Digit 2023 Proceedings

Diffusion Interest Group In Information Technology

12-1-2023

Empowering Domain Experts in Developing AI: Challenges of bottom-up ML development platforms

Stefan Thalmann
University of Graz, stefan.thalmann@uni-graz.at

Vladislav V. Fomin
Vilnius University, Vladislav.Fomin@knf.vu.lt

Isabel Ramos
University of Minho, iramos@dsi.uminho.pt

Thomas Gremsl
University of Graz, Thomas.Gremsl@uni-graz.at

Eugene Mourzine
Rockwell Automation, emourzi1@ra.rockwell.com

Empowering Domain Experts in Developing AI: Challenges of bottom-up ML development platforms

Research-in-Progress Paper

Stefan Thalmann
University of Graz
Graz, Austria
Stefan.Thalmann@uni-graz.at

Vladislav V. Fomin
Vilnius University
Kaunas, Lithuania
Vladislav.Fomin@knf.vu.lt

Isabel Ramos
University of Minho
Guimarães, Portugal
iramos@dsi.uminho.pt

Thomas Gremsl
University of Graz
Graz, Austria
Thomas.Gremsl@uni-graz.at

Eugene Mourzine
Rockwell Automation
Cleveland, OH
emourzi1@ra.rockwell.com

Abstract

Recent trends in AI development, exemplified by innovations like automated machine learning and generative AI, have significantly increased the bottom-up organizational deployment of AI. No- and low-code AI tools empower domain experts to develop AI and thus foster organizational innovation. At the same time, the inherent opaqueness of AI, complemented by the abandonment of requirement to follow rigorous IS development and implementation methods, implies a loss of oversight over the IT for individual domain experts and their organization, and inability to account for the regulatory requirements on AI use. We build on expert knowledge of no- and low-code AI deployment in different types of organizations, and the emerging theorizing on weakly structured systems (WSS) to argue that conventional methods of software engineering and IS deployment can't help organizations harness the risks of innovation-fostering bottom-up developments of ML tools by domain experts. In this research in progress paper we review the inherent risks and limitations of AI - opacity, explainability, bias, and controllability - in the context of ethical and regulatory requirements. We argue that maintaining human oversight is pivotal for the bottom-up ML developments to remain "under control" and suggest directions for future research on how to balance the innovation potential and risk in bottom-up ML development projects.

Keywords: AI, Bottom-up development, domain experts, low-code, no-code, autoML.

Empowering Domain Experts in Developing AI: Challenges of bottom-up ML development platforms

Research-in-Progress Paper

Introduction

Artificial intelligence (AI) has pushed the boundaries of what was once thought possible, experiencing a remarkable surge in adoption across the business landscape (Chui and Malhotra 2018). Machine learning (ML), in particular, is rapidly emerging as the heart of enterprise data science (Burt 2018). Nonetheless, it is crucial to acknowledge that AI and ML stand apart significantly from traditional algorithmic information systems (IS) in how the two breeds of technologies are developed, implemented, and maintained. Traditional IS are built on the foundation of unambiguous algorithmic logic developed by humans analyzing business processes. In contrast AI leverages data to construct ML models, where decision rules often elude human comprehension, rendering AI somewhat opaque to a black box (Castelvecchi 2016). These distinctions have not only fostered skepticism within society but also prompted swift responses from regulators worldwide, leading to the drafting of new AI regulations such as the European Union's (EU) AI Act (European Commission 2021). These regulatory efforts aim to ensure the safety, ethics, and reliability of AI systems as the reliance of society and economy on artificial agents continues to escalate. Considering this expanding dependence, it becomes essential to investigate consequences and risks associated with how AI systems are developed and deployed (Osoba and Welser IV 2017).

Until now, the development of IS has been led by experts. In the case of AI systems development, AI experts are responsible for tasks like data preprocessing, algorithm selection, feature determination, and defining hyperparameters. Consequently, it is reasonable to assume that AI experts possess a comprehensive understanding of the limitations inherent in AI systems and have the capacity to shape "proper" learning of ML from specific datasets. Furthermore, regulatory efforts have primarily centered on these expert-driven, top-down structures to enforce compliance. However, even in the case of expert-driven AI development, the traditional ISD methods can't be applied in their entirety, resulting in gaps and uncertainties when it comes to such tasks as system testing, certification, governance.

