Research proposal:
Computer-Aided Design for safety-critical engineering systems: a Large Language Model approach

Olivier Cornes
Zurich

ABSTRACT

Safety and other ilities are critically important in modern engineering systems. Meeting requirements represent a significant fraction of development costs and schedules. Unlike other engineering disciplines, safety engineering does not have computational tools to assist designers and analysts, especially for radically new concepts. Attempting to predict unknown hazards through physical simulation is not computationally feasible. However, avoiding repeating accidents is. A central issue is that historical data is often too large and complex to process by humans in order to extract relevant de- sign information. The data relevant for safety and other ilities are usually stored as natural language (i.e. accident reports, regulations, news articles, online reviews). Classical natural language processing (NLP) has limited capability to process such texts. However, recent performance breakthroughs with Large Language Models (LLMs) suggest that the application of NLP in Computer-Aided Design is within reach. A design knowledge base could be built automatically from historical data (accident data/ regulatory documentation) and interacted with intuitively. The research presented here will focus on:

- How LLMs can be trained to retrieve relevant historical data and requirements for a given design/analysis problem?
- How an LLM-based system may be designed for safety-critical applications (explainability, reproducibility, transparency)?

The contribution of this research is the theoretical foundations of a new type of computer-aided design software targeted at improving the ilities (in particular safety) for the design and analysis of complex systems.

INTRODUCTION

Safety and other ilities are driving aspects of modern systems (Ref. 1). Societys most crucial engineering systems tend to be safety-critical (for example, transportation, energy conversion and storage and medical devices) Because of the risks involved (human and financial), research and development for these critical systems is limited and slow compared to other products (for example, consumer products). New technical advances such as rapid prototyping and advanced manufacturing are integrated slowly. Building safety into products represents a significant fraction of development costs and schedules (roughly 90% in the case of aircraft). Unlike other engineering disciplines (electrical, structural, fluid mechanics), safety has limited computational tools to assist during design, especially in radically new designs. Computer-Aided Design (in all its forms) has significantly accelerated the design and analysis of new products. Al- though safety engineering has evolved significantly to deal with safety in complex systems, identifying possible hazards is still central to safety engineering (Ref. 2), (Ref. 3). Hazard identification for a given design task is currently undertaken without computer assistance and is primarily based on engineering judgment. Hazard identification requires creative thinking, and identifying previously unknown hazards requires cognitive capabilities (such as compositionality) that are out of reach of current computing. However, preventing past accidents from reoccurring is achievable through lesser means, such as pattern recognition. The data encoding experience relevant for safety is usually in natural language form. The goal of this research is to investigate if modern natural language processing techniques (Large Language Models in particular) can identify relevant hazards from the history of similar design problems. Although a wide variety of ilities could be improved using such techniques, this work focuses on safety. The domain of application is restricted to the design of electro-mechanical systems with aerospace applications.

RESEARCH QUESTIONS

The underlying assumptions for this research are:

- Although new technology has the potential to create new types of hazards, most of them can be avoided by learning from historical data.
- Safety-relevant data is stored in natural language form (i.e. lessons learned, accident reports, news articles, regulations)
- A given design or system architecture can be expressed in natural language form

- Once a relevant hazard is brought to the attention of the de- signer, the necessary steps will be taken to mitigate it.

The broader research question is:

Can recent developments in artificial intelligence improve the development process of safety-critical engineering systems?

Given the scope of the research and the proposed methodology, this can be reformulated into:

Can the design of safety-critical electromechanical systems be assisted by Large Language Model-based computational tools by informing the designer of relevant hazards and requirements?

LITTERATURE REVIEW

Safety Engineering

Current safety engineering methods (for example, Functional Hazard Analysis, Fault Tree Analysis, Failure Mode, effects and criticality analysis) (Ref. 3) work well for classical electromechanical systems but encounter limitations as systems become complex. This is especially true with the increased use of software/autonomy (Ref. 2). In addition, these approaches do not scale: each safety aspect must be analysed by an expert on the system in question. New methods such as STPA (Ref. 2) deal with complexity far better, as they:

- View the system globally, and focus on interactions within and around the system
- Approach safety as a control problem, i.e. the system must be controllable/steered away from undesirable states, such as accidents

However, they still do not address the issue of scalability: the analysis must still be performed by experts to identify the hazards (undesirable states) to be avoided. Although removing humans completely from safety-critical decision making is undesirable, the assistance of an expert system would significantly accelerate and improve safety-engineering work.

