

# Large Language Models: Expectations for Semantics-driven Systems Engineering

Robert Buchmann<sup>1</sup>, Johann Eder<sup>2</sup>, Hans-Georg Fill<sup>3</sup>, Ulrich Frank<sup>4</sup>,  
Dimitris Karagiannis<sup>5</sup>, Emanuele Laurenzi<sup>6</sup>, John Mylopoulos<sup>7</sup>,  
Dimitris Plexousakis<sup>8</sup>, Maribel Yasmina Santos<sup>9</sup>

<sup>1</sup> Babeş-Bolyai University of Cluj-Napoca, Romania

<sup>2</sup> University Alpen-Adria Klagenfurt, Austria

<sup>3</sup> University of Fribourg, Switzerland

<sup>4</sup> University of Duisburg-Essen, Germany

<sup>5</sup> University of Vienna, Austria

<sup>6</sup> FHNW - University of Applied Sciences and Arts Northwestern Switzerland

<sup>7</sup> University of Toronto, Canada

<sup>8</sup> Institute of Computer Science, FORTH and University of Crete, Greece

<sup>9</sup> University of Minho, Portugal

robert.buchmann@econ.ubbcluj.ro, johann.eder@aau.at  
hans-georg.fill@unifr.ch, ulrich.frank@uni-due.de  
dk@dke.univie.ac.at, emanuele.laurenzi@fhnw.ch  
jm@cs.toronto.edu, dp@ics.forth.gr  
maribel@dsi.uminho.pt

**Abstract.** The hype of Large Language Models manifests in disruptions, expectations or concerns in scientific communities that have focused for a long time on design-oriented research. The current experiences with Large Language Models and associated products (e.g. ChatGPT) lead to diverse positions regarding the foreseeable evolution of such products from the point of view of scholars who have been working with designed abstractions for most of their careers - typically relying on deterministic design decisions to ensure systems and automation reliability. Such expectations are collected in this paper in relation to a flavor of systems engineering that relies on explicit knowledge structures, introduced here as “semantics-driven systems engineering”.

The paper was motivated by the panel discussion that took place at CAiSE 2023 in Zaragoza, Spain, during the workshop on Knowledge Graphs for Semantics-driven Systems Engineering (KG4SDSE). The workshop brought together Conceptual Modeling researchers with an interest in specific applications of Knowledge Graphs and the semantic enrichment benefits they can bring to systems engineering. The panel context and consensus are summarized at the end of the paper, preceded by a proposed research agenda considering the expressed positions.

**Keywords:** Large language models, systems engineering, conceptual modeling, knowledge engineering

## 1 Introduction

Large Language Models (LLMs) and the hype surrounding associated products like ChatGPT promised and brought disruption in content creation, workflows, quality assurance and engineering processes therefore debates over their benefits, pitfalls and potential evolution have been dominant scrutiny issues both in technology-focused social media and in scientific communities for the last couple of years [1-4].

Reports on LLM adoption and use often focus on concerns related to ethics, safety, transparency and responsibility - which are of course major issues relevant across application domains. Although we are following the debate on such foundational issues [5-7] and the regulatory initiatives associated with them<sup>1</sup>, this paper's perspective focuses on systems engineering as an application domain for LLMs, and the value that LLMs can bring to this.

We believe that ethics, safety and accountability should be core to LLM systems themselves rather than delegated as "separated concerns" to external guardrails of information systems that rely on them. The terms involved in the discussions on responsibility should be more clearly defined, since their meaning is often assumed, or repurposed. Is "explainability" about obtaining some ontological insight or about identifying the network layer that produced a certain result? Is the responsibility of data to be fair, or of the system ingesting the data, or of the system that performs tasks based on these data? Exact scopes and purposes are still ill-defined as the discourse tends to blur boundaries. Are LLMs expected to simultaneously report facts (for which hallucination should be avoided), generate creative content (for which hallucination may be desirable), make decisions, automate actions and compensate for already existing societal biases?

Considering that all authors of the present paper possess a background in conceptual modeling and related forms of engineered knowledge, we focus on the interplay between LLMs and conceptual structures for the benefit of semantics-driven systems engineering and associated scholarly activities. Semantics-driven systems engineering relies on declarative knowledge expressed in structures that are founded on well-defined theories. New knowledge is produced by applying automated reasoning methods, following mainly the symbolic AI paradigm.

The LLM-Symbolic AI interplay has been a recurring topic in both tech and scientific events - sometimes in a vendor-specific setting or as part of larger events such as the Knowledge Graph Conference<sup>2</sup>, where it was suggested that Knowledge Graphs adoption can be accelerated by LLMs [8-9]. Also related to our perspective is the research stream towards the combination of machine learning and knowledge engineering approaches [10-12] - see the several editions of the AAAI-MAKE workshop<sup>3</sup>, part of the AAAI Spring Symposium series; the Neuro-Symbolic AI initiative [13-14] aiming to hybridize engineered knowledge with knowledge learned from neural networks; the

---

<sup>1</sup> see the EU AI Act, <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

<sup>2</sup> <https://www.knowledgegraph.tech/>

<sup>3</sup> <https://www.aaai-make.info/>

recently launched Neuro-Symbolic Artificial Intelligence journal<sup>4</sup>; domain-specific manifestations of the symbolic-subsymbolic convergence [15]; or, the recent proposal of G. Marcus and D. Lenat on a potential LLM-Cyc synergy [16].

This paper will introduce the KG4SDSE workshop's interpretation on *Semantics-driven Systems Engineering* (SDSE) - in contrast to the traditional Model-driven Engineering (MDE) - and will relate it to concerns and expectations classified in two categories - firstly, regarding LLMs as technology and products; secondly, regarding users and how they see and adopt LLMs. On this basis, a research agenda will be formulated in Section 4. The paper ends with concluding quotes from contributors and stresses the importance of design-oriented approaches when researching LLM-based information systems.

## 2 Semantics-driven Systems Engineering

### 2.1 From Model-driven Engineering to Semantics-driven Systems Engineering

We start by characterizing the notion of SDSE in the way it was introduced by the KG4SDSE workshop where the hereby synthesized ideas were initiated. The notion of SDSE will be further instantiated by a toy example manifesting the key characteristics.

Traditionally, model-driven software engineering [17] and the more recent, industrial notion of low-code platforms [18-20], relied on fixed/stable metamodels that allow us to treat diagrammatic models as a form of code or configuration - e.g. generating Java classes from UML class diagrams, or running workflows in engines based on process diagrams. Such use cases are enabled by stable or standardized conceptual schemas to facilitate transformation or interoperability of models. If needed, roundtrip engineering for maintaining code and the synchronization of diagrammatic models may be added.

The notion of Semantics-driven Systems Engineering - initially discussed as a flavor of model-driven engineering labelled "model-aware engineering" in [21] - goes back to a looser interpretation on the bindings between model contents and running code/software. In that loose interpretation a developer simply reads and understands models as a "knowledge base" (rather than as visual symbols of code patterns) and tries to implement what they understand. This "model-aware" flavor implies a knowledge management approach and is not automated as long as human cognition must act as the mediator between the model-based documentation and the engineered system.

However, in the age of Knowledge Graphs and Large Language Models, this knowledge-centric interpretation, which has been somewhat left behind by the "code generation" desideratum of traditional model-driven engineering, can be revisited. It can take novel forms, by capturing the human interpretability/mediation and turning it into machine interpretability relying on semantic information retrieval and machine reasoning. To support this, models must be able to capture and operationalize arbitrary

---

<sup>4</sup> <https://www.iospress.com/catalog/journals/neurosymbolic-artificial-intelligence>

domain-specific semantics that can deviate unpredictably from standard metamodels. Therefore, they cannot be simply compiled into code because of the evolving nature of the method for capturing such semantics - which typically employs a domain-specific modeling language (DSML). Instantiations of this idea from our previous work are “Semantic Meta-Model” [22], “Linked Open Models” [23], “Ontology-based Meta-Meta-Model” [24], where the meta-model contributes to a knowledge graph schema and visual model elements become instances of the schema. In [25] and [26] agile aspects have been incorporated to enable the creation of knowledge graph-based DSMLs or model-driven knowledge graphs.

