Book Review

Systems Engineering and Artificial Intelligence

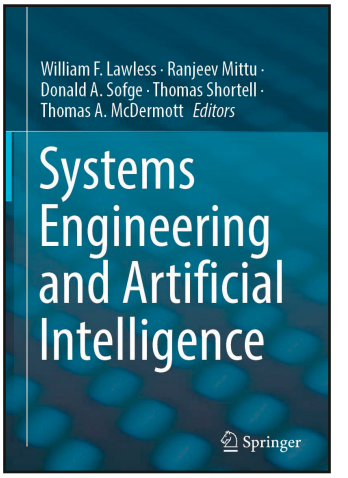
William F. Lawless, Ranjeev Mittu, Donald A. Sofge, Thomas Shortell, and Thomas A. McDermott

Springer 2021, ISBN: 978-3-030-77282-6

Jesus Garcia Herrero
Universidad Carlos III de Madrid
Colmenarejo, Spain
jgherrer@inf.uc3m.es

INTRODUCTION

This book presents a wide collection of analysis and examples around the design of Artificial Intelligence (AI) and Machine Learning (ML) systems from the point of view of Systems Engineering (SE). Some relevant aspects covered in the book are, among others, verification and validation of complex systems based on AI/ML, autonomy, emergent behavior, and human-machine teaming. It contains 25 chapters, providing a rich variety of views, disciplines, and examples in different domains, with an extensive analysis of literature on these topics.



Artificial Intelligence is one of the most disruptive technologies in recent years, boosted in part by the decision of big technological companies to integrate it into their business models. It has already shown a significant economic impact worldwide, with even more economic and social impacts to come [1], [2]. According to the Gartner CIO 2019 survey, the volume of organizations that have implemented AI has grown by 270% in the period 2015–2019 [3].

The basic idea behind traditional ML methods is training computer algorithms with data collected in a domain to learn a certain behavior (e.g., self-driving cars) so that an outcome can be produced by the computer algorithm when it is presented with a novel situation [4]. A methodological approach is needed to put these systems in complex, dynamic situations, after following appropriate testing and evaluation methodology. In the Gartner forecast report for 2021 [5], the importance of incorporating AI engineering into business strategy is cited to make investments in AI profitable by improving performance, scalability, interpretability, and reliability of AI-based models.

The book motivates thinking about the open challenges and important issues to develop intelligent, autonomous systems. Interaction and collaboration in human-machine teams, including context sharing to improve mutual understanding, is an interesting view, contrasting with the fear of autonomous, opaque machine learning algorithms that eventually may outperform human skills. This need to improve the understanding of autonomy is associated with the timely decisions that may need to be made faster than humans can process [6], mentioning as examples military scenarios and the push for quicker command, control, and communication upgrades, and also the common use of AI in transport systems like self-driving cars, trucks, ships, or subways.

The application of AI/ML raises several concerns and questions for SE. The usual procedures and metrics of formal verification, certification, and risk assessment must be defined for autonomous systems at the design, operation, and maintenance stages [7], [8]. Specifying performance metrics for emergent behavior opens interesting questions, such as how systems engineers shall assure that the "pieces work together to achieve the objectives of the whole" [9], how to define metrics to assess the risks associated with collaboration, and also how they can be calculated [10].

A central aspect is defining interdependence from a system's perspective, analyzing the interactions and interfaces among subsystems, and dealing with the whole system across its life cycle. Interdependence is a very relevant term in SE, AI, and the science of human–machine teamwork. As hypothesized by the editors, "the best teams maximize interdependence to communicate information via constructive and destructive interference", the optimum team size occurs when they are freely able to choose to minimize redundant team members [11], [12].

CHAPTERS REVIEW

Given the big size of the book, only a few chapters have been selected for this review, to summarize their contents and give a more detailed idea of the book's scope. This selection is a "sample" in the sense that it would serve for this purpose of presenting the main ideas and offer very interesting illustrative examples focused on SE, autonomy, or human-machine interaction. Readers are encouraged to go through all of the chapters to get a rich collection of thoughts, practices, and examples from different perspectives.

