



Introducing Semantic AI

Ingredients for a sustainable Enterprise AI Strategy

Table of Contents

AI has reached Enterprises	3
Semantic AI as an advanced perspective on AI	5
Six core aspects of Semantic AI	6
Data Quality	6
Data as a Service	6
No black-box	6
Hybrid approach	6
Structured data meets text	6
Towards self optimizing machines	7
Benefiting from Semantic AI along the Data Lifecycle	8
The need for Semantic Data Lakes	8
Example: Which service requests can we automate?	9
Data Lifecycles and Semantics	9
Semantics and Linked Data along a traditional Data Life Cycle	9
The Linked Data Lifecycle	10
Semantic Knowledge Graphs	11
Technical Deep Dive: How to build Knowledge Graphs	12
Towards Semantic AI	14
Summary / Key statements of this paper	15
References	16

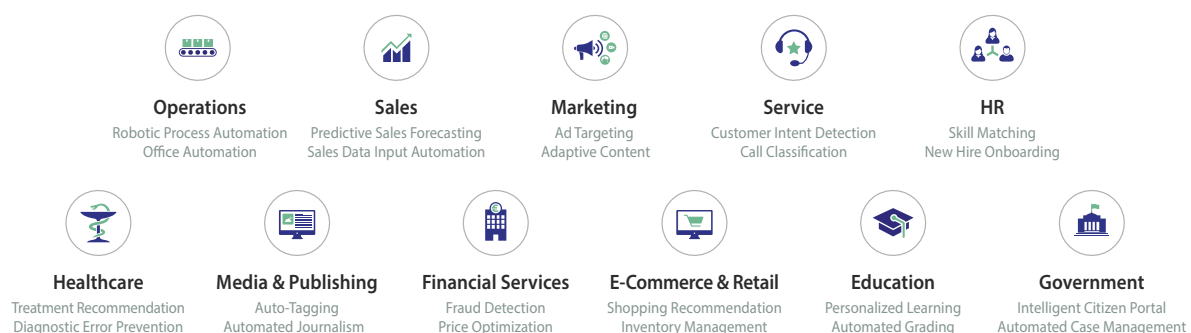
AI has reached Enterprises

In recent years, the topic Artificial Intelligence (AI) became dominant as technology trend and is now influencing almost every business domain. The professional world is changing. Almost everybody from business over IT to subject matter experts is obliged to understand to a certain degree how this technology functions and has to start to question his own role and way of working within this new context. In the light of this development, AI should not be reduced to 'yet another new technology', but should be discussed on an organizational, legal, and ethical layer as well ¹⁾.

Many organizations are just beginning their AI journey. Pilots have shown that AI applications can help to save costs, for example by automating customer support processes where agents get supplemented or replaced by self services and highly functionable chatbots. The cost saving opportunities are equalized by huge investments in building up AI capabilities to extend the automatisations across further business processes, but it's not about costs only: AI is a main driver for innovation and growth. Game-changing applications can be found in all industries and span across all functions.

A PLETHORA OF USE CASES ACROSS FUNCTIONS & INDUSTRIES

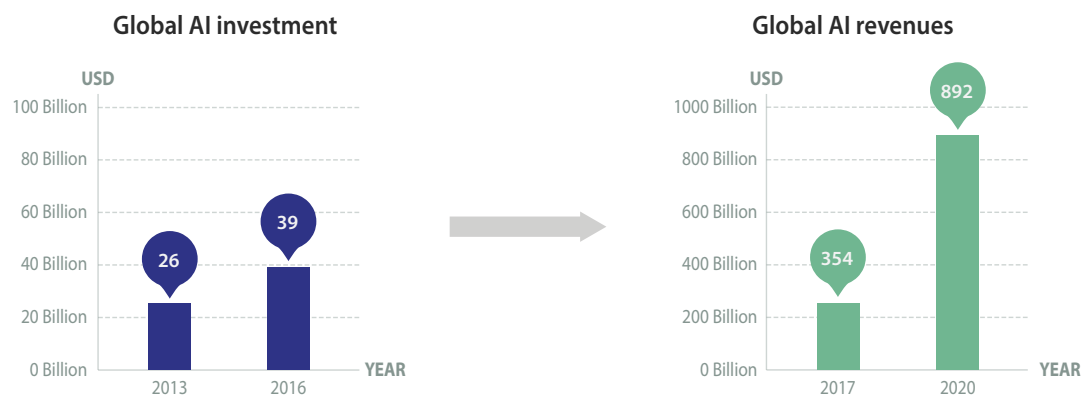
The potential of AI spans across industries and function. The digital workplace is becoming enhanced by smart applications. Business leaders have to decide where to start. They need to keep pace with the technological advancement and how their competitors embrace it.