In this sense, the core aspect of SDSE is that what the human developers used to “interpret” by looking at visual models can now be transferred (to some extent) to various forms of knowledge retrieval and machine interpretability. We envision that the coupling between agile DSMLs, knowledge graphs and LLMs can be a key enabler to achieve this, keeping conceptual modeling as a relevant systems engineering practice but shifting its role and the way models are involved in engineering processes.

## 2.2 Semantics-driven Systems Engineering: an Exemplar

In order to distill and illustrate the key characteristics of SDSE, we use a toy exemplar based on the OMiLAB Digital Innovation environment [27] – a teaching and experimentation installation having conceptual modeling methods and digital engineering as core preoccupations. One component of the installation is a robotic arm that can be involved in a scenario mimicking a cooking process. The cooking is performed according to recipes described by conceptual modeling means - flowcharts of actions mapped on ingredient positions - for which the BEE-UP modeling tool [28] is used. Different configurations underlying different approaches for enacting model-driven behavior are summarized by the three setups in Fig.1 suggesting the progression from model-driven engineering to SDSE:

**Setup A. Model-driven engineering (in the general sense)** represents the configuration where each element of the diagrammatic cooking recipe generates a piece of code that triggers execution of the corresponding cooking step - technically, an HTTP call to the API that manages the robotic movements. At every moment, the robotic arm is only “aware” of the current execution step and its parameters, following the flowchart prescription; at best, it can collect a log of those steps.

**Setup B. Semantics-driven engineering** represents the configuration where the cooking recipes are treated as a semantic repository, essentially a Knowledge Graph conforming the Linked Open Models patterns discussed in [23] and facilitated by a diagram-to-RDF export feature available in BEE-UP. The robotic arm now takes a more “knowledgeable” role as it runs semantic queries (SPARQL) on the Knowledge Graph comprising the diagrammatic recipes and relevant data properties (e.g. ingredient positions, richer ingredient descriptions). It can still achieve the stepwise execution by querying the sequence of steps, but it can also retrieve semantically richer information – depending also on the richness of descriptions enabled in the modeling tool (BEE-UP being developed through a metamodeling approach). Rules can also be added - to enforce semantic validation (e.g. are the vegan recipes really avoiding animal by-

products?), to reason on alternatives (if ingredient X is not available, can it be replaced with a ingredient Y?), to highlight semantic patterns (e.g. which are the recipes using ingredient X after ingredient Y?), to generate reports associated with the enacted behavior (e.g. report the ingredient consumption for a selected meal). The robotic behavior is not only endowed with flexibility and decision-making, but it can also be augmented with a suite of reporting or controlling apps that share the same knowledge base, possibly updated with execution traces for subsequent analysis. In response to potentially evolving requirements, the Knowledge Graph can be enriched on diagrammatic level by metamodeling means or directly on repository level (linking to ingredient taxonomies, OWL axioms etc.). In short, modeling contributes to a knowledge base to be consulted by a cyber-physical artifact, instead of abstracting some executable code patterns.

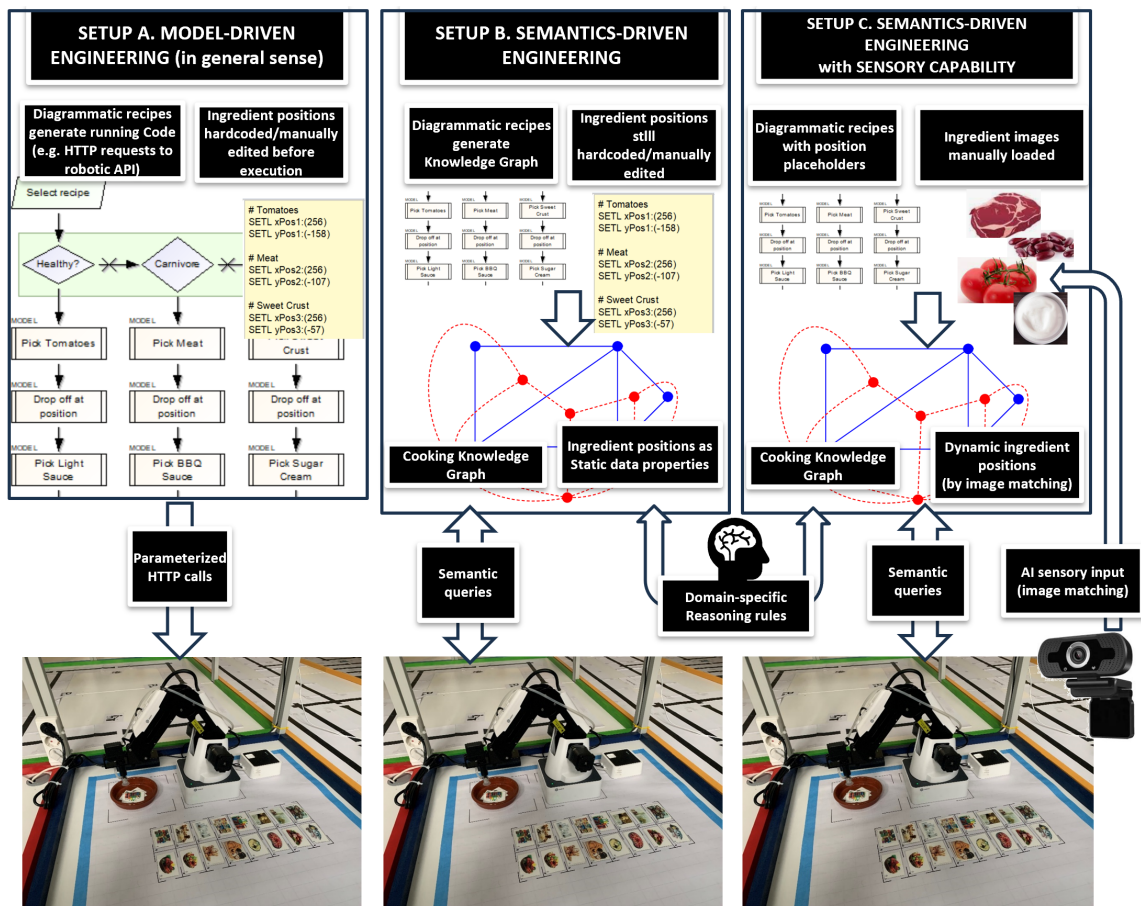


Figure 1. From Model-driven engineering to Semantics-driven engineering

**Setup C. Semantics-driven engineering with sensory capability** further extends SDSE with the ability to recognize the ingredients (laminated image cards) by means of computer vision – by matching them against user-provided images in the modeling environment to dynamically detect their positions, or missing ingredients. This complements the Knowledge Graph of Setup B with sensory data capture from the environment, a minimalist setup to illustrate the benefits of combining prescriptive knowledge with data-driven machine intelligence.

### 2.3 Potential LLM Roles in Semantics-driven Systems Engineering

Relative to the SDSE toy example previously introduced, LLMs can be envisioned as potentially contributing to several of the involved components, highlighted in Fig. 2 – from the end user-facing role to engineering roles:

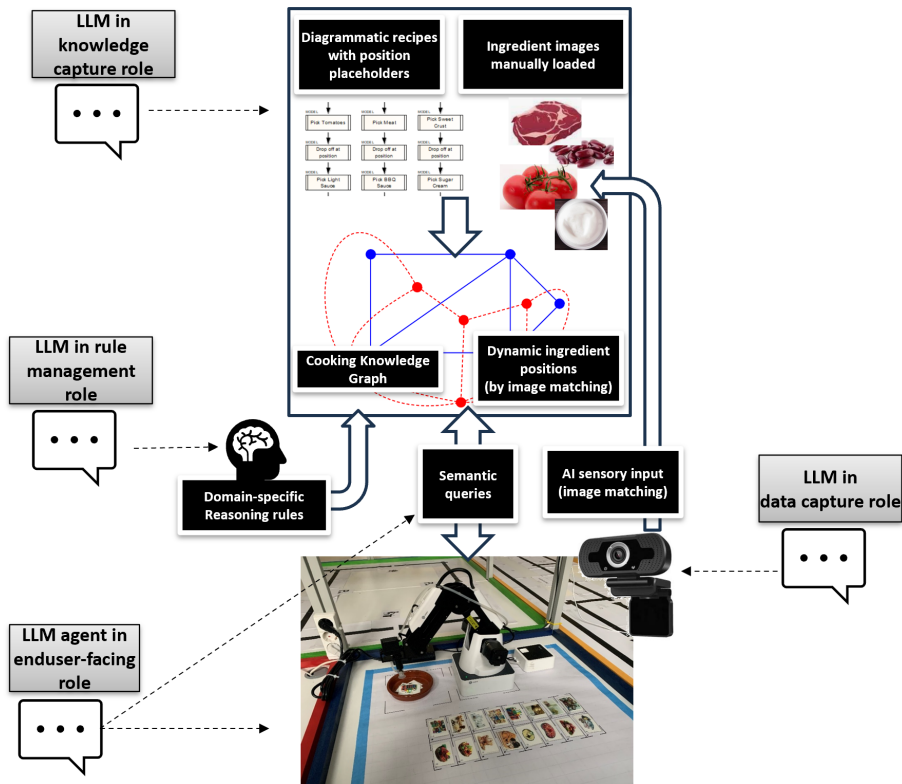


Figure 2. Potential LLM roles in semantics-driven engineering

- The end user-facing role can manifest as a natural language interface provided to the customer of the cooking service for specifying the desired recipe, or picking one from a recommender that converts the recipe graphs to natural language summaries. This can also propagate to the semantic queries that the robotic device can execute -

already some early results have been reported with SPARQL generation out of natural language requests [29]; alternatively, retrieval-augmented generation (RAG) [30] can be employed to contextualize the conversation with content from the structured knowledge repository and any related external sources that can be prepared and indexed in support of the augmentation (in BEE-UP, diagrammatic content can be annotated with RDF triples and links to external URIs);

- The domain-specific rules, regardless of having a restrictive or a productive purpose are, in current experimentation, either added manually to a semantic graph repository or, more laymen oriented, visually described in an improvised DSML for expressing ingredient alternatives or incompatibilities. Rule learning [31] can contribute here, although a powerful SPARQL generation solution may also tackle those semantic queries that can effectively act as rules;
- Similarly, LLMs can contribute to the data capture role, complementing sensory capabilities with data collected from the conversational interface;
- Finally, the knowledge acquisition and structuring role implies the use of LLMs for knowledge engineering [32], with prominent instances of this being knowledge graph population [33][34] or discovering implicit ontologies that underly conversational content [35]. Although it can be argued that diagrammatic representations may be circumvented by direct interaction between LLMs and knowledge graphs, we believe that knowledge visualization has its reserved role in achieving the sense-making and communication goals that are at the core of the conceptual modeling discipline [36]. Visual models critically contribute to cognitive effectiveness of certain types of content, e.g. procedural knowledge in the form of process diagrams [37]. This raises requirements for LLMs to contribute not only through their interplay with knowledge graphs in the various combinations roadmapped by the literature [33], but as part of a multi-modal “triangle” that also involves diagrammatic knowledge representation (see early reports available in [1] or [38]).

The distinction enforced here between end user-facing and engineering-facing roles may be rendered irrelevant, depending on the prompt engineering capabilities that are involved. Such skills become more and more common as users/customers of LLM products engage in prompting strategies to gradually refine specification of their problem or solution requirements, ultimately contributing on an engineering level to get the most out of an LLM-based service. On the other hand, certain fine-tuning approaches such as LoRA (low-rank adaptation of LLM) [39] are expected to require engineering expertise in the long term.

### 3 Expectations from Large Language Models

LLMs, as they have become known to the public through the advent of ChatGPT in November 2022, provide fascinating properties that even experienced computer scientists had not expected. LLMs offer great capabilities for tasks such as the generation of text in various languages on any topic, the redaction of text to mimic specific styles or tones, the summarization of text or the generation of code in multiple languages based

on requirements that can then be refined. Multi-modal versions of LLMs are further capable of processing visual input, e.g. to identify objects in images [40].

The quality and fluency of the responses generated by ChatGPT led to the expectation that an artificial agent would now be available to solve problems of any kind in the sense of General Artificial Intelligence and may even reach human-like capabilities. While this expectation and the fears resulting from it in the public perception have clearly not been met so far, the actual benefits and properties of LLMs still need to be explored by the scientific community and industry alike.

The contributors of this paper explored the tools and the topic from their individual perspectives as academics and researchers, formulating both current observations and expectations from the rapid evolution of the technology. The main ideas are synthesized in the next sections - first for the LLM-based systems themselves, and then for the users or engineers adopting them.

The criteria for distinguishing the two categories of expectations is rooted in the socio-technical nature of information systems, i.e. we need to look both at the technical artifacts and how they are perceived, adopted or engineered. Information Systems scholars are commonly approaching systems from one of these perspectives, while also considering constraints or requirements from the other perspective. Following this, a research agenda will be structured in Section 4 by the same criteria.

### 3.1 Expectations from LLM Technology/Products

**Supporting conceptual modeling.** Experiments have been recently reported in journals that publish conceptual modeling research, such as EMISAJ [1] and Software and Systems Modeling [38], on how LLMs can generate or process diagrammatic content of various model types - primarily UML class diagrams, BPMN diagrams or ER models, with some impressive results suggesting that ChatGPT is aware of certain abstraction patterns present in such structured visualizations. Does this mean that the systems engineering community can already delegate abstraction and design tasks to LLM-based products?

Diagrammatic representations are typically better structured than natural language and some forms of model-driven engineering (in the sense of code generation or configuration generation) have been reliable for a long time. It is expected for LLMs to facilitate an extended roundtrip cycle involving natural language descriptions, diagrammatic blueprints and code patterns. Future model editors may offer natural language interactions to create and adapt model contents, thus having direct impact on the barriers of modeling that are a concern for the enterprise modeling community [41]. Conversely, model contents can be communicated by LLM tools to laypersons unfamiliar with modeling methods or their notations.

On the other hand, SDSE as defined in Section 2 needs agilely evolving languages and sometimes narrow-scoped DSMLs with only limited example corpora available for training. This suggests a need for LLMs to be prepared to handle the abstraction mechanisms used in metamodeling, multi-level modeling and similar knowledge engineering paradigms. Even if this may remain a barrier, immediate benefits are foreseeable in tasks such as knowledge acquisition, conceptualization or requirements elicitation - i.e.

providing natural language means to laypersons who get involved in such tasks. Current LLM systems are perceived as means of knowledge distribution from the system to human users. By expanding this use case towards domain conceptualization, it would be important for LLMs to be able to ask the right questions (and to ask surprising questions) in order to support its own knowledge ingestion and ultimately to make possible the bidirectional conversational refinement that typically manifests in business analysis or requirements elicitation.

Due to the novelty of the technology, the scientific community needs to conduct experiments and explore which LLM properties are most suitable to scenarios pertaining to the automation of modeling, model use or even DSML engineering.

**Making semantics explicit.** Truthfulness is not a strong point of LLMs for a variety of reasons (including an inherent mechanism of mimicking human falsehoods), and some benchmarks suggest that this will not be overcome simply by scaling [42]. Hallucination can be limited by explicit truth-making abstractions embedded as inference support or guardrails [16], and the recent contextualization approach of retrieval-augmented generation (RAG) [30].

The dichotomy (and complementarity) between "slow" and "fast" thinking, initially introduced in a psychology context [43] has been lately adopted as a metaphor for the hybridization of engineered knowledge and learned knowledge towards more efficient enterprise knowledge management - e.g. see a Deloitte white paper showing an industry perspective on this in [44]. This translates to a need to hybridize and agilely shift between fast data-driven pattern recognition AI and a priori logic-based, knowledge-intensive AI with explicitly represented semantics - at least for highly reusable conceptual structures e.g. causality, containment, temporal patterns. Certain fundamental concepts, already found in upper ontologies, can be expected to be invariants for a discipline such as systems engineering and there's no reason for them to be re-learned from training corpora or fine-tuning efforts. They can be embedded as deterministic invariants and natural priors; their "invariant" quality can be subjected to multiple levels of flexibility depending on how reusable they are, possibly involving already available knowledge repositories such as the Cyc project<sup>5</sup>.

