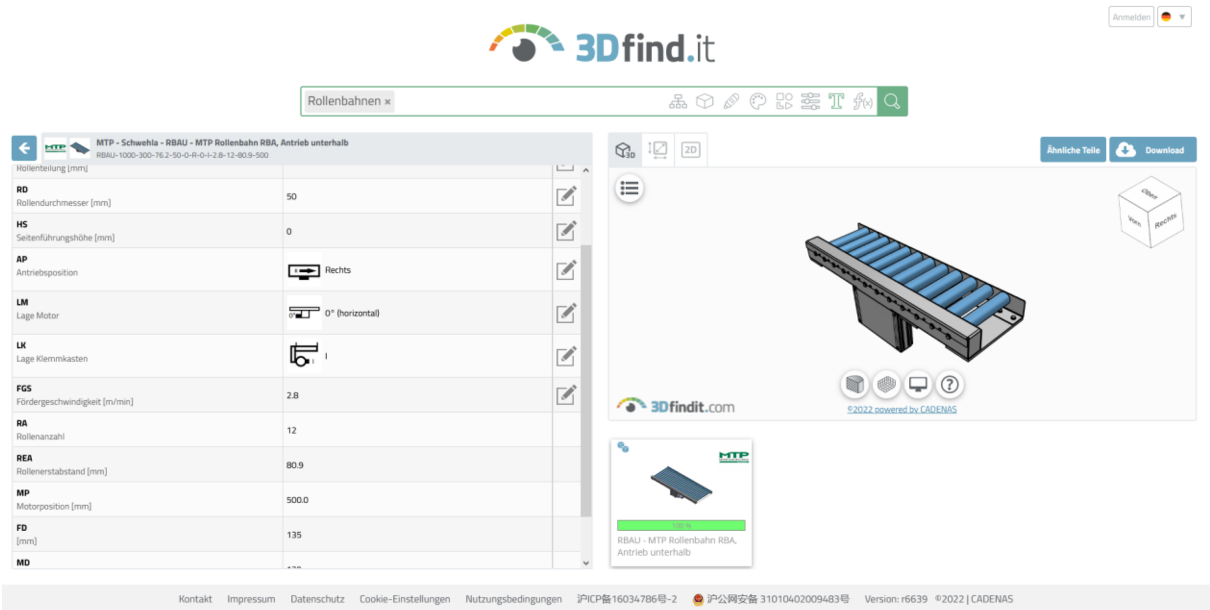

Figure 9. E-catalog from 3Dfind.it.[78]

Depending on the company philosophy, these product configurators are only accessible to verified partners or to all interested parties. Another form of product dissemination via configurators is the use of portals for e-catalogues, such as www.3Dfindit.com from the provider CADENAS. The offerings here include products from a wide range of industries, including materials handling technology and automation technology components. As

[77] http://www.kbkhk.demag-designer.com

[78] https://www.3dfindit.com/digitaltwin?lang=de_DE&path=schwehla%2Fmtp_rollenbahnen%2Frollenbahnen_angetrieben%2Frbau_asmtab.prj&lineid=10&linesubid=-1&varsettransfer=AAACdHiclZLJDoIwEEDvfgVylwrG5TA2QQE11iVAMB6L9tAEMWHR37dSl5R4qJfO63TeTJoWut5uHh_2F3vhHTlPCySji74w5ELGOnil9z48aKUsSpaZsGBsJzJtfVGdt9QC9sUlGdCuzZMit3kjF4bkKzmpUIQ3CQiCGhBadpJo639MJwSAA1AE2B6N8XEyQ_2F47u4pfmzljfQ0mJVGo8sR8MKPVUb6oxaRqqk47h71Qk1HLL5fw5Zq85KwwkWrQs51uSXhr5PjT4_2FCj8Ai0XQtw_3D_3D

configuration results, the user can download documentation and drawings or 3D models of the products in addition to an article number.

A few suppliers also provide their customers with product libraries as CAD add-ons or stand-alone tools. Herein, there are for example extensive 2D CAD blocks for AutoCAD from the supplier TGW. Interroll also offers its sales partners and customers an AutoCAD add-on, a tool for plant design that already contains configuration rules for Interroll products (see Figure 10). This means that the technical components of the plant are already specified in the layout. Depending on the type of contractual relationship, even price information can be obtained.

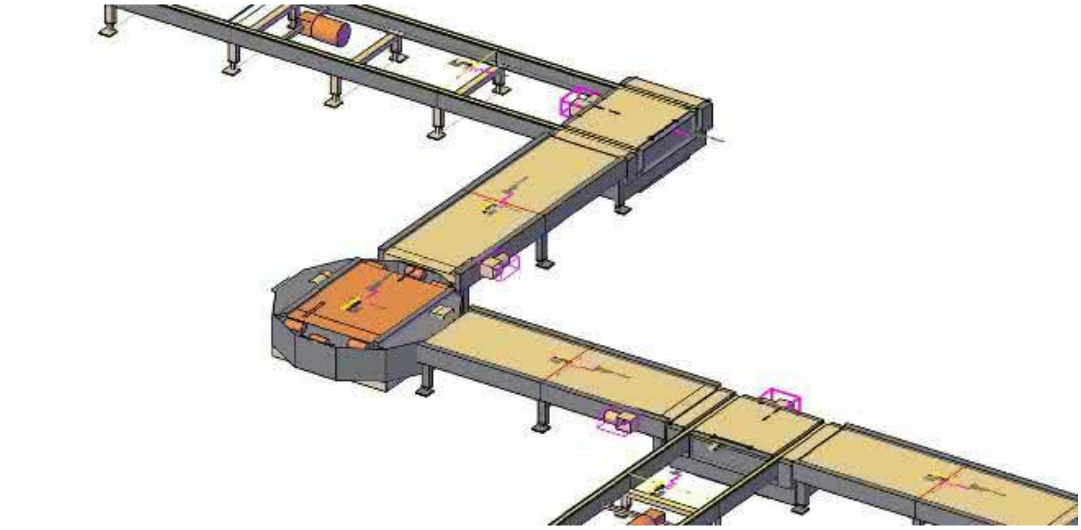

Figure 10. Result of Interrolls Layouter (AutoCAD addOn)[79].