Source:
<https://appliedai.com>

According to a survey from the consulting firm McKinsey ²⁾, global investment into AI technologies reached 39 billion USD in 2016, which is three times more than in 2013. This growth is expected to continue exponentially in the upcoming years, which also implies the creation of new jobs. Enterprises are willing to invest into AI, because significant revenue streams are anticipated from smart applications. In 2017, already \$354B of revenues were generated, for 2020 \$892B are expected ³⁾.

AI DELIVERING ON THE BUSINESS PROMISE



Sources:
McKinsey Global Institute (2017): Artificial Intelligence - The Next Digital Frontier
Statista (2018): Revenues from AI market worldwide from 2015 to 2024

This White Paper provides an overview of the current status of Artificial Intelligence adoption from a business and technology perspective. Decision makers will also learn about fundamental technology concepts, which will enable them to steer the strategic discussions about their business in a potentially less risky direction. In this context, it is essential to differentiate technology hype from actual business needs and to be aware of the maturity level of technologies, as well as the execution capabilities of one's own organization.

In this paper we will take a closer look at various types of AI approaches currently used. We will set a focus on a hybrid approach called 'Semantic AI', which makes use of machine learning like most contemporary AI efforts, but in combination with natural language processing and semantic technologies. We will also discuss why enterprises need AI Governance, and how Semantic AI differs from other frameworks in this respect.

Semantic AI as an advanced perspective on AI

Semantic AI is an approach that comes with technical and organizational advantages. It's more than 'yet another machine learning algorithm'. It's rather an AI strategy based on technical and organisational measures, which get implemented along the whole data lifecycle. Semantic AI provides a foundation for an enterprise-wide rollout of AI. In this chapter we will introduce 'Six core aspects of Semantic AI', and we will see, that as an integrated AI strategy, it unfolds its effects on various levels.

Most recently, many companies have started to learn how to enhance digital services, products and processes with AI by building prototypes. A project setup with a clearly defined outcome limits the risk of wasted resources and supports a gradual learning curve. Building up knowledge about these technologies and the usage of new solutions on an organisational level is the foundation of a solid business strategy. Developing an AI strategy must be more than ever an agile process, which is not governed by a top-down approach.

Introducing AI-based automation into workflows should not be the responsibility only of information architects, software developers or data scientists. Also subject matter experts need to have a stake in the development and maintenance of intelligent applications. Management needs to understand the workings of AI-enabled environments too. Organizations want to benefit from superior technology and at the same time don't want to get dependent on it. The perspective to build a "black box", which can be operated only by a few engineers is rightfully considered a major organizational risk.

Semantic AI combines thoroughly selected methods and tools that solve the most common use cases such as classification and recommendation in a highly precise manner. Current experience shows that AI initiatives often fail due to the lack of appropriate data or low data quality. A semantic knowledge graph is used at the heart of a semantic enhanced AI architecture, which provides means for a more automated data quality management. "Managing data in support of AI is not a one-off project, but an ongoing activity that should be formalized as part of your data management strategy"⁴⁾. Increase of data quality is also achieved due to the fact that subject matter experts without mathematical or software engineering skills can understand the logics behind data processing, and can contribute with their domain-specific knowledge. Combining conventional AI methodologies (e.g., Deep Learning) with semantic technologies also allows to establish a better division of work between differently specialized stakeholders.