Expert Systems

The idea of encoding knowledge into computers is not new: expert systems have been experimented with since the early 1970s (Ref. 4). The application of Expert-Systems to engineering sprouted the field of Knowledge-Based Engineering (KBE) (Ref. 5). Modern KBE suffers from a tendency to develop black-box applications (Ref. 6), (Ref. 7), which is not acceptable for critical decision making, where explainability and traceability are necessary. In addition to this, expert systems remain expensive to develop and maintain (Ref. 7). In order to be effective, such systems should be built and updated with little to no human intervention.

Natural Language Processing & knowledge

Natural Language Processing (NLP) is the critical discipline for this research project since data relevant for safety is mostly in natural language form. Classical NLP is based on short-range dependency between words (for example, word embeddings, Latent Dirichlet Allocation). This is insufficient for processing technical documentation, which requires capturing longer-range dependency. LLMs have shown significant potential in various NLP tasks (Ref. 8), including the ones related to knowledge graphs and question answering (Ref. 9).

RESEARCH DESIGN (METHODOLOGY)

Predicting hazardous events is computationally not feasible because it would require capturing the full complexity of the real world. Despite enormous advances in computing power, predicting unknown unknowns is not possible. However, avoiding repeating accidents is:
The main issue is that the historical data is too large (be it as accident reports or as regulations) for a single or even a group of individuals to parse to extract relevant design information. The data relevant to safety is stored as natural language (accident reports, regulations, news articles, and online reviews). This research aims to apply modern natural language processing techniques to automatically build and maintain expert systems for safety engineering. Unlike many other tasks, safety-related decision making needs to be transparent (explainable, traceable). The proposed approach to ensure explainability is to combine subsymbolic models (LLMs) with symbolic models (knowledge graphs). Unlike many other tasks, safety-related decision making needs to be transparent (explainable, traceable). The proposed approach to ensure explainability is to combine subsymbolic models (LLMs) with symbolic models (knowledge graphs).

The research will be driven around:

- Developing symbolic models that capture the type of knowledge necessary for hazard identification and safety-requirement selection
- Understanding how LLMs can be adapted to build and interface with knowledge graphs for tasks such as explainable question answering
- Develop combined symbolic/subsymbolic architectures that can retrieve relevant historical documents (i.e. accident reports) and requirements for a design problem (at first, formulated as a description/question).

Project Limitations (Scope)

Even though the research presented here can be applied many critical domains, such as the identification of financial, geopolitical, or healthcare risks, the scope is at first restricted to the design and development of electromechanical systems in the

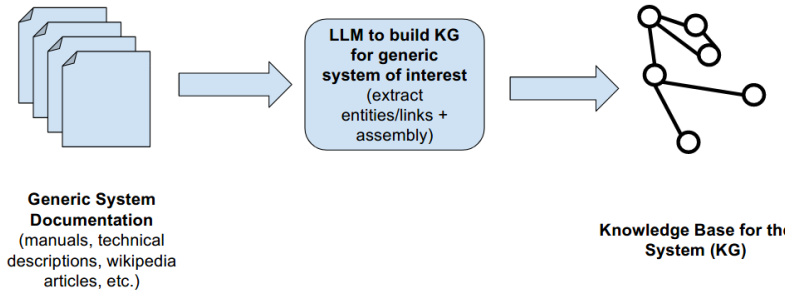

Fig. 1: *Illustration of a hybrid subsymbolic/symbolic architecture, where a LLM (subsymbolic) is used to build a KG (symbolic) model).*