Furthermore, industry-specific planning and visualization tools exist in which parameterizable generic model libraries for production machinery (e.g. Vistable), intralogistics and warehouses (e.g. taraVRbuilder) or automation technology systems (e.g. Visual Components) can be supplemented or replaced by manufacturer-specific modules. Product configurators are often available as an online tool. In contrast, tools that also enable the design of a plant layout are desktop installations.

[79]https://www.materialfluss.de/fordertechnik-und-komponenten/materialflussloesungen-effektiver-planen.htm

6 InnoSale contribution beyond state-of-the-art

The main objective of the InnoSale project is to develop novel, data-driven solutions to support the B2B sales process of complex industrial products aiming at both improved customer experience and smooth sales process. In InnoSale research and development work, several topics are addressed from various perspectives to be able to benefit from the recent technological advances. In this section, we provide a short summary of the anticipated progress beyond the state-of-the-art in the InnoSale project.

At the service design and use case identification level, the InnoSale project aims to identify tasks where AI-driven solutions provide real value for the user and the customer. This is paramount, as the adoption of AI-driven solutions necessitates both investments for the development and data access as well as resources for usage. We believe that there is great potential for utilization of large data pools (historical sales data both in structured and unstructured format) with AI methods to be able to extract information to support the sales experts and to input actionable insight to the product configurators, automating selected sales tasks.

The B2B sales organizations already use a variety of systems, tools and services that collect and process data. The InnoSale tools will be modular, interfacing with the current B2B sales systems and able to use data from external services as well. All the InnoSale features will be built on data sources that are accessible, reliable and contain representative data. Thus, data sharing and aggregation have an important role in the research and development work. InnoSale will increase the flexibility of product configurators by introducing an extensible application design, e.g., a knowledge acquisition component and an inference engine that support different types of logic for data evaluation. These AI-driven configurators will be easy to use and able to adapt — both according to the customer requirements and the sales expert knowledge as well as their preferences. Special emphasis will be given to improved visualization functionalities. The configurator will for example generate specifications or models for several product alternatives and presentations using 3D-enabled technologies

AI methods will be developed to increase customer understanding via novel customer segmentation methods. For example, AI methods can identify similarities of current customer requests to former requests, offerings or projects by searching various heterogeneous data sources, such as ERP, CRM and CAD systems. Tools for automated semantic transformation between domain-specific vocabularies used by customers and manufacturers will be developed (ontology-based creation of custom product configurations). The sales expert will be equipped with novel tools supporting them in the pricing process (estimating optimal prices, considering various customers and domain-specific challenges, regional criteria, and production capacities of the producer) and offer-negotiation phase (NLP methods to identify facts supporting the offer content).

7 Conclusion

In this deliverable, we provide both scientific and industrial state-of-the-art information regarding different technologies related to various phases along the B2B sales funnel. The B2B sales process of complex configurable industrial products is often as complex as the products to be sold. There are manifold opportunities for improvement via the adoption of novel AI-driven solutions.

The research and development process in InnoSale project presumes both understanding of the current B2B sales processes and the tools used as well as technology expertise from software development to AI research. In InnoSale project, our use case partners possess first-hand knowledge of current ways of working and in this deliverable; we provide a wider perspective, especially on the adoption and potential of AI-driven solutions along the B2B sales funnel. Combining the results of the literature review with the user requirements derived from our project partners, we will be able to identify the optimal targets for the development of AI methods regarding knowledge base design, existential rule-reasoning, semantic search, named entity recognition, customer segmentation and fuzzy logic -based pricing methods.

In the context of InnoSale, the product configurator has a central role. Interaction with different internal and external systems must be reliable. For the users, the configurator must provide a smooth user experience with novel 3D visualization solutions. The configurator should also be able to utilize the actionable data-driven insight provided by the AI methods to automate specific phases of the sales process.

8 Abbreviations

AI	Artificial Intelligence
AMAN	Adaptive Multi-Attention Network
API	Application Programming Interface
AR	Artificial Reality
ASP	Answer Set Programming
BDVA	Big Data Value Association
BERT	Bi-directional Encoder Representation from Transformer
BOW	Bag of Words
B2B	Business to business
B2C	Business to Customer
CPQ	Configure, Price, Quote
CRM	Customer Relation Management
DAX	Data Analysis Expressions
DL	Deep Learning
DGAN	Directed Graph Attention Network
EDI	Electronic Data Interchange
ERP	Enterprise Resource Planning
ESB	Enterprise Service Bus
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
IDS	International Data Spaces
IIoT	Industrial Internet of Things
iPaaS	Integration Platform as a Service
JSON	JavaScript Object Notation
KB	Knowledge Base
KBC	Knowledge Base Construction
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory
MES	Manufacturing Execution System
ML	Machine Learning
MLNN	Multi-layer Neural Networks
MQTT	Message Queuing Telemetry Transport
NER	Named Entity Recognition
NLP	Natural Language Processing
OCR	Optical character recognition
OEM	Original Equipment Manufacturer
OPC-UA	Open Platform Communications – Unified Architecture
OWL	Web Ontology Language