Six core aspects of Semantic AI

Data Quality

Semantically enriched data serves as a basis for better data quality and provides more options for feature extraction. This results in higher precision of prediction & classification calculated by machine learning algorithms. With knowledge graphs in place, an advanced data model can be used that allows to make data interpretable and reusable in various contexts. (See also: [PoolParty Semantic Classifier](#))

Data as a Service

Linked data based on W3C Standards can serve as an enterprise-wide data platform and helps to provide training data for machine learning in a more cost-efficient way. Instead of generating data sets per application or use case, high-quality data can be extracted from a knowledge graph or a semantic data lake. Through this standards-based approach, also internal data and external data can be automatically linked and can be used as a rich data set for any machine learning task ¹³).

No black-box

In sharp contrast to AI technologies that ‘work like magic’ where only a few experts really understand the underlying techniques, Semantic AI seeks to provide an infrastructure to overcome information asymmetries between the developers of AI systems and other stakeholders, including consumers and policymakers. Semantic AI ultimately leads to AI governance that works on three layers: technically, ethically, and on the legal layer ¹).

Hybrid approach

Semantic AI is the combination of methods derived from symbolic AI and statistical AI. Virtuously playing the AI piano means that for a given use case various stakeholders, not only data scientists, but also process owners or subject matter experts, choose from available methods and tools, and collaboratively develop workflows that are most likely a good fit to tackle the underlying problem. For example, one can combine entity extraction based on machine learning with text mining methods based on semantic knowledge graphs and related reasoning capabilities to achieve the optimal results.

Structured data meets text

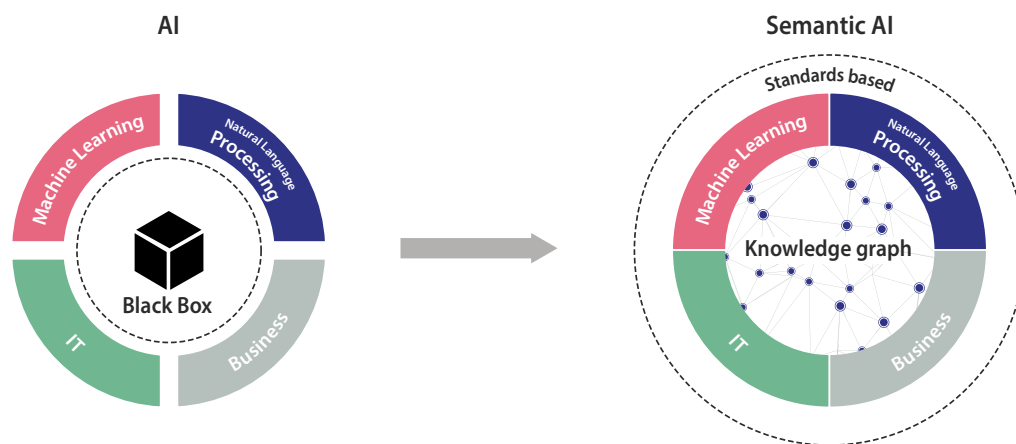
Most machine learning algorithms work well either with text or with structured data, but those two types of data are rarely combined to serve as a whole. Semantic data models can bridge this gap. Links and relations between business and data objects of all formats such as XML, relational data, CSV, and also unstructured text can be made available for further analysis. This allows us to link data even across heterogeneous data sources to provide data objects as training data sets which are composed of information from structured data and text at the same time.

Towards self optimizing machines

Semantic AI is the next-generation Artificial Intelligence. Machine learning can help to extend knowledge graphs (e.g., through ‘corpus-based ontology learning’ or through graph mapping based on ‘spreading activation’), and in return, knowledge graphs can help to improve ML algorithms (e.g., through ‘distant supervision’). This integrated approach ultimately leads to systems that work like self optimizing machines after an initial setup phase, while being transparent to the underlying knowledge models. Graph Convolutional Networks (in progress) promise new insights ¹⁴⁾.

SEMANTIC AI: CONNECTING STANDARDS-BASED TECHNOLOGY WITH PEOPLE

Successful Enterprise AI adoption depends on highly specialized and talented professionals, who can train machines and leverage data to generate value. Sustainable AI needs to be created with management and subject matter experts together. Semantic AI provides access and transparency to the underlying data models of AI-driven applications. A knowledge graph boosts data quality and provides a solid foundation for AI strategy and execution.