**Recognizing a refined notion of "context".** Semantics also requires a reconsideration of what "context window" means for AI systems. For an LLM it is typically about the number of tokens, but for knowledge and data management it is about capturing explicit contextual relationships – e.g. the 6W interrogations (what, when, where, who, how, why) or other radial chains of relationships, not necessarily expressed in the neighboring text tokens. Context-awareness must reconcile the token-oriented and the graph-oriented understandings of "context" to allow for more agile LLMs that can formulate content not only in terms of narrative styles but also in terms of qualifying contexts.

It must be noted that symbolic AI models are logic-based and not all of them are conceptual models. Description Logics are definitely conceptual in the sense that they are founded on the notions of concepts and roles. But other models, such as ones based

---

<sup>5</sup> <https://cyc.com/>

on First-Order Logic are not. This is important because the great modeling deficit of LLMs is that they are statistical, rather than conceptual and rely on massive training sets. This results in using the Web for training data, and it translates to LLMs not knowing what is true vs false, or right vs wrong. This can be contrasted with the learning we went through as students which is based on learning concepts starting from simpler ones, e.g., turning, parking and backing up, before moving to complex ones, e.g., driving. This recommends a learning method that is more in line with educational theories of learning and cognitive development, such as those of Jean Piaget, as opposed to the amorphous training process of LLMs where an autonomous vehicle is supposed to learn driving without having any knowledge of turning or backing up. Such a shift would mean that LLMs will require less training data, also that LLMs have a lot to benefit from conceptual modeling enabling some form of “symbolic thought” [45].

**Considering resource efficiency.** The statistical approach of LLM aims to minimize the need for the mentioned prior “hardcoded” knowledge, while consuming extensive resources to derive patterns by statistical learning, even if some of that knowledge is reasonably invariant. On the other hand, natural intelligent systems seem to strive for a balance between the energy consumption involved in learning, acquisition and memorization of invariant shared knowledge - this reflects the traditional dichotomy between symbolic and sub-symbolic AI, and the above-mentioned dichotomy between “slow thinking” and “fast thinking” discussed in psychology and knowledge management. Once resource efficiency comes into play, research is called to investigate the trade-offs between what is worth learning and what is worth embedding as priors in the form of engineered knowledge. The resource-agnostic ambition of learning everything may be energy-detrimental, even if having a 24/7 system capable of cheaply distributing knowledge can be seen as a form of energy saving on the consumer side – thus a trade-off between personal effort and environmental cost.

It has been noted that LLMs such as ChatGPT seem to show some ability pertaining to logic and abstraction rather than being limited to statistical predictions of text sequences - it remains to be discussed to what extent this ability is learned, at what cost and what part of that cost can be removed by explicit knowledge engineering. Further details are needed about the internal design decisions of LLM-driven system, in order to determine the extent to which such systems are really limited to “learned knowledge” or whether that is just a dominant idea obscuring the fact that some engineered knowledge base already plays a part, such as in graph-based RAG configurations. Technical reports on LLMs typically focus on the quantitative scaling (e.g. how much the parameters or the token context has expanded) often leaving out such design decisions.

The ambition of “Green AI” as recently surveyed [46] seems to entirely neglect symbolic AI and knowledge engineering - mitigations are discussed there strictly in terms of machine learning. Experiments on energy saving typically discuss trade-offs about the neural network structure (e.g. removal of neurons with some drop in accuracy) [47], also implying that AI and deep learning are seen by some as synonyms. The notion of Green AI should expand with investigations tackling energy trade-offs in neuro-symbolic or RAG architectures. Machine learning energy consumption has been granularly decomposed and analyzed [48], making it possible to trace where and to what extent

symbolic structures can be a facilitator for resource efficiency, besides acting as a reasoning aid.

**Managing confidence and trust levels.** While it is expected that explainability, quality checks, truthfulness and sound referencing of the baseline information underlying generated content will remain a challenge for the near future [42], perhaps some immediate measures can be taken in the sense of avoiding high confidence in the delivery of counterfactual information, or suggesting levels of confidence for different parts of the generated content.

Providers of LLM-based services must consider means of communicating to end-users the distinction between "guessed" information, reproduced information and other means by which the content has been distilled. Only the occasional generic disclaimer (often ignored by users misunderstanding the technology) is not sufficient. References to sources and explanations for the generated content remain a strong desiderate for the further development of LLMs. Moreover, LLMs are expected to recognize LLM-generated text to avoid automated re-ingestion that can lead to an inflation of unverified content that eliminates humans from the knowledge co-creation loop and runs into the dangers of "model collapse" [49].

### 3.2 Expectations from Users and Systems Engineers

**Recognizing LLM limitations with respect to systems engineering tasks.** The engineering of information systems leveraging on the application of LLMs need to consider the (at least currently) inherent shortcomings of LLMs and the risk of hallucinated software packages [50]. For solutions generated by LLMs adequate test capabilities must be developed. In some application areas tests can easily check whether a solution is valid – e.g. whether some variable instances satisfy an equation, or whether a schedule satisfies some temporal constraints. In other areas testing requires more effort. For supporting code generation, for an example, the development of test cases such as in human based software engineering should be common practice. Requesting explanations and searching for sources could also be achieved with the help of LLMs to check the plausibility of answers. Prompt engineering will be a necessary craft to not only elicit solutions from an LLM but also to obtain explanations and arguments for the validity of the solutions. Finally, confidence in the results can be increased by orchestrated consulting of several LLMs independently to compare the results they return. Such a strategy could, for an example, be employed for natural language queries on databases. In general, the social systems we invented to manage human faultiness - requesting proofs and arguments for statements, checking results generated, triple redundancy check, or independent peer review, etc. - can be useful for designing systems while being aware of or being able to assess LLM limitations. It is critical to develop metrics and frameworks to assess the quality of generated content - the stage was already set by promising results that can inspire a specific research direction, e.g. RAGAs [51].

**Training LLMs with DS(M)Ls.** The reported successes with "understanding" UML, BPMN or ER models is clearly facilitated by the fixed structures and grammars imposed by well-known metamodels and the large corpora of examples or model serializations available to training. The situation is more challenging with DSMLs of narrow scope and limited reuse - their corpora of examples are inherently limited, and trainings must be organized by the specific communities interested in them. The conceptual modeling community must devise training protocols as well as benchmarks, possibly as extensions to existing language and method engineering methodologies such as [26] and [52].

**Gaining awareness on inherent limitations of stochastic systems.** Even without LLMs, scientific truth has been under attack in socio-political fora during recent years. Adding to this societal problem technological mechanisms of hallucinatory truth-making can be detrimental if AI-generated content that is fundamentally stochastic is perceived as being deterministic and authoritative. LLMs perform probabilistic predictions based on existing content, thus amplifying "as is" content through rearview mirrors against emergent knowledge that provides, from a probabilistic perspective, outliers.

Transferring this issue to automation projects, engineers must be fully aware when deterministic workflow patterns are most appropriate, in contrast to when stochastic decision support makes sense. Inferring causality from correlation is a major barrier of LLMs requiring complementing ingredients [53] and process automation works with fundamental temporal and causal relationships that should be reliably stable. Existing process automation approaches are sufficiently mature in their deterministic approaches, and in certain applications any process execution success rate lower than 100% is just unacceptable. Rather than looking towards a full overhaul of the process automation paradigm, process engineers should devise recipes for the orchestration of deterministic and stochastic patterns that cede control to LLMs only in the control points where that is affordable and useful.

**Educating AI usage models.** In contrast to often circulated assumptions, LLMs cannot be considered per se as a trustful source of knowledge since so far it cannot be traced how particular responses relate to the data that they have been trained on. In addition, the training data – in case it is being disclosed – may contain biased, false or even hallucinated data, not necessarily checked prior to feeding it into an LLM. At best, additional mechanisms applied during the training – e.g., such as reinforcement learning from human feedback – may prevent an LLM from delivering harmful or biased information. However, this is not the case for all LLMs as it requires considerable financial means. The output generated by LLMs so far needs to be assessed critically and anyone using the output of an LLM should be aware of the nature of the generative mechanisms.