PAL	Predictive Analysis Library
PWA	Progressive Web Apps
QA	Question Answering
RDF	Resource Description Framework
REST	Representational State Transfer
RFM	Recency, Frequency, Monetary
RNN	Recurrent Neural Networks
ROI	Return of Investment
SFA	Sales Force Automation
SFMC	Salesforce Marketing Cloud
SLA	Service Level Agreements
SME	Small and Medium-sized Enterprises
STP	Segmentation, Targeting, and Positioning
TOAN	Transformer-based Operation-level Attention Network
UI	User Interface
VR	Virtual Reality

9 References

[1] Syam, Niladri; Sharma, Arun. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Industrial marketing management, 2018, 69: 135-146.

[2] Cuevas, Javier Marcos. The transformation of professional selling: Implications for leading the modern sales organization. Industrial Marketing Management, 2018, 69: 198-208.

[3] Hall, Kelly R., et al. Understanding salesperson intention to use AI feedback and its influence on business-to-business sales outcomes. Journal of Business & Industrial Marketing, 2021

[4] Mattila, Malla; Yrjölä, Mika; Hautamäki, Pia. Digital transformation of business-to-business sales: what needs to be unlearned? Journal of Personal selling & sales Management, 2021, 41.2: 113-129

[5] Prieto, Fábio; BRAGA, H. Contribution of Artificial Intelligence in B2B Sales: A Danfoss Case Study. International Journal of Advanced Engineering Research and Science, 2021, 8.4: 225-234.

[6] Saura, Jose Ramon; Ribeiro-Soriano, Domingo; Palacios-Marqués, Daniel. Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. Industrial Marketing Management, 2021, 98: 161-178

[7] Tura, Amanuel Diriba; Mamo, Hana Beyene; Gemechu, Wendimu Fanta. Mathematical modeling and parametric optimization of surface roughness for evaluating the effects of fused deposition modeling process parameters on ABS material. International Journal of Advanced Engineering Research and Science, 2021, 8.5: 049-057

[8] Dubinsky, Alan J. A factor analytic study of the personal selling process. Journal of personal selling & sales management, 1981, 1.1: 26-33.

[9] Fraccastoro, Sara; Gabrielsson, Mika; Pullins, Ellen Bolman. The integrated use of social media, digital, and traditional communication tools in the B2B sales process of international SMEs. International Business Review, 2021, 30.4: 101776.

[10] Bowen, Melanie, et al. Social media in B2B sales: Why and when does salesperson social media usage affect salesperson performance?. Industrial Marketing Management, 2021, 96: 166-182

[11] Paschen, Jeannette; Wilson, Matthew; Ferreira, João J. Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. Business Horizons, 2020, 63.3: 403-414.

[12] Jelinek, Ronald. All pain, no gain? Why adopting sales force automation tools is insufficient for performance improvement. Business Horizons, 2013, 56.5: 635-642.

[13] Kristjansdottir, Katrin, et al. The main challenges for manufacturing companies in implementing and utilizing configurators. Computers in Industry, 2018, 100: 196-211.

[14] Singh, Jagdip, et al. Sales profession and professionals in the age of digitization and artificial intelligence technologies: concepts, priorities, and questions. Journal of Personal Selling & Sales Management, 2019, 39.1: 2-22.

[15] Mahlamäki, Tommi, et al. Adoption of digital sales force automation tools in supply chain: Customers' acceptance of sales configurators. Industrial Marketing Management, 2020, 91: 162-173

[16] Blazek, Paul; Pilsl, Klaus. The importance of analyzing configurator usage data to leverage product innovation and sales performance. 9th International Conference On Mass Customization and Personalisation, Sep. 23 -25. 2020

[17] Leszkiewicz, Agata; Hormann, Tina; Krafft, Manfred. Smart Business and the Social Value of AI. In: Smart Business and Better Management. Emerald Group Publishing Ltd., 2021.

[18] Schwartz, Reva, et al. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. 2022. NIST Special Publication 1270.

[19] Ntoutsi, Eirini, et al. Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020, 10.3: e1356.

[20] Rohaan, D.; Topan, E.; Groothuis-Oudshoorn, C. G. M. Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. Expert Systems With Applications, 2022, 188: 115925.

[21] Frost & Sullivan, "Global Retail Artificial Intelligence (AI) Growth Opportunities - AI Solutions will Drive the Future Growth Potential of the Retail Space for Technology Providers," 2022.

[22] Frost & Sullivan, "Innovations in Artificial Intelligence, Internet of Things, and Natural Language Processing," 2020.

[23] Frost & Sullivan, "A Benchmarking System to Spark Companies to Action — Innovation that Fuels New Deal Flow and Growth Pipelines," 2022.

[24] Frost & Sullivan, "Innovations in AI, Blockchain, and Business Intelligence," 2019.

[25] Frost & Sullivan, "The Role of Bots in Financial Services," 2018.