Benefiting from Semantic AI along the Data Lifecycle

Data is the fuel of the digital economy and the underlying asset of every AI application. Semantic AI addresses the need for interpretable and meaningful data, and it provides technologies to create this kind of data from the very beginning of a data lifecycle.

Companies possess and constantly generate data, which is distributed across various database systems. When it comes to the implementation of new use cases, usually very specific data is needed. Questions arise, if this data is available and if so, where. In many cases, valuable data could even be inferred automatically, if various data sources would get linked. Applications usually evolve and will require additional data from somewhere else. Generating data for a specific application doesn't mean that data workflows in the source system will be replaced. This can lead to data duplication and error-proneness in an organization.

Those few examples already spell out the complexity of agile data management. It is by no means a technical responsibility only, but illustrates the importance of a central data governance framework for digitizing an enterprise including its products and services.

The need for Semantic Data Lakes

Fishing data out of a data lake is a challenging task:

- Was really the right data chosen for the use case?
- Wouldn't there be other, more suitable data that might provide the desired outcome?
- Shouldn't we combine structured and unstructured data to fulfil the task?
- Isn't there data which actually belongs together, but is treated separately?
- Is the selected data complete or did we miss relevant data?
- Do we know the provenance of data? And if not, who will know about the data quality?
- Do we have the right to reuse the data in the given usage scenario?

Without a data catalog, available data can't reach up to the expected business potential. Semantic AI puts a strong focus on data linking and enrichment. It provides all necessary methods and tools to establish a professional information management and data governance. Thereby, all organizational stakeholders have an overview of available data and can define and steer data-driven initiatives in a much more precise way. AI is powerful when applied to high quality data. This is not only about maintaining records correctly, but enriching and linking it with additional information. Semantic AI pays tribute to the importance of generating, maintaining, and increasing data quality at any step of the data lifecycle.

Example: Which service requests can we automate?

Imagine a telecom provider, who wants to reduce costs by implementing self-service portals in its customer service department. In order to come up with a prioritized plan for strategy execution, the management wants to know:

- How many different types of service requests can be distinguished?
- For which products and services do we have the most incoming requests?
- Where do we lack or need to improve written information to automate the process?

According to a typical data lifecycle, the process would start with collecting the appropriate data for this use case. Data scientists would request to get access to available service tickets. However, only a subset will be relevant. How can be determined which subset this is? If the data would be enriched with meaningful metadata and would be linked to other resources, subject matter experts without any specific knowledge about the underlying datasets could provide guidance where to start.

Data Lifecycles and Semantics

There is no common understanding of what a data lifecycle is and which aspects should be highlighted or rather should be missed out ¹⁰⁾. A data lifecycle that includes the concept of linked and semantically enriched data can be developed in two ways.

Semantics and Linked Data along a traditional Data Life Cycle

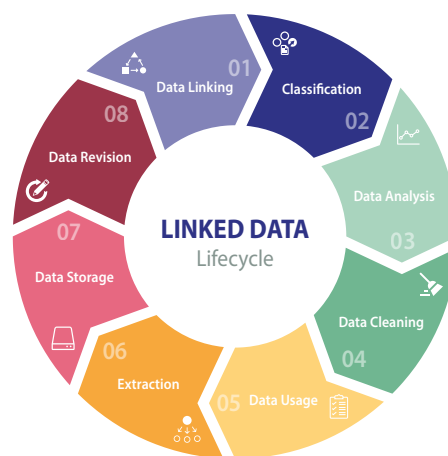
‘Data Capture’ or ‘Data Entry’ as typical first steps, followed by ‘Data Maintenance’ and ‘Data Synthesis’ are frequently the initial three steps used in life cycle models, before data gets used (‘Data Usage’). Any of these steps could potentially benefit from a semantic layer and the use of metadata based on standards. Examples:

- Capturing data from standards-based sources reduces data maintenance efforts. Such data could be extracted from internal or also external sources, e.g. from the linked open data cloud ¹¹⁾.
- When new data values are created by human operators or devices (‘Data Entry’), enterprises would benefit from using controlled vocabularies that are typically part of a semantic knowledge graph. Instead of ‘semantic re-engineering’ at the end of the life cycle, semantically rich data would be generated from the beginning.
- Extract-transform-load processes are typically executed during ‘Data Maintenance’. The less heterogeneous captured data is, the less costs are caused. Semantic AI strategies help to develop towards uniform data models in enterprises.
- Data Acquisition often involves contracts that govern how an organization is allowed to use the data. Semantic AI technologies can support the automated clearance of rights issues in the creation of derivative data works ¹²⁾.