From an educator perspective, the KG4SDSE workshop conclusion does not contradict the general idea that such systems should be embraced rather than banned (some recent guidelines are available directly from OpenAI<sup>6</sup>). Students must be encouraged

---

<sup>6</sup> <https://openai.com/blog/teaching-with-ai>

to critically assess, compare, and refine the results thus obtained, rather than taking them as ground truth. Although prompt engineering may be seen as a mechanism of continuous correction of a system prone to providing erroneous output, it may also be a key mechanism for corrective input or for empowering the human agent as a moral authority. However, this cannot be expected as default user behavior - training programs (in association with some kind of "AI driving license") may become a citizen-oriented necessity, to complement higher level regulatory acts such as the EU AI Act. LLMs can be manipulated not only in the interest of the prompters trying to glean a recommendation, answer or solution, but also in the interest of third parties' [54]. Moreover, there is always the risk of perceiving generative AI as a fact base - an anecdotal example during the workshop discussion referred to students contesting exams with incorrect multiple-choice responses obtained from ChatGPT - and justifying that "ChatGPT made me do it". Such a scenario goes beyond content generation, towards *action generation* in the human world - with non-discerning citizens becoming actuators of an oracle whose confidence persuades them in the absence of any sense of skepticism. Other concerns have been invoked related to the scientific publishing process, where not only article content, but also peer reviews are delegated to LLMs - potentially leading to a ChatGPT paper being evaluated by a ChatGPT review, thus rendering pointless the scientific dissemination process as a human endeavor of knowledge exchange.

**Updating authorship models.** The authorship relation for generated content is clearly disrupted. Taxonomies of contributor roles<sup>7</sup> have been a concern in scientific authoring for some time and integrity-oriented tools for preventing content misuse are well established. However, they are insufficient to manage the contribution of LLMs to scientific authoring - forbidding it can miss opportunities, while misusing it leads to sophisticated forms of plagiarism.

A question arises about the possibility of recognizing tools like ChatGPT as explicit co-authors of the content they produce - e.g. for summarizing background information, partly supporting literature surveys, recommending reusable experimental protocols found in the literature or helping with the structural organization of a paper. The productivity of scientific content creation may benefit, but good practices must be devised for proper handling of LLM's generative contribution - for the sake of productivity, and not as a replacement for the human endeavor of advancing science.

Novel authoring relationships can be devised - current policies dismiss the "author" quality of AI - see the CEUR-WS policy [55] or Elsevier's policy [56] - but they do mention different levels of AI assistance. More refined authorship models may distinguish content author from styling author, accountable author, organizer and bundler of resources and additional roles expanding the traditional contributor taxonomies.

---

<sup>7</sup> <https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement>

## 4 Research Agenda

In this section we formulate a research agenda that synthesizes the points previously raised. It is an agenda that serves the interest of the Conceptual Modeling discipline while looking at the larger context of Information Systems. Since these are fundamentally socio-technical systems, they can be approached from either the social or the technical side, while considering constraints, design rationale or requirements from the other side. Following the same criteria as for the expectations in the previous section, we distinguish in the following artifact-centric and social/organizational/human-centric research challenges:

### Artifact-oriented research challenges:

- Conceptual modeling is not only about knowledge structuring, but also about *knowledge visualization*. Although the dominant discussion about knowledge conversion focuses on the interplay between natural language and machine-readable knowledge graphs, the graphical expression of knowledge by means of diagramming has always been central to achieve the conceptual modeling goals of sense-making, communication and complexity reduction. Novel modeling methods are called to streamline knowledge conversion between textual, graphical and formal knowledge representation. At the same time, multi-modal LLM services are called to ensure knowledge consistency between the textual expression and the graphical expression, with metamodels potentially taking on the role of ontologies in RAG-like configurations and large repositories of visual models prepared for training and graph embeddings. Computer vision may play a role in the recognition of visual symbols, but the primary concern would be to leverage types and relationships, i.e. the dual nature of graphical modeling languages - as both visually comprehensible and machine-readable (with some non-visual attributes) by ways of model serialization formats;
  - Fine-tuning (e.g. LoRA [39]) is a major research direction towards having LLMs adapted to narrow-scoped purposes and the domain-specificity of choice. This corresponds to the long-term preoccupation of conceptual modeling with *domain-specific conceptualization* and agile adaption of modeling languages - typically approached through dedicated language engineering and method engineering approaches [26][52] that sacrifice generalization and consensus for the benefit of satisfying situational requirements or evolving requirements. LLM fine-tuning or RAG techniques must take into account the case of *evolving semantics*, reflecting the challenges of managing evolution in ontologies and DSML metamodels;
    - Extensive experimentation is reported about abstracting away technical query languages (SQL, SPARQL etc.) under a *natural language querying layer* capable of generating platform-specific queries under the hood. Results accuracy do not fulfil the deterministic expectation of the typical query client [57] and query complexity in successful demonstrations seems rather limited. However for simple, unambiguous retrieval needs it is an approach that promises to boost information retrieval productivity and this benefit can be transferred to model querying and model-based report generation (e.g. business process queries [58] being a prominent application area);

- *Prompt engineering as a modeling procedure*, recently reported with some promising results in [59], can be an alternative to the drag-and-drop user experience of conceptual modeling. Conversely, *modeling methods for prompt engineering* are also being investigated [60]. Design-oriented research must look into the adequacy of modeling grammars for such purposes, including domain-specific ones, and to formulate reusable techniques for making them effective for arbitrary requirements and modeling languages.;

- *Multi-perspective modeling* is a traditional approach to complexity reduction in enterprise modeling [61] (and for other inherently complex systems). It involves the enabling of shifting contexts by partitioning the metamodels or modeling grammars into distinguishable, possibly overlapping, viewpoints, raising challenges for maintaining the multi-view consistency. The extent to which such a flexible decomposition/recomposition of views can be supported by LLMs or RAG-supported LLMs is critical for involving them in the complexity reduction mission of conceptual modeling;

- *Model-driven engineering (MDE) is commonly understood as deterministic* – the same diagrammatic design decisions should not lead to different enacted behaviors across different deployments. When variation is desirable, it should be precisely controlled –by parameterization or environmental data acquired through sensory capabilities. Therefore, the uncertainty injected by LLMs along MDE processes must be carefully assessed by method engineering research and confined only to where it is affordable;

- The trade-offs in hybrid architectures (neuro-symbolic, graph-based RAG) must be scrutinized from an *energy-efficiency perspective*, to map as precisely as possible the ideal boundary between contextual priors/invariants that should be engineered and what is worth learning from data and natural language content, rather than pursuing the infinite scaling ambitions of ubiquitous learning.

#### **User/organizational-oriented research challenges:**

- In the traditional modeling tools, the *encoding and decoding of knowledge is symmetric*, i.e. relying on the same symbol mappings. However, it becomes asymmetric when modeling is performed through a conversational interface, or when model contents are textually summarized. Research must investigate comparative comprehension and cognitive effectiveness between the symmetrical and asymmetrical approaches. This concern also relates to perceived usefulness and ease-of-use studies, in the spirit of TAM (technology acceptance model) [62-63], or other types of studies concerned with the user experience of being involved in multi-modal knowledge conversion and transfer;

- The literature on knowledge management established the seminal SECI (socialization-externalization-combination-internalization) *knowledge conversion cycle* [64-65] at a time when the only participants in knowledge transfers were human employees and a generic notion of information systems. The SECI model must be revisited through the lens of involving LLM agents in any of its phases - e.g. human-chatbot socialization, multi-modal knowledge externalization or combination, conversational knowledge retrieval for internalization;

- The *level of agency attributed to LLM products* – from simple task automation, to decision-making, to communication and ultimately to behavior enactment – requires novel managerial, decision-support and work systems frameworks (or at least guidelines). They should be well informed about the stochastic nature of LLMs and their possible augmentations and adaptations (RAG, LoRA) in relation to expected deterministic outcomes, for example in business process automation. This also extends on a technical level to content management systems that must ensure distinction between human-authored content, machine-authored content and any relevant degrees in between – possibly explicitly marked by metadata to control use or provenance, for example to prevent LLMs getting trained on their own content towards “model collapse” [49], or on content that a human author would prefer to not expose to training corpora. Beyond the previously mentioned SECI updates, novel work theories and LLM adoption guidelines must ensure that the stochastic-deterministic or the generative-factual balances will be manageable in future work systems, which in turn requires updates towards novel AI-enhanced work system theories e.g. [66];

- Shifting the previous point to individual level, we are already facing a phenomenon of AI anthropomorphizing [67] which may be related to the excess of confidence in certain generative content and the misinterpretations of LLM reasoning capabilities – for example when using the umbrella term of “chain of thought” to blur the distinction between step-wise pattern matching and step-by-step logical problem solving [68]. There is a need for converging technical research that probes reasoning manifestations [69] with user perception studies on how those manifestations, and their “emergent” nature, lead to anthropomorphizing and also, paradoxically, to blind trust [70] that would not be granted to a human source of knowledge;

- The hereby discussed flavors of SDSE, as a specialization of model-driven engineering, may inspire *new software engineering methods that must be observed empirically* – i.e., to assess and measure how they play out when involving LLMs in different roles along the engineering process (as suggested by Fig. 2). This can incorporate various interpretations of waste or costs of systems engineering (e.g. from an energy efficiency viewpoint) in diverse configurations and trade-offs, as we already suggested from an artifact-oriented perspective.