[26] Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. Forecasting, 2(3), 267–283. https://doi.org/10.3390/forecast2030015

[27] Sheikh, A., Ghanbarpour, T., & Gholamiangonabadi, D. (2019). A Preliminary Study of Fintech Industry: A Two-Stage Clustering Analysis for Customer Segmentation in the B2B Setting. Journal of Business-to-Business Marketing, 26(2), 197–207. https://doi.org/10.1080/1051712X.2019.1603420

[28] Kandeil, D. A., Saad, A. A., & Youssef, S. M. (2014). A two-phase clustering analysis for B2B customer segmentation. Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014, 221–228. https://doi.org/10.1109/INCoS.2014.49

[29] Janssens, B., Bogaert, M., Bagué, A., & Van den Poel, D. (2022). B2Boost: instance-dependent profit-driven modelling of B2B churn. Annals of Operations Research. https://doi.org/10.1007/s10479-022-04631-5

[30] Yang, J., Liu, C., Teng, M., Liao, M., & Xiong, H. (2016). Buyer targeting optimization: A unified customer segmentation perspective. Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, 1262–1271. https://doi.org/10.1109/BigData.2016.7840730

[31] Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 37(7), 5259–5264. https://doi.org/10.1016/j.eswa.2009.12.070

[32] Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. Industrial Marketing Management, 62, 100–107. https://doi.org/10.1016/j.indmarman.2016.08.003

[33] Rotovei, D., & Negru, V. (2019). Data Driven Sales Prediction Using Communication Sentiment Analysis in B2B CRM Systems. Proceedings - 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2019, 174–181. https://doi.org/10.1109/SYNASC49474.2019.00032

[34] Maulina, N. R., Surjandari, I., & Rus, A. M. M. (2019). Data mining approach for customer segmentation in B2B settings using centroid-based clustering. 2019 16th International Conference on Service Systems and Service Management, ICSSSM 2019, 2–7. https://doi.org/10.1109/ICSSSM.2019.8887739

[35] Bohanec, M., Kljajić Borštnar, M., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. Expert Systems with Applications, 71, 416–428. https://doi.org/10.1016/j.eswa.2016.11.010

[36] Jahromi, A. T., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. Industrial Marketing Management, 43(7), 1258–1268. https://doi.org/10.1016/j.indmarman.2014.06.016

[37] Colias, J. V., Park, S., & Horn, E. (2021). Optimizing B2B product offers with machine learning, mixed logit, and nonlinear programming. Journal of Marketing Analytics, 9(3), 157–172. https://doi.org/10.1057/s41270-021-00113-y

[38] Chen, K., Hu, Y. H., & Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. Information Systems and E-Business Management, 13(3), 475–494. https://doi.org/10.1007/s10257-014-0264-1

[39] D'Haen, J., Van Den Poel, D., & Thorleuchter, D. (2013). Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. Expert Systems with Applications, 40(6), 2007–2012. https://doi.org/10.1016/j.eswa.2012.10.023

[40] Yan, J., Gong, M., Sun, C., Huang, J., & Chu, S. M. (2015). Sales pipeline win propensity prediction: A regression approach. Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015, 854–857. https://doi.org/10.1109/INM.2015.7140393

[41] Liu, C., Zhang, K., & Xiong, H. (2015). Sequential pattern analysis with right granularity. IEEE International Conference on Data Mining Workshops, ICDMW, 2015 (January), 1178–1184. https://doi.org/10.1109/ICDMW.2014.164

[42] Gentner, D., Stelzer, B., Ramosaj, B., & Brecht, L. (2018). Strategic Foresight of Future B2B Customer Opportunities through Machine Learning. Technology Innovation Management Review, 8(10), 5–17. https://doi.org/10.22215/timreview/1189

[43] Sekiguchi, A., & Tsuda, K. (2014). Study on web analytics utilizing segmentation knowledge in business to business manufacturer site. Procedia Computer Science, 35(C), 902–909. https://doi.org/10.1016/j.procs.2014.08.164

[44] Meire, M., Ballings, M., & Van den Poel, D. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. Decision Support Systems, 104, 26–37. https://doi.org/10.1016/j.dss.2017.09.010

[45] Jamjoom, A. A. (2021). The use of knowledge extraction in predicting customer churn in B2B. Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00500-3

[46] Vlachos, M., Vassiliadis, V. G., Heckel, R., & Labbi, A. (2016). Toward interpretable predictive models in B2B recommender systems. IBM Journal of Research and Development, 60(5–6), 1–12. https://doi.org/10.1147/JRD.2016.2602097

[47] Mark, T., Niraj, R., & Dawar, N. (2012). Uncovering Customer Profitability Segments for Business Customers. Journal of Business-to-Business Marketing, 19(1), 1–32. https://doi.org/10.1080/1051712X.2011.593022

[48] De Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K., & Phan, M. (2021). Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. Industrial Marketing Management, 99(October), 28–39. https://doi.org/10.1016/j.indmarman.2021.10.001

[49] Hannah Bast, Björn Buchhold and Elmar Haussmann (2016), "Semantic Search on Text and Knowledge Bases", Foundations and Trends® in Information Retrieval: Vol. 10: No. 2-3, pp 119-271. http://dx.doi.org/10.1561/1500000032

[50] Rhuggenaath, J., Akcay, A., Zhang, Y., & Kaymak, U. (2019). Fuzzy Logic based Pricing combined with Adaptive Search for Reserve Price Optimization in Online Ad Auctions. 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). doi:10.1109/fuzz-ieee.2019.8858975

[51] Tanak Coşkun, G., & Yılmaz Yalçıner, A. (2021). Determining the best price with linear performance pricing and checking with fuzzy logic. Computers & Industrial Engineering, 154, 107150. doi:10.1016/j.cie.2021.107150

[52] Johnsson, Charlotta. "ISA 95-how and where can it be applied." Technical Papers of ISA 454 (2004): 399-408.