The emergence of user-friendly, no- and low-code AI development platforms like AutoML, allows non-IS- or data-professionals to craft AI models with just a few clicks in graphical user interfaces (GUIs) (Xanthopoulos et al. 2020). On the one hand, the emergence of such platforms promises to bring data science to a broader audience of non-AI specialists. On the other hand, the possibility to develop AI models with minimal understanding of and control over the development process (Polzer and Thalmann 2022) opens a pandora box of "AI out of control" (Fomin and Mosakas 2023). The skepticism and risks associated with developing AI systems by professionals are simply exacerbated with the introduction of no- and low-code platforms for ML. Algorithmic IS can always be controlled for bias (errors) by cross-checking the input and output data against a predefined set of instructions (Kalman 1960a). In essence, for system users, the ability to mitigate execution bias becomes an insurmountable challenge. One of the primary reasons behind this inability to exert control lies in the fact that unlike the traditional IS, the algorithmic logic of which is "frozen" by the developers, ML systems continue to "learn" - to improve its execution logic (Lyytinen et al. 2020), sometimes at speed which by far outpaces the human ability to keep track of it.

While bottom-up software development by non-experts has traditionally been supported by no- or low-code tools, it primarily served individual decision support, thus requiring from the organizational management little caution to the risk, ethics or HCI issues. However, recent trends in organizational computing necessitate reconsidering the risks and issues associated with the bottom-up IT development and deployment phenomena: the already visible trends of increased significance and integration of AI within organizational processes, the growing proliferation of no- or low-code tools for AI development, and, as a result, the bypassing of established rigorous and control-centric patterns of interaction between IT experts, the company management, and the domain experts in the development and deployment of (individually engineered) IT solutions. The recent theorizing on weakly structured systems (WSS) (Fomin et al. 2023) suggest the bottom-up initiatives “rewrite” organizational rules on IT use, thus challenging organizational ability and accountability to risk-, ethics-, and oversight-related controls.

Consequently, we assert that the dual shift to 1) different computational paradigm and 2) different organizational approach to IS development, changes the socio-technical dynamics between business processes and IT and introduces challenges for the management of IT, on the one hand, and the regulatory requirements for risk and bias management, on the other hand. To address these challenges, novel theoretical and managerial approaches to bottom-up ML development are required, these recognizing the weakly structured nature (Fomin et al. 2023) of the development and implementation processes, and focused on specialized training for employees involved in the development and deployment of such systems.

Bottom-UP software development based on ML

In classical software engineering for algorithmic systems, domain experts and requirements engineers collaborate to analyze business processes, identify pertinent elements, and recognize recurring patterns that need to be incorporated into the planned IS. In conjunction with system engineers, they formalize their gained knowledge through the creation of models that outline the functionality of the IS. Formal specifications such as ERM, BPMN or UML models then serve as the blueprint for the implementation team, guiding them in the construction of the IS. Additionally, these models constitute the foundational framework for various IT risk management endeavors, including IT audits. This process relies heavily on human involvement and follows well-established principles. Over time, a range of productivity tools, such as code repositories or code sharing platforms, and software engineering methodologies, such as the Waterfall or Scrum approach, have emerged to facilitate and enhance these practices.

Another significant stream of algorithmic system development involves the utilization of no- or low-code tools, designed with the intention of empowering domain experts (Luo et al. 2021). In this approach, domain experts formalize the requirements and functionalities of the envisaged IS by drawing upon their domain-specific expertise. These tools provide standardized interfaces that facilitate the easy expression of ideas, typically through user-friendly graphical interfaces and drag-and-drop functionality (Xanthopoulos et al. 2020). This process fundamentally relies on human knowledge and the capacity of individuals to translate their knowledge into model specifications for the no- or low-code system. The second phase involves automation through tools that undergo rigorous scrutiny to ensure the correct generation of code. Nonetheless, human software experts can readily validate the generated code and perform comprehensive checks and audits as an additional layer of quality assurance.