- Data Synthesis as the creation of data values via inductive logic, using other data as input, is a natural fit with semantic graph technologies providing a rich set of ontology-based inferencing mechanisms.
- Data itself may be (part of) a product or service offered by the enterprise. ‘Data Usage’ and its potentially successive step ‘Data Publication’ may benefit from semantic metadata, eg. from context-sensitive automatisms to mesh information chunks with each other for further reuse.
- In most models two additional phases, ‘Data Archival’ and ‘Data Purging’, are included. Both can highly benefit from a semantic knowledge graph in place. Turning data into knowledge obviously deserves prerequisites that data objects and documents will receive some meaningful and documented labels before they get dumped in a data lake.

REACHING ENTERPRISE GOALS WITH THE LINKED DATA LIFECYCLE

Companies have to treat data as first class citizens, as all applications depend on high quality data. Linking data and enriching it with semantic metadata enables companies to build smart applications that perform along the whole data lifecycle.



The Linked Data Lifecycle

As an alternative to traditional data life cycle models, some more specific models have been developed, which set a clear focus on all activities towards the creation of linked data and a semantic knowledge graph. These models should be considered as a complementary overlay of the traditional data life cycle models. Thus, they provide a clear picture of which activities, tools, skills and methodologies should be in place before enterprises are able to develop their data governance towards Semantic AI. As illustrated, most steps of the linked data lifecycle, for example interlinking or classification, are not highlighted in conventional life cycle models ¹⁵⁾.

Semantic Knowledge Graphs

Semantic knowledge graphs help to make data interpretable for man and machine due to explicit semantics based on standards. It's all about things, not strings: Any business object represented in a knowledge graph receives a unique address and can then easily be integrated in a bot, service, or application. Terms and strings are no longer used to build ambiguous metadata, instead, a semantic layer on top of all content and data assets works like a multi-dimensional index.

THE KNOWLEDGE GRAPH AS TURBOCHARGER FOR SEMANTIC AI

A knowledge graph represents a domain on a metalevel and defines which metadata should be used to enrich available data repositories. It significantly increases data quality and allows to discover new insights due to extensively linked data. A knowledge graph also helps interdisciplinary stakeholders to work together on AI use cases as data organization becomes visible.



Semantic graphs can build the backbone of any information architecture, not only on the web. They can enable entity-centric views also on enterprise information and data. Such graphs of things contain information about business objects (such as products, suppliers, employees, locations, research topics, ...), their different names, and relations to each other. Information about entities can be found in structured (relational databases), semi-structured (XML), and unstructured (text) data objects. Nevertheless, people are not interested in containers but in entities themselves, so they need to be extracted and organized in a reasonable way.

Machines and algorithms make use of semantic graphs to retrieve not only simply the objects themselves but also the relations that can be found between the business objects, even if they are not explicitly stated. As a result, 'knowledge lenses' are delivered that help users to better understand the underlying meaning of business objects when put into a specific context.

The ability to take a view on entities or business objects in different ways when put into various contexts is key for many knowledge workers. For example, drugs have regulatory aspects, a therapeutical character, and some other meaning to product managers or sales people. One can benefit quickly when only confronted with those aspects of an entity that are really relevant in a given situation. This rather personalized information processing has heavy demand for a semantic layer on top of the data layer, especially when information is stored in various forms and when scattered around different repositories.

Recommender engines based on semantic graphs can link similar contents or documents that are related to each other in a highly precise manner. The same algorithms help to link users to content assets or products. This approach is the basis for ‘push-services’ that try to ‘understand’ users’ needs in a highly precise way.