## 5 The Panel: Key Statements and Consensus

This paper synthesized discussions and subsequent ideas initiated during the panel "Expectations from Large Language Models for Semantics-driven Systems Engineering", part of the workshop on "Knowledge Graphs for Semantics-driven Systems Engineering" (KG4SDSE)<sup>8</sup> held at CAiSE 2023 - the 35th International Conference on Advanced Information Systems Engineering<sup>9</sup> organized during 12-16 June 2023 by Universidad San Jorge in Zaragoza, Spain.

---

<sup>8</sup> [https://www.omilab.org/activities/events/caise2023\\_kg4sdse/](https://www.omilab.org/activities/events/caise2023_kg4sdse/)

<sup>9</sup> <https://caise23.svit.usj.es/>

The contributors are selected experts representing diverse research areas that involve knowledge engineering, conceptual modeling and systems engineering approaches:

- D. Karagiannis - having worked extensively at the intersection between domain-specific conceptual modeling and knowledge representation [23], typically in the context of OMiLAB activities<sup>10</sup>;
- U. Frank - author of the MEMO language and other seminal contributions to multi-perspective enterprise modeling [61];
- D. Plexousakis - having long term involvement in semantic query languages, knowledge representation and automated reasoning [71-72];
- J. Mylopoulos - author of foundational works inspiring today's Knowledge Graph technology, e.g. the Telos knowledge representation language [36];
- H. G. Fill - author of the SeMFIS modeling method for semantic annotation of diagrammatic models [73];
- J. Eder - having seminal contributions to temporal modeling, reasoning and constraints management [74];
- R. Buchmann (co-chair of the KG4SDSE workshop) - specializing in the interplay between Enterprise Modeling and Knowledge Graphs [75];
- E. Laurenzi (chair of the HybridAIMS workshop also held during CAiSE 2023<sup>11</sup>) - investigating the intersection of Enterprise Modeling and applied Hybrid AI [11-12,76];
- M. Y. Santos - contributing with perspectives from the field of Big Data analytics and business intelligence [77].

All the authors that provided input to this paper have worked predominantly with design-oriented approaches, employing design both as method and as an artifact to support the engineering of methods, tools, languages, ontologies, (meta)data management pipelines. While the majority of LLM-focused discussions during the preceding months have centered on the nature of the underlying technology, on advertising proofs-of-value demonstrators or on raising issues about ethics and responsibility, this panel focused on formulating expectations for the foreseeable evolution of this paradigm in relation to systems analysis, design and engineering - based on the contributors' own experience and experimentation with LLM-based products.

We finalize the paper with some key statements from each contributor to synthesize the spirit of the original discussion and of the positions hereby collected:

*LLMs have great potential to contribute to the construction and understanding of conceptual models. However, LLMs are statistical, as opposed to conceptual models; this deficiency can be ameliorated by adopting learning theories of educators such as Jean Piaget (J. Mylopoulos)*

---

<sup>10</sup> <https://www.omilab.org/nodes/>

<sup>11</sup> <http://hybridaims.com/>

*[LLM systems are] analogous to certain people who answer something at all costs, even when they don't have the answer - we despise this behavior from humans, should we allow it to a system? (J. Eder).*

*These are great times, and we should be enthusiastic as the advent of LLM is not something that threatens us, but rather something that poses a great challenge and creates a great opportunity. [...] Although LLMs require a lot of energy, they also distribute knowledge cheaply in a system available 24/7 (U. Frank).*

*LLMs may produce false and incorrect statements with a high degree of confidence. The lack of moral transparency and a sense of right and wrong leads to reduced trust. Ultimately the decision-making responsibility should remain with the humans, or humans in the loop (D. Plexousakis).*

*An important danger lays on data privacy, as sensitive data can be ingested in LLMs that would not comply with GDPR, which requires transparency with respect to how personal data is used as well as the right to delete it (H. G. Fill).*

*LLMs are trained on massive data sets of different sources and of varied quality. Veracity requires traceable and trustworthy data sets. Value, as veracity, needs reliable associations between these accurate data sets (M. Y. Santos).*

*LLMs are impressive in quickly generating new content but give bad results on things they have not been trained for. Knowledge graphs (KGs) perform at a constant quality level for the things they represent, and domain-specific models (DSMs) support the decision-making of domain experts. Still, both KGs and DSMs come with a high engineering effort and cost. The complementary strengths and weaknesses of LLMs, KGs and DSMs make them ideal to be combined to generate value in various application domains (E. Laurenzi).*

*LLMs have to reach a stage where they display meta-conceptual capabilities - being able to manage multiple abstraction layers and ultimately being able to generate meta-structures. Human intelligence uses abstraction to reduce complexity in a domain for a specific purpose, and the same ability should be targeted by Generative AI (D. Karagiannis).*

*The Semantic Web community must expand efforts towards supporting LLM training with availability of Linked Open Data, conceptual structures and even reasoning rule-bases. LLMs must reason under an open world assumption and training effort/costs should be reduced by leveraging open data and epistemological invariants. Conversely, the expectation is to bridge the gap between natural language queries and formal data queries of any complexity, thus expanding the value and usability of semantic datasets (R. Buchmann).*

In terms of a consensus, we point to the need for defusing a damaging synonymy between Artificial Intelligence and Machine Learning, or even Generative AI. The reduction of the AI field to only a few (popular) technologies or approaches is noticeable not only in the general public discourse but has also contaminated scientific literature, e.g. [46], leaving entire streams of research and achievements out of the AI label and out of what is superficially perceived as the AI scope - something that inevitably confuses junior researchers. On the other hand, the systems engineering and conceptual modeling communities must be more explicit in formulating value propositions for the AI field on multiple levels of granularity – from high level architecting to specific technical ingredients (e.g. for RAG).

We advocate the need to devise AI systems where the statistical matching/ranking approaches and grammar structures employed by ChatGPT [78] are complemented by deterministic design of invariants (time, causality etc.) and their associated logic-based mechanisms, something that can be deployed in multiple architectural configurations that SDSE is called to articulate and experiment with. A possible instantiation of these is the LLM-Cyc synergy recently advocated by the father of the Cyc knowledge base, the late Douglas Lenat [16]. Other works [79] suggest taking advantage of the proliferation of Knowledge Graphs for enriching the output of LLMs. Earlier works explicitly argued for a need to pair conceptual modeling with machine learning [80]. Therefore, we hereby contribute also to the position aiming to defuse the artificial dichotomy and the sense of competition that can be gleaned from statements such as “there’s no sign that knowledge engineering will ever be able to compete with machine learning outside of a few niche areas.” [81, p.36]. After all, Plato’s work on epistemology did not aim to compete with common human speech, but to strengthen its capabilities of argumentation and reasoning against rhetorical sophism, a situation to be mirrored by the field of Artificial Intelligence.

For the Information Systems field, this also implies that scientific research on AI-driven systems cannot stay at the level of empirical observation of black-box systems displaying some characteristics of intelligence, it needs to give equal importance to internal design decisions and to probe technical characteristics, looking for possible trade-offs for balancing different mechanisms of intelligence. Our design-focused stance is in line with an earlier memorandum for design-oriented science in information systems research [82] that finds now a very specific interpretation in LLM-focused research.

## **Acknowledgements**

We are thankful to OMiLAB NPO, Germany, and FORTH, Greece, for sponsoring the workshop where the ideas summarized here were initially debated. We also thank the workshop audience for engaging in the panel discussion, for sharing their own experience or raising inspiring questions, and not lastly the CAiSE 2023 organizers for hosting the event.