AI systems rely on the underpinning of ML algorithms, which acquire knowledge from data and autonomously identify patterns within that data (Castelvecchi 2016). Consequently, data stemming from business processes serves as the input for these ML algorithms, enabling the automatic generation of AI models. A frequently acknowledged challenge in this context is the inherent opaqueness of AI, which often results in the development of black-box models (Castelvecchi 2016). However, within the domain of AI research, a multitude of approaches have been developed to keep control upon this process. These strategies encompass data preprocessing to identify and reduce biases, the in-depth interpretation and validation of generated models (e.g., through eXplainable AI or causality approaches) (Gashi et al. 2022), the comprehensive documentation of AI procedures (Königstorfer and Thalmann 2022), and even the establishment of AI audits (Polzer and Thalmann 2022). Consequently, while the entire process spanning from business processes to the construction of the IS is automated, it is possible for human experts to check and control this process. It's crucial to acknowledge, however, that this rigorous approach, characterized by human oversight, demands highly skilled experts, resources, and time. As such, it predominantly finds application in critical use cases where the stakes are high, and the imperative for transparency and accountability is paramount (Königstorfer and Thalmann 2020).

Top Down AI development projects conducted by AI experts have recently been complemented by the growing trend of bottom-up ML development, enabled by tools like AutoML, which leverage no- or low-code interfaces. The primary aim of these approaches is “to democratize data science” and empower domain experts (Hutter et al. 2019). Similar to classical no- or low-code development platforms, these tools aim to reduce the reliance on IT experts. Notably, there has been a recent surge in platforms marketed as “data science as a service,” which have gained popularity. From an organizational standpoint, these bottom-up approaches foster innovations by unearthing hidden knowledge within organizational data, enabling domain experts to harness it for various purposes (Polzer and Thalmann 2022).

The bottom-up, AutoML-based AI development by domain experts requires minimal to no prior ML knowledge. In contrast to algorithmic systems, where functionality can often be validated in advance, domain experts developing ML systems have limited or no visibility into the inner workings of this system (Polzer and Thalmann 2022), and the correctness of the produced ML model heavily depends on the quality and nature of the input data. Consequently, it can be inferred that domain experts with limited AI knowledge face substantial challenges in comprehending the evolving models and assimilating the newly created knowledge. Furthermore, monitoring and controlling this process pose significant challenges as well, including the high risk of introducing bias, given that data used for training were not "proofed" by data scientists. As the application of such systems within business processes generates new data which, in turn, are used to support the continued ML, biases contained within the original data, become reinforced.

Figure 1 shows the three different approaches to system design and the implications for control of the process. Whereas the human has full control of the process in the traditional approach (visualized on the left), the human transfers the implementation of the conceptual model to a no- or low-code system (visualized in the middle) and finally all steps are transferred to the AutoML tool in our third scenario (visualized on the right).

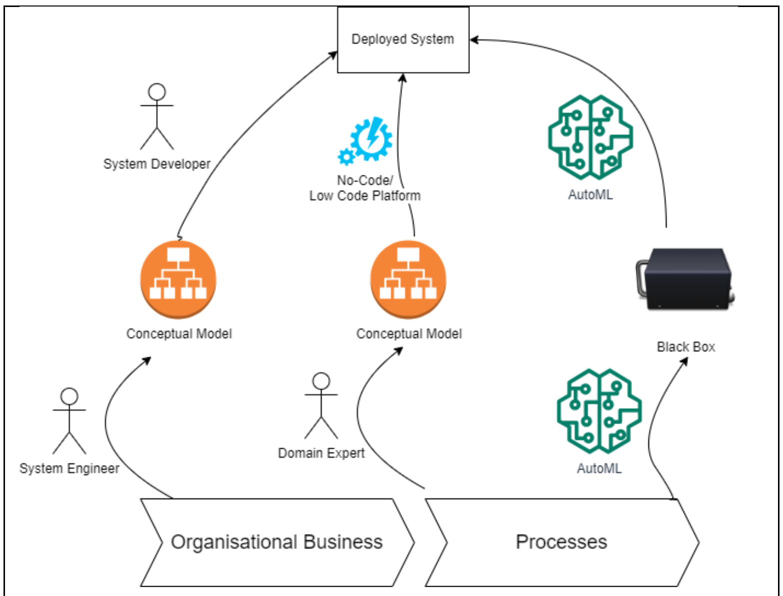

Figure 1. Three different approaches to system design and the implications for control of the process