[53] Åkerman, Magnus. Implementing shop floor IT for Industry 4.0. Chalmers Tekniska Hogskola (Sweden), 2018

[54] Givehchi, Omid, et al. "Interoperability for industrial cyber-physical systems: An approach for legacy systems." IEEE Transactions on Industrial Informatics 13.6 (2017): 3370-3378.

[55] Nakagawa, Elisa Yumi, et al. "Industry 4.0 reference architectures: State of the art and future trends." Computers & Industrial Engineering 156 (2021): 107241.

[56] P. Drahoš, E. Kučera, O. Haffner and I. Klimo, "Trends in industrial communication and OPC UA," 2018 Cybernetics & Informatics (K&I), 2018, pp. 1-5, doi: 10.1109/CYBERI.2018.8337560.

[57] M. Iglesias-Urkia, A. Orive, M. Barcelo, A. Moran, J. Bilbao and A. Urbieta, "Towards a lightweight protocol for Industry 4.0: An implementation based benchmark," 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), 2017, pp. 1-6, doi: 10.1109/ECMSM.2017.7945894.

[58] Haddara, Moutaz. "ERP systems selection in multinational enterprises: a practical guide." International Journal of Information Systems and Project Management 6.1 (2018): 43-57.

[59] Krizanic, Snjezana, Tanja Sestanj-Peric, and Katarina Tomicic-Pupek. "The changing role of ERP and CRM in Digital Transformation." Economic and Social Development: Book of Proceedings (2019): 248-256.

[60] Hoyle, J.A., Dingus, R. & Wilson, J.H. An exploration of sales forecasting: sales manager and salesperson perspectives. J Market Anal **8,** 127–136 (2020). https://doi.org/10.1057/s41270-020-00082-8

[61] J. Chawla, A. K. Ahlawat and G. Goswami, "A Review on Web Services Interoperability Issues," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2018, pp. 1-5, doi: 10.1109/UPCON.2018.8596930.

[62] Schwee, J.H.; Sangogboyer, F.C.; Kjærgaard, M.B.: Anonymizing building data for data analytics in cross-organizational settings. Proceedings of the International Conference on Interner of Things Design and Implementation, ACM. New York 2019

[63] Benjamin, M.; Gagnon, P.; Rostamzadeh, N.; Pal, C.; Bengio; Y.; Shee, A.: Towards Standardisation of Data Licenses: The Montreal Data License. arXiv preprint arXiv:1903.12262, 2019.

[64] D. Booth, H. Haas, F. McCabe, E. Newcomer, M. Champion, C. Ferris, and D. Orchard. (2004, February) Web services architecture. [Online]. Available: https://www.w3.org/TR/ws-arch/

[65] A. Archip, C. Amarandei, P. Herghelegiu, C. Mironeanu and E. şerban, "RESTful Web Services – A Question of Standards," 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), 2018, pp. 677-682, doi: 10.1109/ICSTCC.2018.8540763.

[66] Halili, Festim, and Erenis Ramadani. "Web services: a comparison of soap and rest services." Modern Applied Science 12.3 (2018)

[67] BDVA Position Paper: Towards a European Data Sharing Space: Enabling Data Exchange and Unlocking AI Potential, 2019

[68] Otto, B., Hompel, M.t., Wrobel, S. (2019). International Data Spaces. In: Neugebauer, R. (eds) Digital Transformation. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-58134-6_8

[69] Borgogno, O.; Colangelo, G.; Data sharing and interoperability: Fostering innovation and competition through APIs. Computer Law & Security Review,Volume 35, Issue 5, 2019.

[70] Bindschaedler, Vincent, Reza Shokri, and Carl A. Gunter. "Plausible deniability for privacy-preserving data synthesis." arXiv preprint arXiv:1708.07975 (2017)

[71] Abiteboul, S., Hull, R., and Vianu, V. Foundations of Databases. Addison-Wesley, 1995.

[72] Baget, J., Leclere, M., Mugnier, M., and Salvat, E. On rules with existential variables: Walking the decidability line. Artif. Intell. 175, 9-10 (2011), 1620–1654

[73] Beeri, C., and Vardi, M. Y. The implication problem for data dependencies. In Automata, Languages and Programming, 8th Colloquium, Acre (Akko), Israel, July 13-17, 1981, Proceedings (1981), S. Even and O. Kariv, Eds., vol. 115 of Lecture Notes in Computer Science, Springer, pp. 73–85.

[74] Gogacz, T., and Marcinkowski, J. All-instances termination of chase is undecidable. In Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II (2014), J. Esparza, P. Fraigniaud, T. Husfeldt, and E. Koutsoupias, Eds., vol. 8573 of Lecture Notes in Computer Science, Springer, pp. 293–304.

[75] Barany, V., Gottlob, G., and Otto, M. Querying the guarded fragment. In Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS

2010, 11-14 July 2010, Edinburgh, United Kingdom (2010), IEEE Computer Society, pp. 1–10.

[76] Gogacz, T., Marcinkowski, J., and Pieris, A. All-instances restricted chase termination. In Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020 (2020), D. Suciu, Y. Tao, and Z. Wei, Eds., ACM, pp. 245–258.

[77] Grau, B. C., Horrocks, I., Krötzsch, M., Kupke, C., Magka, D., Motik, B., and Wang, Z. Acyclicity notions for existential rules and their application to query answering in ontologies. J. Artif. Intell. Res. 47 (2013), 741–808.

[78] Carral, D., Dragoste, I., and Krötzsch, M. Restricted chase (non)termination for existential rules with disjunctions. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017 (2017), C. Sierra, Ed., ijcai.org, pp. 922–928.