## References

1. Fill, H. G., Fettke, P., Köpke, J. (2023): Conceptual Modeling and Large Language Models: impressions from first experiments with ChatGPT. *Enterprise Modelling and Information Systems Architectures (EMISAJ) – International Journal of Conceptual Modeling*, 18: 1-15.
2. Jalil, S., Rafi, S., LaToza, T. D., Moran, K., Lam, W. (2023): ChatGPT and software testing education: promises and perils, in *IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, Dublin, Ireland, 2023, pp. 4130-4137, doi: 10.1109/ICSTW58534.2023.00078.
3. White, J., Hays, S., Fu, Q., Spencer-Smith, J., Schmidt, D. C. (2023): ChatGPT prompt patterns for improving code quality, refactoring, requirements elicitation, and software design, doi: 10.48550/arXiv.2303.07839
4. Feuerriegel, S., Hartmann, J., Janiesch, C., Zschech, P. (2023): Generative AI. *Business & Information Systems engineering*, doi: 10.1007/s12599-023-00834-7
5. Stahl, B. C. (2023): Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports* 13:7586
6. Ray, P. P. (2023): ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, *Internet of Things and Cyber-physical Systems*, 3:121-154, doi: 10.1016/j.iotcps.2023.04.003
7. Zhou, J., Muller, H., Holzinger, A., Chen, F. (2023): Ethical ChatGPT: concerns, challenges, and commandments, doi: 10.48550/arXiv.2305.10646
8. Ontotext (2023): Reflections on the Knowledge Graph Conference 2023, <https://www.ontotext.com/blog/reflections-on-the-knowledge-graph-conference-2023/>
9. Stardog (2023): LLM will accelerate knowledge graph adoption, <https://www.stardog.com/blog/llm-will-accelerate-knowledge-graph-adoption/>
10. Schmid, C., Laurenzi, E., Michelucci, U., Venturini, F. (2023): Explainable AI for the olive oil industry. In: *Proceedings of BIR 2023*. LNBIP vol 493. Springer, Cham. [https://doi.org/10.1007/978-3-031-43126-5\\_12](https://doi.org/10.1007/978-3-031-43126-5_12)
11. Rodorf D., Käser, J., Crego, A., Laurenzi, E. (2023): A hybrid intelligent approach combining machine learning and a knowledge graph to support academic journal publishers addressing the reviewer assignment problem (RAP). In: *Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE2023)*, California, USA, March 27-29, 2023.
12. Witschel H.F., Pande C., Martin A., Laurenzi E., Hinkelmann K. (2021) Visualization of patterns for hybrid Learning and reasoning with human involvement. In: *New Trends in Business Information Systems and Technology*. Studies in Systems, Decision and Control, vol 294. Springer, Cham. [https://doi.org/10.1007/978-3-030-48332-6\\_13](https://doi.org/10.1007/978-3-030-48332-6_13)
13. Tiddi, I., Schlobach, S. (2022): Knowledge graphs as tools for explainable machine learning: A survey. *Artif. Intell.* 302, 103627.
14. Van Harmelen, F., ten Teije, A. (2019): A Boxology of Design Patterns for Hybrid Learning and Reasoning Systems. *J. Web Eng.* 18, 97–124.
15. Opdahl, A., Tessem, B., Dang-Nguyen, D. T., Motta, E., Setty, V., Throndsen, E., Tverberg, A., Trattner, C. (2023): Trustworthy journalism through AI, *Data & Knowledge Engineering* 146:102182
16. Lenat, C., Marcus, G. (2023): Getting from generative AI to trustworthy AI: what LLMs might learn from Cyc, <https://arxiv.org/abs/2308.04445>
17. Brambilla, M., Cabot, J., Wimmer, M. (2017): *Model-driven software engineering in practice*, Springer.

18. Bock, A., Frank, U. (2021): Low-code platform, *Business & Information Systems Engineering*, 63:733-740, doi: 10.1007/s12599-021-00726-8
19. Di Ruscio (2022): Low-code development and model-driven engineering: Two sides of the same coin?, *Software and Systems Modeling* 21:437-44, doi: 10.1007/s10270-021-00970-2
20. Curty, S., Harer, F., Fill, H. G. (2023): Design of blockchain-based applications using model-driven engineering and low-code/no-code platforms: a structured literature review, *Software and Systems Modeling*, doi: 10.1007/s10270-023-01109-1
21. Buchmann, R. A., Cinpoeru, M., Harkai, A., Karagiannis, D. (2018): Model-aware software engineering - a knowledge-based approach to model-driven software engineering, *Proceedings of ENASE 2018, Portugal, SciTe Press*, pp. 233–240.
22. Hinkelmann, K., Gerber, A., Karagiannis, D., Thönssen, B., Van der Merwe, A., Woitsch, R. (2016): A new paradigm for the continuous alignment of business and IT: Combining enterprise architecture modelling and enterprise ontology. *Computers in Industry*. Vol. 79, pp. 77-86
23. Karagiannis, D., Buchmann, R. A. (2016): Linked Open Models: extending Linked Open Data with conceptual model information. *Information Systems* 56: 174-197
24. Hinkelmann, K., Laurenzi, E., Martin, A., & Thönssen, B. (2018). Ontology-based meta-modeling. In R. Dornberger (Ed.), *Business Information Systems and Technology 4.0*. Springer International Publishing. [https://doi.org/10.1007/978-3-319-74322-6\\_12](https://doi.org/10.1007/978-3-319-74322-6_12)
25. Laurenzi, E. (2024). Agile and ontology-based meta-modelling for the creation and maintenance of enterprise knowledge graph schema, *Enterprise Modelling and Information Systems Architectures* 19(SI), <https://doi.org/10.18417/emisa.19.6>
26. Karagiannis, D. (2022): Conceptual modelling methods: the AMME agile engineering approach, in *Domain-specific conceptual modeling - concepts, methods and ADOxx tools*, Springer pp. 3-21
27. Karagiannis, D., Buchmann, R. A., Utz, W. (2022): The OMiLAB digital innovation environment: Agile conceptual models to bridge business value with digital and physical twins for product-service systems development, *Computers in Industry* 138:103631, <https://doi.org/10.1016/j.compind.2022.103631>.
28. OMiLAB NPO (2023): The BEE-UP tool. <https://bee-up.omilab.org/activities/bee-up/>
29. Al Hasan Rony, M. R., Kumar, U., Teucher, R., Kovriguina, L., Lehmann, J. (2023): SGPT: A Generative Approach for SPARQL Query Generation From Natural Language Questions. *IEEE Access* 10:70712-70723, 2022, doi: 10.1109/ACCESS.2022.3188714
30. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, ., Sun, J., Guo, Q., Wang, M., Wang, H. (2024): Retrieval-augmented generation for large language models: a survey, <https://arxiv.org/abs/2312.10997>
31. Wu, H., Wang, Z., Wang, K., Ghiasnezhad Omran, P., Li, J. (2023): Rule learning over knowledge graphs: a review, *Transactions on Graph Data and Knowledge* 1(1):7:1-7:23
32. Allen, B. P., Stork, L., Growth, P. (2023): Knowledge engineering using large language models, *Transactions on Graph Data and Knowledge* 1(1):3:1-3:19
33. Pan, J. Z., Razniweski, S., Kalo, J. C., Singhania, S., Chen, J., Dietze, S., Jabeen, H., Ome-liyanenko, J., Zhang, W., Lissandrini, M., Biswas, R., de Melo, G., Bonifati, A., Vakaj, E., Dragoni, M., Graux, D. (2023): Large language models and knowledge graphs: opportunities and challenges, *Transactions on Graph Data and Knowledge* 1(1):2:1-2:38
34. Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., Wu, X. (2023): Unifying large language models and knowledge graphs: a roadmap, *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2024.3352100