Technical Deep Dive: How to build Knowledge Graphs

These days, many organisations have begun to develop their own knowledge graphs. One reason is to build a solid basis for various machine learning and cognitive computing efforts. For many, it remains still unclear where to start. The Simple Knowledge Organization System (SKOS) offers a simple way to start and opens many doors to extend a knowledge graph over time.

- *Standardised:* SKOS is a standards-based ontology, which was published by the World Wide Web Consortium (W3C) in 2009 ⁵⁾.
- *Future-proof:* SKOS is part of a larger set of open standards, which is also known as the Semantic Web ⁶⁾. The usage of open standards for data and knowledge models eliminates proprietary vendor lock-in.
- *Wide range of applications:* SKOS is primarily used to build and to make controlled vocabularies like taxonomies, thesauri or business vocabularies available as a service ⁷⁾. This builds the basis for a wide range of applications, starting from semantic search and text mining, ranging to data integration, data analytics and machine learning.
- *Graph-based:* SKOS concepts can be related and linked to other concepts and instances of ontologies. By these means, SKOS constitutes the nucleus of a decentralised enterprise-wide knowledge graph.
- *Cost-efficient and incremental approach:* Any SKOS-based taxonomy, thesaurus, or controlled vocabulary can be extended and enriched by additional ontologies step-by-step, thus context-sensitive views on the same node can be created when needed. SKOS-based vocabularies can be used as a starting point for a cost-efficient development of more extensive semantic knowledge graphs.
- *Actionable content:* SKOS models, no matter if linked to more expressive ontologies or not, can be queried with SPARQL, or validated by the use of SHACL, a recently issued standard for describing and validating RDF graphs. By that means, knowledge models become actionable

and can help to find answers in unstructured content, trigger alerts or to make better decisions.

- *Widely adopted:* Hundreds of SKOS vocabularies are available on the web ⁸⁾. Large international bodies like the EU, UN, or The World Bank make use of SKOS to make their knowledge available to external and internal stakeholders as well. Many Fortune 500 companies have already adopted SKOS for internal use.
- *Mandatory:* SKOS, RDF and other standards can, for instance, be required in EU public procurement. ⁹⁾

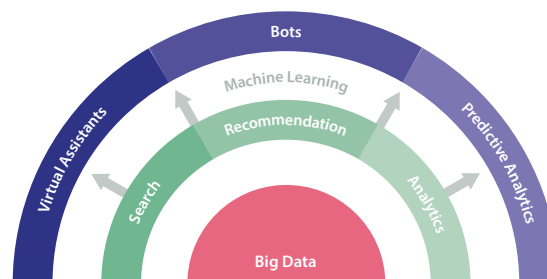
Towards Semantic AI

AI has not arrived unexpected. Companies with experience in implementing search, recommendation and analytics technologies have now started to look into Machine Learning to build AI-based information systems. An increasing level of automation level is made possible by more mature algorithms like Deep Learning, the ever increasing computing power and the option to use machine learning as a service. Advances in natural language processing or in image processing change how digital communication channels will work in the near future. Virtual assistants and chatbots are on the rise, have an impact on the workplace and can be definitely categorized as a “game changer”. The new generation of conversational user interfaces is able to execute working steps for employees or customers.

However, enterprises that have already built search engines over various data repositories or product recommendation engines know that data quality is key for personalized digital experiences. New technology methods and tools definitely bear potential for disruptive innovation, but they depend not only on good data quality, but also on advanced data models that allow to make data interpretable and reusable in specific domains and various contexts. This task cannot be outsourced, but remains to be a core competency of any organization. Organizations will only reach the vision of a broadly established and well functioning Enterprise AI, when they invest in professional information management, which is in practice often heavily neglected. With the explosion of data and the expectations to make use of it, data governance becomes mandatory. Otherwise, organizations start automating fragmented data, which will end up in automated chaos. Semantic AI seems to be a promising option to tackle this challenge.

ENTERPRISE AI: IT'S AN EVOLUTION, NOT A DISRUPTION

Companies are still learning to utilize data for smart applications.
With Machine Learning, we are entering a new generation of smart applications.



Summary / Key statements of this paper

Developing an AI strategy must be more than ever an agile process, which is not governed by a top-down approach.