[79] Karimi, A.; Zhang, H.; and You, J. 2021. Restricted chase termination for existential rules: A hierarchical approach and experimentation. Th. Pract. Log. Program. 21(1):4–50.

[80] Leclere, M.; Mugnier, M.; Thomazo, M.; and Ulliana, F. 2019. A single approach to decide chase termination on linear existential rules. In Barcelo, P., and Calautti, M., eds., 22nd International Conference on Database Theory, ICDT 2019, 2019, Portugal, volume 127 of LIPIcs, 18:1–18:19. Schloss Dagstuhl - Leibniz-Zentrum f ̈ur Informatik.

[81] M. Benedikt, G. Konstantinidis, G. Mecca, B. Motik, P. Papotti, D. Santoro, and E. Tsamoura. Benchmarking the chase. In Proc. 36th Symposium on Principles of Database Systems (PODS'17), pages 37–52. ACM, 2017.

[82] Calì, A., Gottlob, G., and Lukasiewicz, T. A general Datalog-based framework for tractable query answering over ontologies. In: Journal of Web Semantics 14 (2012), pages 57–83.

[83] Aref M., Cate B., Green, T. J., Kimelfeld, B., Olteanu, D., Pasalic, E., Veldhuizen, T. L., and Washburn, G. Design and implementation of the LogicBlox system. In Proc. 2015 ACM SIGMOD International Conference on Management of Data, pages 1371–1382. ACM, 2015.

[84] Baget, J., Leclère, M., Mugnierm M., Rocher, S., and Sipieter, C. Graal: A toolkit for query answering with existential rules. In Proc. 9th International Web Rule Symposium (RuleML'15), volume 9202 of LNCS, pages 328–344. Springer, 2015.

[85] Benedikt, M., Leblay, J., and Tsamoura, E. PDQ: proof-driven query answering over web-based data. Proc. VLDB Endowment (PVLDB), 7(13):1553–1556, 2014.

[86] Bonifati, A., Ileana, I., and Linardi, M. Functional dependencies unleashed for scalable data exchange. In Proc. 28th International Conference on Scientific and Statistical Database Management (SSDBM'16), pages 2:1–2:12. ACM, 2016.

[87] Geerts, F., Mecca, G., Papotti, P., and Santoro, D. That's all folks! LLUNATIC goes open source. Proc. VLDB Endowment (PVLDB), 7(13):1565–1568, 2014.

[88] Nenov, Y., Piro, R., Motik, B., Horrocks, I., Wu, Z., and Banerjee, J. RDFox: A highly-scalable RDF store. In Proc. 14th International Semantic Web Conference (ISWC'15), Part II, volume 9367 of LNCS, pages 3–20. Springer, 2015.

[89] Urbani, J., Jacobs, C., and Krötzsch, M. Column-oriented Datalog materialization for large knowledge graphs. In Proceedings of the 30th AAAI Conference on Artificial Intelligence, pages 258–264. AAAI Press, 2016.

[90] Faber, W. An Introduction to Answer Set Programming and Some of Its Extensions. In: Manna, M., Pieris, A. (eds) Reasoning Web. Declarative Artificial

Intelligence. Reasoning Web 2020. Lecture Notes in Computer Science, vol 12258. Springer, Cham.

[91] Gelfond, M., Lifschitz, V. The Stable Model Semantics For Logic Programming. Logic Programming Volume 2, 2000.

[92] Leone, N., Pfeifer, G., Faber, W., Eiter, T., Gottlob, G., Perri, S., Scarcello, F. The DLV System for Knowledge Representation and Reasoning. ACM Transactions on Computational Logic 7(3), 499–562 (2006)

[93] Gebser, M., Leone, N., Maratea, M., Perri, S., Ricca, F., & Schaub, T. (2018, July). Evaluation Techniques and Systems for Answer Set Programming: a Survey. In IJCAI (Vol. 18, pp. 5450-5456).

[94] Gebser. M., Schaub, T., and Thiele, S. Gringo: A new grounder for answer set programming. In LPNMR'07, volume 4483 of LNCS, pages 266–271. Springer, 2007.

[95] Alviano, M., Dodaro, C., Faber, W., Leone, N., and Ricca, F. WASP: A native ASP solver based on constraint learning. In LPNMR13, volume 8148 of LNCS, pages 54–66. Springer, 2013.

[96] Faber, W., Leone, N., and Perri, S. The intelligent grounder of DLV. In Correct Reasoning: Essays on Logic-Based AI in Honour of Vladimir Lifschitz, volume 7265 of LNCS, pages 247–264. Springer, 2012.

[97] Simons, P., Niemelä, I., and Soininen. T. Extending and implementing the stable model semantics. Artif. Intell., 138(1-2):181–234, 2002.

[98] Gebser, M., Kaufmann, B., and Schaub, T. Conflict-driven answer set solving: From theory to practice. Artif. Intell., 187-188:52–89, 2012.

[99] Giunchiglia, E., Lierler, Y., and Maratea, M. Answer set programming based on propositional satisfiability. Journal of Automated Reasoning, 36(4):345–377, 2006.

[100] Lin, F., and Zhao, Y. ASSAT: Computing answer sets of a logic program by SAT solvers. Artif. Intell., 157(1-2):115–137, 2004.

[101] Janhunen, T., and Niemelä, I. Compact translations of non-disjunctive answer set programs to propositional clauses. In LP, KR, and NMR, volume 6565 of LNCS, pages 111–130. Springer, 2011.