35. Saba W. S. (2023): Stochastic LLMs do not understand language: towards symbolic, explainable and ontologically based LLMs, Proceedings of ER 2023, LNCS 14320, Springer, pp. 3-19
36. Mylopoulos, J., Borgida, A., Jarke, M., Koubarakis, M. (1990): Telos: representing knowledge about information systems. *ACM Transactions on Information Systems* 8(4): 325-362
37. Mendling, J., Malinova, M. (2022): Experimental evidence on the cognitive effectiveness of diagrams. *Proceedings of ISICO 2021, Procedia Computer Science* 197, pp. 10-15.
38. Cámara, J., Troya, J., Burgueño, L., Vallecillo, A. (2023): On the assessment of generative AI in modeling tasks: an experience report with ChatGPT and UML, *Software and Systems Modeling*, 22: 781–793.
39. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W. (2021): LoRA: low-rank adaptation of large language models. <https://arxiv.org/abs/2106.09685>
40. Zhu, D., Chen, J., Shen, X., Li, X., Elhoseiny, M. (2023): MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models, *Arxiv Pre-Print: 10.48550/arXiv.2304.10592*
41. Sandkuhl, K., Fill, H. G., Hoppenbrouwers, S., Krogstie, J., Leue, A., Matthes, F., Opdahl, A., Schwabe, G., Uludag, O., Winter, R. (2018): From expert discipline to common practice: a vision and research agenda for extending the reach of enterprise modelling, *Business and Information Systems Engineering*, 60(1): 69-80.
42. Lin, S., Hilton, J., Evans, O. (2022): TruthfulQA: measuring how models mimic human falsehoods, <https://arxiv.org/abs/2109.07958>
43. Kahneman, D. (2012): *Thinking, Fast and Slow*, Penguin.
44. Deloitte (2023): Knowledge graphs for financial services, <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/risk/deloitte-nl-risk-knowledge-graphs-financial-services.pdf>
45. Piaget, J. (1972): *The Psychology of Intelligence*. Totowa.
46. Verdecchia, R., Sallou, J., Cruz, L. (2023): A systematic review of Green AI. *WIREs data mining and knowledge discovery*, 13(4), e1507. <https://doi.org/10.1002/widm.1507>
47. Zhang, B., Davoodi, A., Hen Hu, Y. (2018): Exploring energy and accuracy tradeoff in structure simplification of trained deep neural networks. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8(4): 836–848.
48. Garcia-Martin, E., Rodrigues, C. F., Riley, G., Grahn, H. (2019): Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing* 134:75-88.
49. Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., Anderson, R. (2023): The Curse of Recursion: Training on Generated Data Makes Models Forget, *arXiv:2305.17493*.
50. Security Week, AI Hallucinated Packages Fool Unsuspecting Developers, <https://www.securityweek.com/ai-hallucinated-packages-fool-unsuspecting-developers/>
51. Es, S., James, J., Espinosa Anke, L., Schockaert, S. (2024): RAGAs: Automated Evaluation of Retrieval Augmented Generation. In *Proceedings of EACL: System Demonstrations*, pp 150–158. Association for Computational Linguistics.
52. Frank, U. (2013): Domain-specific modeling languages: requirements analysis and design guidelines. In: *Domain engineering*, Springer, p.133-157
53. Jin, Zh., Liu, J., Lyu, Zh., Poff, S., Sachan, M., Mihalcea, R., Diab, M., Schölkopf, B. (2023): Can large language models infer causation from correlation? <https://arxiv.org/abs/2306.05836>
54. Kumar, A., Lakkaraju, H. (2024): Manipulating large language models to increase product visibility. <https://arxiv.org/abs/2404.07981>

55. CEUR-WS (2023): CEUR-WS policy on AI assisting tools, <https://ceur-ws.org/ACADEMIC-ETHICS.html>
56. Elsevier (2023): Elsevier policy on the use of AI in writing, <https://www.elsevier.com/about/policies/publishing-ethics-books/the-use-of-ai-and-ai-assisted-technologies-in-writing-for-elsevier>
57. Sequeda, J., Allemang, D., Jacob, B. (2023): A benchmark to understand the role of knowledge graphs on large language model's accuracy for question answering on enterprise SQL databases, <https://arxiv.org/abs/2311.07509>
58. Polyvyanyy, A. (ed.) (2022): Process querying methods. Springer, <https://doi.org/10.1007/978-3-030-92875-9>.
59. Härer, F. (2023): Conceptual model interpreter for large language models. In ER 2023 Companion Proceedings (ER 2023 Forum), CEUR-WS 3618, [https://ceur-ws.org/Vol-3618/forum\\_paper\\_11.pdf](https://ceur-ws.org/Vol-3618/forum_paper_11.pdf).
60. Clariso, R., Cabot, J. (2023): Model-Driven Prompt Engineering. Proceedings of MODELS 2023, pp. 47-54, IEEE CS
61. Frank, U. (2014): Multi-perspective enterprise modeling: foundational concepts, prospects and future research challenges. *Software & Systems Modeling* 13(3):941-962
62. Saif, N., Khan, S. U., Shaheen, I., Alotaibi, F. A., Alnfai, M. M., Arif, M. (2024): Chat-GPT: validating Technology Acceptance Model (TAM) in education sector via ubiquitous learning mechanism. *Computers in Human Behavior* 154:108097.
63. Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A., Soomro, R. B. (2024): Extended TAM based acceptance of AI-powered ChatGPT for supporting metacognitive self-regulated learning in education: a mixed-methods study. *Heliyon* 10(8):E29317.
64. Nonaka, I., von Krogh, G. (2009): Tacit knowledge and knowledge conversion: controversy and advancement in organisational knowledge creation theory. *Organization Science* 20: 635-652, <https://doi.org/10.1287/orsc.1080.0412>.
65. Adesina, A. O., & Ocholla, D. N. (2019). The SECI Model in Knowledge Management Practices: Past, Present and Future. *Mousaion*, 37(3), <https://hdl.handle.net/10520/EJC-1c79a03422>
66. Alter, S. (2022): Agent responsibility framework for digital agents: roles and responsibilities related to facets of work. In Proceedings of BPMDS and EMMSAD 2022, LNBIP 450, Springer, pp. 237-252.
67. Alabbed, A, Javornik, A, Gregory-Smith, D. (2022): AI anthropomorphism and its effect on users' self-congruence and self-AI integration: a theoretical framework and research agenda. *Technological Forecasting and Social Change* 182:121786
68. Stechly, K., Valmeekam, K., Kambhampati, S. (2024): Chain of thoughtlessness: an analysis of CoT in planning. <https://arxiv.org/abs/2405.04776>
69. Panas, D., Seth, S., Belle, V. (2024): Can large language models put 2 and 2 together? Probing for entailed arithmetical relationships. <https://arxiv.org/abs/2404.19432>
70. Church, K. (2024): Emerging trends: when can users trust GPT, and when should they intervene? *Natural language engineering* 30(2):417-427.
71. Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D. (2002): RQL: a declarative query language for RDF. Proceedings of WWW 2002, USA, ACM, pp. 592-603
72. Basina, N., Patkos, T., Plexousakis, D. (2021): ECAVI: an assistant for reasoning about actions and change with the event calculus. *Domain-specific Conceptual Modeling*, Springer, pp. 457-477
73. Fill, H. (2017): SeMFIS: a flexible engineering platform for semantic annotations of conceptual models, *Semantic Web* 8(5): 747–763.

74. Eder, J., Panagos, E., Rabonovich, M. (1999): Time constraints in workflow systems. Proceedings of CAiSE 1999, Germany, Springer, pp. 286-300
75. Buchmann, R. A., Karagiannis, D. (2016): Enriching Linked Data with semantics from domain-specific diagrammatic models. *Business & Information Systems Engineering* 58(5): 341-353
76. Laurenzi, E., Hinkelmann, K., van der Merwe, A. (2018): An agile and ontology-aided modeling environment. Proceedings of PoEM 2018, Austria, Springer, pp. 221-237
77. Santos, M. Y., Costa, C. (2020): *Big Data: Concepts, Warehousing and Analytics*. River Publishers
78. Wolfram, S. (2023): *What is ChatGPT doing... and why does it work?* Wolfram Media
79. Mountanonakis, M. and Tzitzikas, Y. (2023): Using multiple RDF knowledge graphs for enriching ChatGPT responses , ECML PKDD 2023 Demos Track, Springer, pp 324–329
80. Maass, W., Storey, V. C. (2021): Pairing conceptual modeling with machine learning, *Data & Knowledge Engineering* 134: 101909
81. Domingos, P. (2017): *The Master Algorithm: how the quest for the ultimate learning machine will remake our world*. Penguin Books
82. Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E. (2011): Memorandum on design-oriented information systems research. *European Journal of Information Systems* 20: 7-10.