CHAPTER 2. "RECOGNIZING ARTIFICIAL INTELLIGENCE: THE KEY TO UNLOCKING HUMAN AI TEAMS" BY PATRICK CUMMINGS, NATHAN SCHURR, ANDREW NABER, CHARLIE, AND DANIEL SERFATY

This chapter was prepared by a team that included an artificial embodiment, "Charlie", in collaboration with other three human co-authors. It presents direct insights generated after coworking with Charlie: how she came into existence, how she operates in public, and how she can be influenced by both human and artificial coworkers and by their contributions.

The chapter starts by distinguishing the two different types of human-IA collaboration and embodiment internal state: sup-

portive collaboration, in which a human and an AI agent together serve as a single member for the team; and participatory collaboration, in which the AI agent is an individual team member, a situation where the AI agent communicates and coordinates with fellow human teammates, a fundamental aspect linked to the progress of the AI field.

The authors present Charlie's embodiment interface and the iterations to refine her communication and representation states driven by feedback. As indicated by the authors, in the case of chatbots, response delays may be acceptable, especially in responses to other panelists. However, to participate with physical and audible queues—gestures used by humans—Charlie had to effectively coordinate the use of the display and audio to achieve a similar presence and clearly represent its internal states.

Many people are aware of sophisticated conversational agents like Watson or AI Debater, thanks to public demonstrations. Moreover, most people interact frequently with conversational agents as customer service chatbots and virtual personal assistants. The developments presented in this chapter around Charlie touch several AI domains, from AI interaction with humans (covering user interface or explainability from AI to human) to integrations into a workplace or team. As reported, Charlie showed advanced capabilities for interaction, such as participating in a panel discussion, speaking during podcast interviews, contributing to a rap battle, catalyzing a brainstorming workshop, and even collaboratively writing the chapter.

> ***"Interaction and collaboration in human-machine teams, including context sharing to improve mutual understanding, is an interesting view contrasting with the fear of autonomous, opaque machine learning algorithms."***

CHAPTER 3. "ARTIFICIAL INTELLIGENCE AND FUTURE OF SYSTEMS ENGINEERING" BY THOMAS A. MCDERMOTT, MARK R. BLACKBURN, AND PETER A. BELING

This chapter was written by experienced systems engineers, who review the transformations expected in their area due to new digital tools for modeling "digital twins", resulting in the integration of data and modeling. They refer to an envisioned long-term outcome, "Human–Machine Co-Learning", referring to a future scenario where both humans and machines will adapt their behavior over time by learning from each other or alongside each other. This new context implies a significant transformation of SE methods, processes, tools, and practices over time.

To achieve this end state, the authors consider several "waves" or eras to be visited by AI and SE disciplines. The first of them includes "Explainable AI", covering technologies and approaches that make the decisions produced by AI systems more transparent to human developers and users. It also includes more transparency and understanding of the methods and tools used to develop AI applications, the underlying data, and the human–machine interfaces that lead to effective decision-making in the type of complex systems SE deals with routinely.

Secondly, the "robust and predictable" wave is to produce systems that learn and may be non-deterministic, but also appropriately robust, predictable, and trustworthy systems, using common aspects to the application of SE practices today. This wave particularly includes both human and machine behaviors in joint decision environments, highly reliant on good human-system design, and presentation of decision information. It also includes the adaptation of test and evaluation processes to co-learning environments.

Finally, the third wave involves systems that adapt and learn dynamically from their environments. In this wave, machine-to-machine and human-to-machine (in both directions) trust will be critical. Trust implies a dependence between the human and machine, and it normally requires the human to understand and validate the performance of the system against a set of criteria in a known context.

The authors identify the key research areas to achieve these waves, such as data collection and curation, ontological modeling, information presentation, digital twin automation, explainability, etc., identifying their use and association with the goals indicated. For instance, the human analysis and decisions will require better understanding and trust in the machine-generated analysis and decisions, or cognitive bias induced in sampled data or algorithms must be reduced to avoid unexpected results of the system making it inappropriate for use.