The requirement of human oversight

Since earliest times, humans have constructed and used technology to make their lives easier, but especially so to be in control in areas, where humans could not be in control (Lyytinen et al. 2020). Technology was always specialized and thus better in specific areas than humans, who are all-rounders (Grunwald, 2019). The narrow application of technology meant that questions of responsibility and control have been clearly distributed. The person or organization using the specific technology is responsible for technology-enabled actions - both prospectively and retrospectively.

With the advent of AI systems, however, narrow specialization of the technique or technology is fading away and systems are being created that are more and more diverse, more and more broadly oriented than towards a very specific purpose and have a potential to exert an influence not only on individual users, but also on organizations and societies at large. Autonomous people can decide independently of the will of others or other influences, they act in a self-determined way (EC AI HLEG 2019a). This value is of central importance for our societies, whether in private or professional contexts, and must be reflected on more intensively, especially in view of AI.

At the same time, from the origins of management theory organizations were theorized as striving for control over individuals. Introduction of informational technologies only increased the possibilities and the drive for control (Zuboff 2015). The elusive boundaries of IT in the age of networked technologies, however, made it increasingly difficult to determine who is responsible for what and when. A call for the inclusion of ethical perspectives in connection with the advancing development of technology was voiced ever louder at each account of unexpected (and often negative) effects of IT on (or misuse of IT by) individuals, firms, and society at large (Coeckelbergh 2020; Kirchschläger 2021).

Based on the central significance of human autonomy (Grimm and Hammele 2019) and according to the fundamental work on trustworthy AI by the HLEG-AI (2019a), the primacy of human action and human oversight is of central importance (Gremsl 2022). If humans were not adequately considered in the design of AI systems or the socio-technical systems within which people use AI, for example in organizations, their autonomy and self-determination could be seriously endangered in the face of technical paradigms. Human supervision has to ensure human autonomy in the socio-technical context and at the same time establish that attributions of responsibility remain possible.

Particularly with the demand for human-centered AI, which is also comprehensively in line with fundamental rights, the question of how this can be ensured in the face of bottom-up AI development is becoming increasingly important. The emergence of no- and low-code AI systems requires rethinking (and adaptation) of previously established ethical and regulatory considerations. For example, the focus on domain experts changes the question of how best to ensure human oversight. Such a need has recently been articulated at the global level particularly frequently by the United Nations.

Existing regulations (e. g. GDPR, AI-Act, which is in discussion (European Commission 2021)) are important factors in the humane use of AI systems, and at the same time imposing a requirement of organizational control over AI developments. In the case of bottom-up AI development, neither the existence of legal regulations, nor the traditional management and ISD methods can ensure human oversight in a comprehensive manner as the implementation of human oversight for each AI development or use case (Kloker et al. 2022). This is particularly evident from the perspective of WSS (Fomin et al. 2023), posing new challenges throughout the entire AI lifecycle. Whereas previous considerations of human oversight in AI development and use were focused on professional developers, when considering a bottom-up development and WSS implementation cases, the focus is shifting towards domain experts.

Challenges for future research

Explainability challenge

AI systems are based on models originating from ML algorithms that can be difficult to interpret and to ascertain the validity of the results. This lack of transparency can lead to a lack of trust in the AI systems and hinder their adoption in critical to human oversight domains such as healthcare or finance (Königstorfer and Thalmann 2020). The established research on technical aspects of eXplainable AI (XAI) needs to be extended with the social dimension, to account for how users can actively engage with the system to obtain explanations, ask questions, and explore alternative scenarios (Saeed and Omlin 2023). Interactive explanations can enhance humans' ability to probe and validate the model's decisions. Furthermore, criteria for evaluating the comprehensibility, fidelity, and usefulness of explanations must also be researched (Kloker et al. 2022). Better user-centric evaluation methods, such as user studies and surveys, to gather feedback on the perceived value and impact of explanations are also needed.