Semantic AI addresses the need for interpretable and meaningful data, and it provides technologies to create this kind of data from the very beginning of a data lifecycle.

Semantic AI is more than ‘yet another machine learning algorithm’. It’s rather an AI strategy based on technical and organisational measures, which get implemented along the whole data lifecycle.

Semantic AI makes use of machine learning like most contemporary AI efforts, but in combination with natural language processing and semantic technologies.

The perspective to build a “black box”, which can be operated only by a few engineers is rightfully considered a major organizational risk.

Current experience shows that AI initiatives often fail due to the lack of appropriate data or low data quality. A semantic knowledge graph is used at the heart of a semantic enhanced AI architecture, which provides means for a more automated data quality management.

Semantically enriched data serves as a basis for better data quality and provides more options for feature extraction.

Through a standards-based approach, also internal data and external data can be automatically linked and can be used as a rich data set for any machine learning task.

Semantic AI seeks to provide an infrastructure to overcome information asymmetries between the developers of AI systems and other stakeholders, including consumers and policymakers.

Semantic AI is the combination of methods derived from symbolic AI and statistical AI.

Most machine learning algorithms work well either with text or with structured data, but those two types of data are rarely combined to serve as a whole. Semantic data models can bridge this gap.

Semantic AI ultimately leads to systems that work like self optimizing machines after an initial setup phase, while being transparent to the underlying knowledge models.

Machines and algorithms make use of semantic graphs to retrieve not only simply the objects themselves but also the relations that can be found between the business objects, even if they are not explicitly stated.

With the explosion of data and the expectations to make use of it, data governance becomes mandatory. Otherwise, organizations start automating fragmented data, which will end up in automated chaos. Semantic AI seems to be a promising option to tackle this challenge.

References

- 01) Gasser, U., and Virgilio A.F. Almeida (2017): “A Layered Model for AI Governance.” IEEE Internet Computing 21 (6) (November): 58–62. doi:10.1109/mic.2017.4180835.
- 02) McKinsey Global Institute (2017): “Artificial Intelligence - The Next Digital Frontier”
- 03) Statista (2018): Revenues from AI market worldwide from 2015 to 2024
- 04) Gartner (2017): “Four Data Management Best Practices for AI” - Roxane Edjlali, Svetlana Sicular
- 05) W3C: SKOS Simple Knowledge Organization System
<https://www.w3.org/2004/02/skos/>
- 06) W3C: Semantic Web Standards
<https://www.w3.org/standards/semanticweb/>
- 07) Priscilla Jane Frazier (2015): “SKOS - A Guide for Information Professionals”
<http://www.ala.org/alcts/resources/z687/skos>
- 08) The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)
<http://bartoc.org/>
- 09) Commission Implementing Decision (EU) 2017/1358
- 10) Malcolm Chisholm (2015): “7 Phases of A Data Life Cycle”
<https://www.information-management.com/news/7-phases-of-a-data-life-cycle>
- 11) The Linked Open Data Cloud (LOD Cloud)
<http://lod-cloud.net/>
- 12) Pellegrini, T., Mireles, V., Steyskal, S., Panasiuk, O., Fensel, A., & Kirrane, S. (2018). Automated Rights Clearance Using Semantic Web Technologies: The DALICC Framework. In Semantic Applications (pp. 203-218). Springer Vieweg, Berlin, Heidelberg.
- 13) Dumontier, M., Wilcke, X., Bloem, P., & de Boer, V. (2017): “The knowledge graph as the default data model for learning on heterogeneous knowledge.” Data Science, 1(2), 39-57.
- 14) Matthew Das Sarma (2018): “How do we capture structure in relational data?” The Gradient (by SAIL - Stanford Artificial Intelligence Laboratory)
<https://thegradient.pub/structure-learning/>
- 15) Auer, S., Bühmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., & Stadler, C. (2012). Managing the life-cycle of linked data with the LOD2 stack. In International semantic Web conference (pp. 1-16). Springer, Berlin, Heidelberg.
https://link.springer.com/content/pdf/10.1007/978-3-642-35173-0_1.pdf