CHAPTER 6. "SYSTEMS ENGINEERING FOR ARTIFICIAL INTELLIGENCE-BASED SYSTEMS" BY JAMES LLINAS, HESHAM FOUAD, AND RANJEEV MITTU

In this chapter, the authors give a historical review of SE for AI-based systems over time. The chapter starts with a brief history of AI and its main categories (narrow, weak, and strong AI) and an interesting taxonomy of all areas of research, organized in AI techniques and problems addressed [13]. Regarding SE, the current engineering challenges of systems-of-systems and enterprise systems are reviewed, how the concept of life cycle has been evolving thanks to Agile development first, and then to "DevOps" methodologies, to increase the link between software development and IT operations.

Software engineering changes are also included in the review, referring to the well-known "Waterfall" development methodology originally proposed for "large computer programs", and its current progression to face challenging aspects

of developing AI/ML systems, such as their strong dependency on the data used. The AI/ML life-cycle model requires dealing with data, selecting a target model, and training and testing it under different configurations and performance metrics. The process must define the logic involved in selecting the data to learn in relation to targeted purposes of the application, requiring non-trivial domain knowledge, and considering non-linear interdependencies [14]. In addition, another key step in SE is formed by test and evaluation processes. Model-Based Test defines models for describing test environments and strategies, generating test cases, etc. to trace the correspondence with requirements and models used in design. In the case of ML, the problem is about model testing for classification to a great degree, and AI is about possibly complex layers of inferencing. Some paths are mentioned for the selection of the test and evaluation processes and metrics for both types of systems.

Finally, the emergent behavior is addressed, as a key property of complex systems, linked to the open issue of explainability in AI and ML. The authors close the chapter by discussing the challenge of AI explanations and explainability, with the aim to solve the Black Box problem through post-hoc analysis, or in an alternative approach using interpretable systems. Impenetrableness of most AI/ML systems comes from difficulties in knowing how inputs are transformed into outputs and which environmental features and regularities are being used.

CHAPTER 8. "RE-ORIENTING TOWARD THE SCIENCE OF THE ARTIFICIAL: ENGINEERING AI SYSTEMS" BY STEPHEN RUSSELL, BRIAN JALAIAN, AND IRA S. MOSKOWITZ

This chapter is focused on engineering design aspects of AI-enabled systems, explained by authors in the military domain, where these systems are becoming pervasive and must face specific challenges. They discuss hierarchical component composition in a system-of-systems context and focus on the stability problems for this type of complex systems, in relation to their level of connectedness. As indicated, system instability appears when emergent behaviors that are not anticipated take place. Moreover, the logic incorporates ML models, which depend on the data used to build them and the data with which it interacts. Therefore, the importance of bounding data for stable learning is highlighted.

Another aspect that is highlighted are the design/engineering problems of interoperability since AI systems usually operate as an element of a multi-component system. The authors refer to the discipline of systems theory, which emphasizes understanding the behavior of the system (e.g., a realized assembly) as a function of the behavior and interaction of its constituent elements (components). Challenges in cascading deployment are particularly relevant to AI systems because the boundaries that typically define system locality can be greatly expanded and obfuscated, leading to emergent system behaviors.

Finally, the presence of uncertainty in any system process opens an opportunity for emergent behavior that expands the boundary of the system's functions. Modern AI is particularly prone to introducing uncertainty into its outputs because of its reliance on ML algorithms. The authors illustrate system engineering problems described in the chapter, using examples of natural language processing (NLP). NLP tasks typically require multiple ML methods to be applied sequentially to achieve the objective of content understanding and are impacted by uncertainty in each step. The NLP system was designed to process the content of weekly activity reports prepared by information science researchers at the Army Research Laboratory, aimed to identify documents about similar topics and present a graphic summary of the relationships found. The example was used to illustrate the described challenges of AI system engineering and how the manipulation of a few hyperparameters made the experimental AI system significantly change its output.

> ***"The application of AI/ML raises several concerns and questions for SE. The usual procedures and metrics of formal verification, certification and risk assessments must be defined for autonomous systems at the design, operational and maintenance stages."***

CHAPTER 10. "DIGITAL TWIN INDUSTRIAL IMMUNE SYSTEM: AI-DRIVEN CYBERSECURITY FOR CRITICAL INFRASTRUCTURES" BY MICHAEL MYLREA, MATT NIELSEN, JUSTIN JOHN, AND MASOUD ABBASZADEH

This chapter describes advances of AI/ML systems to detect cyberphysical anomalies, illustrated with the development of GE's Digital Ghost, a system aimed at improving the security, reliability, and resilience of the power grid in the United States. The design of threat detections for Digital Ghost included machine learning algorithms in combination with deep domain knowledge of the physics for the systems to establish a matrix of credible cyber-attacks, naturally occurring faults, and vulnerabilities.