There is often a trade-off between the accuracy and explainability of AI systems (Crook et al. 2023). More complex models may achieve higher accuracy but at the cost of reduced explainability. Research on how to find the right balance between accuracy and explainability is a challenge. Especially as this balance depends on user preferences and the use case itself. Moreover, AI is often deployed at scale, processing large amounts of data and making numerous predictions or decisions (Soklaski et al. 2022). Providing explanations for each individual output in a scalable and efficient manner can be a challenge, especially when dealing with real-time applications or high-throughput systems.

Decision-makers themselves can introduce biases while interpreting the results produced by the AI system (Rastogi et al. 2022). While designing explanations that are clear, concise, and tailored to the target audience is desirable, determining what is clear and concise for the community can be a challenge and may necessitate educating that audience.

These challenges escalate significantly when AI systems are developed on no- or low-code platforms by domain experts within organizations (Schneider et al. 2023). Over time, these systems can begin to proliferate uncontrollably, producing results that may even be contradictory and difficult to validate by those other than the ones who developed them. Furthermore, AI models will evolve with time and input data, potentially starting to produce results that connect information that the user does not have access to or does not understand how to connect. This will mean that human learning will be harder to achieve in organizational environments that use self-evolving AI models (Mosqueira-Rey et al. 2023; Nimmi et al. 2022). On the other hand, ensuring that all locally developed AI systems explain their results can have significant impacts on the speed and accuracy of decision-making in these sociotechnical systems, where human agents interact with AI agents to decide and carry-out tasks.

The challenges listed above point to the need to explore techniques for generating model-agnostic explanations in no- or low-code ML systems. These techniques should allow for generating explanations for different types of AI models, regardless of their complexity or underlying algorithms. Additionally, the explanations should be understandable and meaningful to users with varying levels of technical expertise.

Opacity challenge

Closely related to the concept of explainability, the concept of opacity was defined as a situation when "ML and/or the number of different programmers involved renders an algorithm opaque even to the programmers." (Etzioni and Etzioni 2016, p. 30). In the traditional computational systems (except for the cases when the algorithm is protected by patent) the computational results could have always been checked against the fixed set of instructions – the algorithm. In an ML system this kind of control cannot be established easily or at all. In other words, the system becomes a black box to the observer (Fomin and Astromskis 2023).

The concept of the black box is juxtaposed to the two core principles of transparency of algorithmic systems: (1) the working principles of the system must be known to its creators, and (2) the working principles can be explained to other parties. Neither of the two transparency principles hold for AI systems, irrespective if they were developed according to rigorous ISD methods by professional data scientists or through a bottom-up ML development effort. Following Ferretti and Blasimme (2018), it can be assumed that three types of opacity occur in bottom-up ML systems: (1) *epistemic opacity,* as it is not possible to have access to or there is not sufficient understanding of the rules governing the computational system, (2) *explanatory opacity,* as it is not clear why an artificial intelligence system provides a specific outcome, and (3) *the lack of disclosure,* as data subjects might be unaware that automated decision-making and profiling activities about them are being carried out.

In the case of bottom-up ML development by domain experts, the opacity challenges become exacerbated as AI development is driven by organizational domain experts without a central management oversight. As the work of domain experts, more often than not, requires interlocking of tasks with one another, risks related to epistemic and explanatory opacity increase exponentially. This problem can be also seen as being reinforced by the nature of the AutoML-like platforms, the use of which eliminates the need for the domain expert to explicate the underlying principles of a business even to oneself, thus "pushing" the boundaries of the implicit knowledge deeper into the cognitive and further away from the normative domains, thus reducing human oversight over the principles of business processes and/or ethical considerations. The diminishing scope of formalized knowledge reinforces epistemic and explanatory opacity, leading to the creating of self-reinforcing vicious cycle of "loss of control".

Bias challenge

For an AI system to produce reliable results, it must be first trained on large amounts of (training) data collected by sensors, IT or humans in real world settings. Big data sets from real world processes have a high probability to be messy and to contain biases. Hence, bias in training data will likely result in AI systems not being able to generalize well and possibly making unfair decisions that can favor some groups over others (EC AI HLEG 2019b). Bias in data can occur due to historical, political, or technical factors that were in play when data were collected, causing the data to be insufficiently balanced or inclusive.