[102] Dal Palù, A., Dovier, A., Pontelli, E., and Rossi, G. GASP: Answer set programming with lazy grounding. Fund. Infor., 96(3):297–322, 2009.

[103] Dao-Tran, M., Eiter, T., Fink, M., Weidinger, G., and Weinzierl, A. Omiga: An open minded grounding on-the-fly answer set solver. In JELIA 2012, pages 480–483, 2012.

[104] Weinzierl, A. Blending lazy-grounding and CDNL search for answer-set solving. In LPNMR 2017, pages 191–204, 2017.

[105] Ellmauthaler, S., Krötzsch, M., & Mennicke, S. (2022, June). Answering Queries with Negation over Existential Rules. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 5, pp. 5626-5633).

[106] Baget, J. F., Garcia, L., Garreau, F., Lefèvre, C., Rocher, S., & Stéphan, I. (2018). Bringing existential variables in answer set programming and bringing non-monotony in existential rules: two sides of the same coin. Annals of mathematics and artificial intelligence, 82(1), 3-41.

[107] Eiter, T., Ianni, G., Lukasiewicz, T., Schindlauer, R., Tompits, H.: Combining answer set programming with description logics for the semantic web. Artif. Intell. 172(12–13), 1495–1539 (2008)

[108] Motik, B., Rosati, R.: Reconciling description logics and rules. J. ACM, 57(5) (2010)

[109] Rosati, R.: Dl+log: Tight integration of description logics and disjunctive datalog. In: Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, pp. 68–78. Lake District of the United Kingdom (2006)

[110] Petasis, G., Cucchiarelli, A., Velardi, P., Paliouras, G., Karkaletsis, V., & Spyropoulos, C. D. (2000, July). Automatic adaptation of Proper Noun Dictionaries through cooperation of machine learning and probabilistic methods. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 128-135).

[111] Sharnagat, R. (2014). Named entity recognition: A literature survey. Center For Indian Language Technology, 1-27.

[112] Li, J., Sun, A., Han, J., & Li, C. (2022). A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering, 34(1), 50-70.

[113] Patawar, M. L., & Potey, M. A. (2015). Approaches to named entity recognition: a survey. International Journal of Innovative Research in Computer and Communication Engineering, 3(12), 12201-12208.

[114] Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3-26.

[115] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[116] Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991.

[117] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

[118] Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., & Li, J. (2019). A unified MRC framework for named entity recognition. arXiv preprint arXiv:1910.11476.

[119] Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., & Li, J. (2019). Dice loss for data-imbalanced NLP tasks. arXiv preprint arXiv:1911.02855.

[120] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[121] Yan, H., Deng, B., Li, X., & Qiu, X. (2019). TENER: adapting transformer encoder for named entity recognition. arXiv preprint arXiv:1911.04474.

[122] Baevski, A., Edunov, S., Liu, Y., Zettlemoyer, L., & Auli, M. (2019). Cloze-driven pretraining of self-attention networks. arXiv preprint arXiv:1903.07785.

[123] Ma, X., Xu, P., Wang, Z., Nallapati, R., & Xiang, B. (2019, November). Domain adaptation with BERT-based domain classification and data selection. In Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019) (pp. 76-83).

[124] Liu, Z., Jiang, F., Hu, Y., Shi, C., & Fung, P. (2021). NER-BERT: A Pre-trained Model for Low-Resource Entity Tagging. arXiv preprint arXiv:2112.00405.

[125] Chowdhary, K.R. (2020). Natural Language Processing. In: Fundamentals of Artificial Intelligence. Springer, New Delhi. https://doi.org/10.1007/978-81-322-3972-7_19

[126] Zhang, Y., Jin, R., & Zhou, Z. (2010). Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning And Cybernetics, 1(1-4), 43-52. doi: 10.1007/s13042-010-0001-0

[127] Blei, D., Ng, A. & Jordan, M. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993-1022.

[128] Sutskever, I., Vinyals, O., Le, Q.V. (2014.). Sequence to sequence learning with neural networks. In: Proceedings of the Advances in Neural Information Processing Systems. Montreal, 3104–3112

[129] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[130] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[131] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559.

[132] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234-1240.

[133] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[134] Schick, T., & Schütze, H. (2020). Exploiting cloze questions for few shot text classification and natural language inference. arXiv preprint arXiv:2001.07676.

[135] Tam, D., Menon, R. R., Bansal, M., Srivastava, S., & Raffel, C. (2021). Improving and simplifying pattern exploiting training. arXiv preprint arXiv:2103.11955.

[136] Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

[137] Malagoli, S; Mastroleo, G.; Magni, C.A.: The use of fuzzy logic and expert systems for rating and pricing firms: a new perspective on valuation. Published in: Managerial Finance, Vol. 33, No. 11 (2007): pp. 836-852.

[138] Min-Ren Yan: A fuzzy logic enhanced bargaining model for business pricing decision support in joint venture projects. Journal of Business Economics and Management, 2011 Vol. 12(2): pp. 234-247.

[139] Morim, A; Fortes, E.S.; Reis, P.; Cosenza, C.; Doria, F.; Goncalves, A.: Think fuzzy system: developing new pricing strategy methods for consumer goods using fuzzy logic. International Journal of Fuzzy Logic Systems (IJFLS) Vol. 7, No. 1, January 2017.