In addition, the authors review the new challenges coming to make human-machine teams effective against any threat, cyber or physical. The authors review research areas related to the design of this system such as explainable AI, invariant learning, and humble AI as critical techniques for improving the data fusion, trustworthiness, and accuracy of AI-driven technology and its application in empowering human-machine teams.

Humble AI is identified as valuable to marry man and machine, addressing aspects such as how the algorithms can alert the operator of a potential decrease in accuracy or confidence in its threat classification results, if an extrapolation is done into

"Interdependence is a very relevant term in Systems Engineering, AI, and the science of human-machine team work. As hypothesized by the editors, 'the best teams maximize interdependence to communicate information via constructive and destructive interference'."

previously unseen operating regions, etc. Subsequently, an operator would see if Digital Ghost should halt operations or continue but express reduced confidence in its results.

Explainable AI, as a feature of AI-based machines to explain the reasoning underlying their decisions in a way understandable to humans, is identified as a key to develop intuitive and trustworthy explanations for decisions provided by AI algorithms. As pointed out, the most successful human-machine teams will collaborate by employing interfaces containing easy-to- understand visualization techniques. This result is essential for machines to be trusted in making autonomous/semi-autonomous decisions, especially for kinetic platforms that are increasingly autonomous, as well as for safety and other mission-critical applications that determine diagnostics and cyber-physical security.

CHAPTER 14. "CONTEXTUAL EVALUATION OF HUMAN–MACHINE TEAM" BY EUGENE SANTOS JR, CLEMENT NYANHONGO, HIEN NGUYEN, KEUM JOO KIM, AND GREGORY HYDE

In this chapter, the authors explain examples of designing human-machine teams for the domains of healthcare and disaster relief. They highlight the importance of explanations in these hybrid teams to improve efficiency and productivity in complex dynamic environments.

For most human-machine team settings, typical metrics used for end users are insufficient to describe performance, and more explanations are crucial because they help understand a team's operational dynamics and identify the shortcomings that individual agents introduce to the team. These explanations allow performance predictions and quickly identify the shortcomings that individual agents (human or machine) introduce to the team.

One of the key terms proposed is interference, used to capture a team's interactional processes. Interference occurs when the goals of one agent affect the goals of the other agents [15], either in positive or negative ways. Interference is likely to arise due to differences in communication mechanisms, roles, capabilities, adaptiveness, and responsibility between humans and machines.

In order to reflect the underlying agent goals and preferences, in relation to behavior, the authors propose the use of rewards in the framework of inverse reinforcement learning (IRL) using a team's past behavior. The rewards are mapped with high-level behavioral attributes (behA) that are connected to a team's performance metrics. These behA provide insights that will then help to explain a team's performance in relation to agent behaviors.

Finally, authors analyze different compositions of team members that complement each other, with experiments to study the effect of each individual's behA into human-machine team interference. The results suggest that predictions of team attributes reflect actual team behaviors, encouraging further research on the lines presented.

CHAPTER 20. "ENGINEERING CONTEXT FROM THE GROUND UP" BY MICHAEL WOLLOWSKI, LILIN CHEN, XIANGNAN CHEN, YIFAN CUI, JOSEPH KNIERMAN, AND XUSHENG LIU

The authors present a detailed use case to analyze human-machine collaboration, using sensors, speech, and gesture inputs. The problem used to exemplify the process is the collaboration between a person and a Sawyer robot to solve physical block-world assembly problems. The human collaborator defines the problem to be solved and gives instructions to the robot, either step-by-step or at a higher level. It is a modular design, where context is maintained on a shared board to keep the information needed for problem-solving and shared for members of the team.