Although scholars and data scientists understand the risks of imperfect data well and data preprocessing is the most important task in data science projects, the domain experts may not be aware of the problem and no- and low-code AI development platforms provide currently little support in this regard. Domain experts may not have either a knowledge of the requirements for or the possibility to determine what establishes a sufficiently large and/or bias-free dataset to complete the training process. In this age of explosive growth of data, especially user data collected from various internet sites, smartphone apps, and other sources, data has become a tradable good. As an easy way to obtain data for training AI systems, data acquisition from a third party carries risks of instilling data-related biases. However, even the use of company-owned data does not preclude sample, measurement, prejudice, or exclusion biases. As domain expert is likely to rely on third-party data, especially repurposed data, for the training, it will likely introduce a bias in the training data (Burkhardt et al. 2019), and hence improper learning and biased results are likely.

Whether original or third-party data are used, to date there is no possibility of automatically controlling for bias in training data, as there are no standards or solutions to categorize, label, or assess data according to how they were collected and whether they are accessible, searchable, or findable (Kinder-Kurlanda 2019). This situation, on one hand, emphasizes the importance of human oversight in the process of building training datasets; on the other hand, it emphasizes the risks stemming from ML model development by domain experts. Data used for model development become an oxymoron - data at the same time can be the fuel and the destroying agent for AI systems.

Controllability challenge

Software engineering and IS research has evolved to develop and reply upon rigorous methods for software projects management, including control for risks, quality, resilience, among other (Bassetti et al. 2004). The concept of controllability was originally introduced by Kalman (1960) demanding that to control a system, one must be able to change inputs and must be able to measure the behavior of the system and the outputs (Kalman 1960). In organizational settings, Kalman's (1960) mathematical conception of control over a system is extended to include interaction with the users of the system and manipulations with data and information, which serve as inputs or are produced as the outputs of the system. A general definition of control of IT systems in organizational settings can be found in the widely adopted COBIT framework[1] of the IS Audit and Control Association (ISACA):[2] "The policies, procedures, practices and organizational structures designed to provide reasonable assurance that business objectives will be achieved and that undesired events will be prevented or detected and corrected." (Heschl 2016).

The interlocking nature of business processes and expert knowledge was also reflected in the socio-technical design of traditional IS. With ML learning from data originating from business processes those boundaries are blurred as it is challenging for domain experts to interpret the arising black box models and to finally learn from them. As the traditional forms of management of an enterprise and technological control are similarly predicated upon the premises of functional simplification and closure, the complex interlocked nature of knowledge, processes, and technologies makes the control function problematic (Bassetti et al. 2004, p. 462; Ciborra 2006). Further, as knowledge and risk are closely related, representing the opposites (Bassetti et al. 2004, p. 463), when moving from the traditional centralized IT development and implementation approach towards a bottom-up development based on the efforts of domain experts, the knowledge of the state of affairs cannot be anymore contained in a centralized manner, hence the organization losing its regulatory grip over the IT use (Fomin et al. 2023).

With the rise of bottom-up ML development platforms, not only control over IT development and deployment becomes problematic at organizational level. As much as technology innovation creates new opportunities for the organization, in the informational age (Zuboff 1988) it creates risks in the society: the well-known phenomena of partisan development is the difficulty in finding the answer to the question "who

[1] COBIT (Control Objectives for Information and Related Technology) is a framework created by the ISACA for IT governance and management. It was designed to be a supportive tool for managers and allows bridging the crucial gap between technical issues, business risks, and control requirements.

[2] https://www.isaca.org

is responsible?" (Bassetti et al. 2004). The democratization of AI development process as promised by AutoML "blurs the boundaries between the processes and activities that can be formally represented and the realm of ignorance" (Ciborra 2006), thus reducing the control-ability of the IT projects and increasing the enterprise- and societal risks. "The more sophisticated, integrated and standardized the technological platforms become, the more they tend to behave autonomously and drift" (Ciborra 2006).