[140] Adeolu Olukorede Dairo and Krisztian Szücs: A Fuzzy Logic Expert System for Pricing Digital Services: The Case of Price Adjustment for a Mobile Service Provider. International Journal of Fuzzy Logic and Intelligent Systems Vol. 20, No. 3, September 2020, pp. 227-241 http://doi.org/10.5391/IJFIS.2020.20.3.227

[141] Jingjing Qiao, Li Wang, Modeling user micro-behaviors and original interest via Adaptive Multi-Attention Network for session-based recommendation, Knowledge-Based Systems, Volume 244, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2022.108567

[142] N. Entezari, E. E. Papalexakis, H. Wang, S. Rao and S. K. Prasad, "Tensor-based Complementary Product Recommendation," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 409-415, doi: 10.1109/BigData52589.2021.9671938.

[143] Wang, Y., & Zatarain, O. A. (2018, July). Design and implementation of a knowledge base for machine knowledge learning. In 2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (pp. 70-77). IEEE.

[144] Rajadhyax D., (2020). A Survey of Methods, Tools and Application of Knowledge Base Construction (KBC). Telecom Business Review A Journal of Symbiosis Institute of Digital and Telecom Management, 13(1):20-26. ISSN: 0973-9114. Retrieved from: https://sidtm.edu.in/wp-content/uploads/2021/06/tbr2020.pdf

[145] Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In Twenty-ninth AAAI conference on artificial intelligence.

[146] Ratner, A., Ré, C., & Bailis, P. (2018). Knowledge base construction in the machine-learning era. Communications of the ACM, 61(11), 95-96.

[147] Razniewski, S., Suchanek, F., & Nutt, W. (2016, June). But what do we actually know? In Proceedings of the 5th Workshop on Automated Knowledge Base Construction (pp. 40-44).

[148] Subasic, P., Yin, H., & Lin, X. (2019). Building knowledge base through deep learning relation extraction and Wikidata. AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering.

[149] Oro, E., & Ruffolo, M. (2008). Towards a system for ontology-based information extraction from PDF documents. R. Meersman & Z. Tari (eds), On the move to meaningful internet systems: OTM 2008. Lecture Notes in Computer Science, vol. 5332. Springer, Berlin, Heidelberg.

[150] Oro, E., & Ruffolo, M. (2009). PDF-TREX: An approach for recognizing and extracting tables from PDF documents. 10th International Conference on Document Analysis and Recognition (pp. 906-910).

[151] Shin, J., Wu, S., Wang, F., Sa, C.D., Ratner, A., Zhang, C., & Ré, C. (2016). Incremental knowledge base construction using DeepDive. The VLDB Journal, 26(1), 81-105.

[152] Ratner, A., Bach, S. H., Ehrenberg, H. R., Fries, J. A., Wu, S., & Ré, C. (2017). Snorkel: Rapid training data creation with weak supervision. Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases, 11(3), 269-282.

[153] Wu, S., Hsiao, L., Cheng, X., Hancock, B., Rekatsinas, T., Levis, P., & Ré, C. (2018). Fonduer: Knowledge base construction from richly formatted data. Proceedings of the 2018 International Conference on Management of Data.

[154] Li, Y., Reiss, F., & Chiticariu, L. (2011, June). SystemT: A declarative information extraction system. In Proceedings of the ACL-HLT 2011 System Demonstrations (pp. 109-114).

[155] Lan, Y., He, G., Jiang, J., Jiang, J., Zhao, W. X., & Wen, J. R. (2021). A survey on complex knowledge base question answering: Methods, challenges and solutions. arXiv preprint arXiv:2105.11644. https://doi.org/10.48550/arXiv.2105.11644

[156] Zhang, C., Lai, Y., Feng, Y., & Zhao, D. (2021). A review of deep learning in question answering over knowledge bases. AI Open, 2, 205–215.

[157] Nguyen, D. B., Abujabal, A., Tran, N. K., Theobald, M., & Weikum, G. (2017). Query-driven on-the-fly knowledge base construction. Proceedings of the VLDB Endowment, 11(1), 66–79.

[158] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yangqiu Song, Seung-won Hwang, and Wei Wang. 2017. KBQA: learning question answering over QA corpora and

knowledge bases. Proc. VLDB Endow. 10, 5 (January 2017), 565–576. https://doi.org/10.14778/3055540.3055549

[159] Sa, C.D., Ratner, A.J., Ré, C., Shin, J., Wang, F., Wu, S., & Zhang, C. (2016). DeepDive: Declarative Knowledge Base Construction. SIGMOD record, 45 1, 60-67.

[160] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K.P., Strohmann, T., Sun, S., & Zhang, W. (2014). Knowledge vault: a web-scale approach to probabilistic knowledge fusion. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.

[161] Hur A., Janjua N., Ahmed M. (2021). A Survey on State-of-the-art Techniques for Knowledge Graphs Construction and Challenges Ahead. IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) (2021): 99-103.

[162] Fu B., Qui Y., Tang C., Li Y., Yu H., Sun J. (2020). A Survey on Complex Question Answering over Knowledge Base: Recent Advances and Challenges. arXiv preprint arXiv:2007.13069. https://doi.org/10.48550/arXiv.2007.13069

[163] Färber M., Rettinger A., (2018). Which Knowledge Graph Is Best for Me? "Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO" in a Nutshell. arXiv preprint arXiv:1809.11099. https://doi.org/10.48550/arXiv.1809.11099

[164] Li F., Chen H., Xu G., Qiu T., Ji F., Zhang J., Chen H. (2020). AliMe KG: Domain Knowledge Graph Construction and Application in E-commerce. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20).

[165] N. Entezari, E. E. Papalexakis, H. Wang, S. Rao and S. K. Prasad, "Tensor-based Complementary Product Recommendation," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 409-415, doi: 10.1109/BigData52589.2021.9671938.