The physical system contains a camera, a depth sensor (in Kinect V2 for Xbox One), and a laptop microphone. The Kinect and the camera are located in a fixed space, overlooking a table-sized interaction space. The robotic arm has its own camera attached to it, near the gripper. The main information sources and processing modules are:

- Data from sensors are interpreted to determine the locations of objects in the collaboration space, in the framework of the interaction process. Block locations are stored relative to the position of the camera.
- For gesture recognition, pointing gestures of a human collaborator are interpreted by means of utilities provided by the Xbox Kinect V2 software, describing a hand's skeleton to determine the direction and location of the pointing actions. It provides a heuristic value that estimates each block's certainty of being identified by a gesture.
- For Speech-to-Text, it uses the Google Cloud service that receives an audio snippet of the speech input and sends back a string representing the spoken text.
- For text parsing, the Stanford CoreNLP library's dependency parser is used to annotate sentences with both universal dependencies and parts-of-speech tags. This information is stored in a Semantic Graph object by the parser.

The information for processing tasks is represented using the Unstructured Information Processing Architecture (UIMA,

2019) developed by the IBM Watson team. UIMA contains the "Common Analysis System", or CAS. Similar to a blackboard architecture, a CAS object serves to capture information in various stages of refinement. Some examples of text parsed in UIMA are shown, capturing information through software components called annotators, so that new pieces of information are stored for future interactions.

SUMMARY

AI, SE, and human–machine teamwork are linked by interdependencies, a key aspect to be considered in the design of complex systems. The book contains a big sample of the research and development areas which are evolving to integrate the advances in AI and ML technologies in complex and critical systems. New systems engineering methods, processes, and tools must be developed to cover the emerging AI and ML technologies and their new applications in order to make these systems reliable, safe and secure.

These challenges must be addressed to move AI/ML systems forward to operate in real conditions, interact tightly with humans, and meet expectations in complex, dynamic situations. The book shows the way towards developing the next generation of AI/ML systems designed with engineering methods to provide assured cost-benefits while achieving desired effectiveness.

REFERENCES

1. Brynjolfsson, E., and Mitchell, T. What can machine learning do? Workplace implications: profound changes are coming, but roles for humans remain. *Science*, Vol. 358, (2017), 1530–1534.
2. Delponte, L. European artificial intelligence (AI) leadership, the path for an integrated vision. European Parliament's Committee on Industry, Research and Energy, 2018.
3. Howard, C., and Rowsell-Jones, A. 2019 CIO Survey: CIOs have awoken to the importance of AI. *Gartner*, Jan. 2019.
4. Raz, A. K., Llinas, J., Mittu, R., and Lawless, W. Engineering for emergence in information fusion systems: a review of some challenges. In *Proceedings of the Fusion 2019*, Ottawa, Canada, Jul. 2019, 1–8.
5. Panetta, K., Gartner top strategic technology trends for 2021. *Gartner*, 2020.
6. Horowitz, B. Introduction of the life cycle-ready AI concept. In *Proceedings of the SERC Workshop: Model Centric Engineering*, Georgetown University, Washington, DC, Apr. 2019.
7. Lemnios, Z. IBM research. In *Proceedings of the SERC Workshop: Model Centric Engineering*, Washington, DC, Apr. 2019.
8. Richards, R. Program manager at DARPA, invited talk. In *Proceedings of the SERC Workshop: Model Centric Engineering*, Washington, DC, Apr. 2019.
9. Thomas, J. INCOSE discussion. In *Proceedings of the SERC Workshop: Model Centric Engineering*, Washington, DC, Apr. 2019.
10. Grogan, P. Game-theoretic risk assessment for distributed systems. In *Proceedings of the SERC Workshop: Model Centric Engineering*, Washington, DC, Apr. 2019.
11. Lawless, W. F. The entangled nature of interdependence bistability, irreproducibility and uncertainty. *Journal of Mathematical Psychology*, 78 (2017), 51–64.
12. Lawless, W.F. The physics of teams: Interdependence, measurable entropy and computational emotion. *Frontiers of Physics*, 5 (2017), 30.
13. Corea, F. *AI Knowledge Map: How to Classify AI Technologies*. Cham: Springer, 2019, pp. 25–29.
14. Amershi, S., et al. Software engineering for machine learning: a case study. In *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, Montreal, Canada, 2019.
15. Castelfranchi, C. Modelling social action for AI agents. *Artificial Intelligence*, 103 (1–2), 157–182.