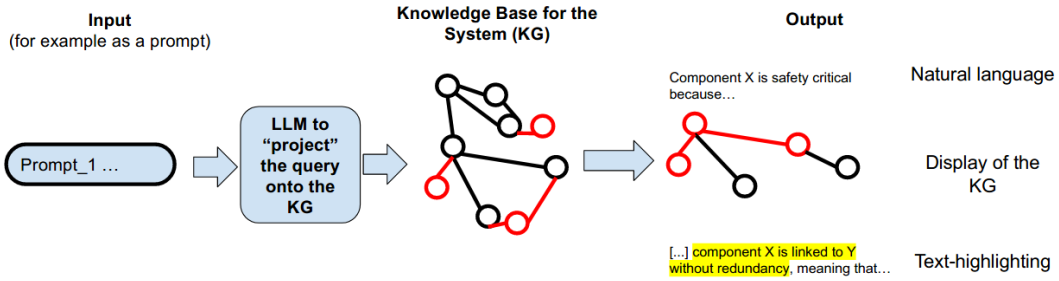

Fig. 2: *Illustration of a hybrid subsymbolic/symbolic architecture, where a LLM (subsymbolic) is used to project a query on a KG (symbolic) model for explainable results*

aerospace context.
Also, the type of input/output data is limited to natural language. Extensions to include other forms of data (images or tabular data) could be added as future work.

Analysis

The original research hypothesis is validated if a computational tool can be built which:

- Can compile domain-specific text data into a knowledge base for safety application automatically
- Can be used to reliably identify (catch) hazards that a known to occur on a given system/design problem. This would be a case study which is either created by humans or a historical case similar to those contained in the training set.
- Is explainable: the decision making process of the algorithm can be understood

Equipment

LLMs require large resources to be trained. These are typically only available to large internet companies (Google, Facebook, OpenAI, etc.). But fine tuning a LLM such as BERT for a specific application requires significantly less resources: A model such as BERT can be fine tuned on single GPU (1 GB of memory and 5 Teraflops) in a matter of minutes (Ref. 10) The most computational intensive task for this research is expected to be fine-tuning LLMs specific tasks such as building knowledge graphs.
The resources necessary for the research is on the order of a single GPU, typically available on Colab (Ref. 11) or an a research cluster such as EULER (ETH) (Ref. 12).

SIGNIFICANCE AND IMPLICATION OF THE STUDY

The impact of the research is twofold:

- The theoretical foundations of a new type of computer-aided design software targeted at the design and analysis of complex systems by automatically identifying safety-relevant information.
- Development of explainable models by combining LLMs with symbolic models such as knowledge graphs.

BIBLIOGRAPHY

*

References

[1] O. L. De Weck, D. Roos, and C. L. Magee, "Engineering systems : meeting human needs in a complex technological world," p. 213, 2011.

[2] Nancy G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge: MIT Press, 2011.

[3] D. Kritzinger, *Aircraft System Safety: Military and civil aeronautical applications*. Cambridge, UK: Woodhead Publishing, 2012.

[4] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*. Penguin Books Ltd, 2019.

[5] C. Dym and R. Levitt, *Knowledge-Based Systems in Engineering*, 1st ed. New York: McGraw-Hill Book Company, jan 1991. [Online]. Available: https://scholarship.claremont.edu/hmc{_}facbooks/15

[6] W. J. Verhagen, P. Bermell-Garcia, R. E. Van Dijk, and R. Curran, "A critical review of Knowledge-Based Engineering: An identification of research challenges," *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 5–15, 2012.

[7] E. J. Reddy, C. N. V. Sridhar, and V. P. Rangadu, "Knowledge Based Engineering: Notion, Approaches and Future Trends," *American Journal of Intelligent Systems*, vol. 5, no. 1, pp. 1–17, 2015.

[8] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems*, vol. 2020-December, may 2020. [Online]. Available: https://arxiv.org/abs/2005.14165v4

[9] M. Yasunaga, H. Ren, A. Bosselut, P. Liang, and J. Leskovec, "QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering."

[10] C. McCormick, "GPU Benchmarks for Fine-Tuning BERT · Chris McCormick," jun 2020. [Online]. Available: https://mccormickml.com/2020/07/21/gpu-benchmarks-for-fine-tuning-bert/

[11] "Welcome To Colaboratory - Colaboratory." [Online]. Available: https://colab.research.google.com/{#}scrollTo=Nma{_}JWh-W-IF

[12] "Euler - ScientificComputing." [Online]. Available: https://scicomp.ethz.ch/wiki/Euler