Discussion

The recent theorizing on WSS implementation (Fomin et al. 2023) argues that IT systems void of embedded support for organizational workflows, when introduced in organizations without prior definition of rules on "what" and "how" with regard to the use of IT, will require "joint-regulation" processes - the bottom-up development or deployment must be complemented by top-down regulatory initiatives to establish a "regulatory belt" for the technology use. In this research in progress paper we suggest that previous research on WSS must be extended to theorize bottom-up AI development and deployment. Specifically, an important avenue for future research is in recognizing and accounting for multiplication of the development and deployment efforts and the required organizational efforts for retaining control and human oversight when centrally led IT development and deployment is replaced by bottom-up initiatives on the organizational scale. Further, also the usage of no- and low-code AI development platforms in non-CS research (here researchers are domain experts) poses many challenges and in particular those of reproducibility (Koenigstorfer et al. 2024).

In this research in progress paper, we attempted to stay within the sociotechnical conceptualization of control-ability in the IT development, deployment, and use. With interlocking of business processes, the move towards cloud- and grid-technologies and data, and elusive boundaries of effects exerted by IT, the notions of "control of technology" and "human oversight over technology" depart from their original risk-management and decision-making connotations to encompass much broader (and more elusive) contexts such as ethics, fundamental human rights, organizational ability to conform to regulatory requirements – all too important to be ignored and too broad to be addressed in a single research paper.

References

Bassetti, P., Ciborra, C., Jacucci, E., and Kallinikos, J. 2004. "ICT Innovation: From Control to Risk and Responsibility," in *IFIP Working Conference on IT Innovation for Adaptability and Competitiveness*, Springer, pp. 461–464.

Burkhardt, R., Hohn, N., and Wigley, C. 2019. "Leading Your Organization to Responsible AI," *McKinsey Analytics*. (https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Leading%20your%20organization%20to%20responsible%20AI/Leading-your-organization-to-responsible-AI.ashx).

Burt, A. 2018. "How Will the GDPR Impact Machine Learning?," *O'Reilly*, , May 16. (https://www.oreilly.com/ideas/how-will-the-gdpr-impact-machine-learning).

Castelvecchi, D. 2016. "Can We Open the Black Box of AI?," *Nature News* (538:7623), p. 20.

Chui, M., and Malhotra, S. 2018. "Notes from the AI Frontier: AI Adoption Advances, but Foundational Barriers Remain," *McKinsey Analytics*. (https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain).

Ciborra, C. 2006. "Imbrication of Representations: Risk and Digital Technologies," *Journal of Management Studies* (43:6), pp. 1339–1356. (https://doi.org/10.1111/j.1467-6486.2006.00647.x).

Coeckelbergh, M. 2020. *AI Ethics*, The MIT Press Essential Knowledge Series, Cambridge, MA: The MIT Press.

Crook, B., Schlüter, M., and Speith, T. 2023. *Revisiting the Performance-Explainability Trade-Off in Explainable Artificial Intelligence (XAI)*, arXiv. (http://arxiv.org/abs/2307.14239).

EC AI HLEG. 2019a. "Ethics Guidelines for Trustworthy AI," Brussels, Belgium: European Commission: High-Level Expert Group on Artificial Intelligence (AI HLEG), April 8. (https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf).

EC AI HLEG. 2019b. "A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines," Brussels, Belgium: European Commission: High-Level Expert Group on Artificial Intelligence (AI HLEG), April 8. (https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341).

Etzioni, A., and Etzioni, O. 2016. "Designing AI Systems That Obey Our Laws and Values," *Commun. ACM* (59:9), pp. 29–31. (https://doi.org/10.1145/2955091).

European Commission. 2021. "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)," *COM(2021) 206 Final*. (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206).

Ferretti, A. S., and Blasimme, A. 2018. "Machine Learning in Medicine: Opening the New Data Protection Black Box," *European Data Protection Law Review (EDPL)* (4:3), pp. 320–332. (https://doi.org/10.21552/edpl/2018/3/10).

Fomin, V. V., and Astromskis, P. 2023. "The Black Box Problem," in *Future Law, Ethics, and Smart Technologies*, J.-S. Gordon (ed.), Brill, pp. 112–125. (https://doi.org/10.13140/RG.2.2.31841.99687).

Fomin, V. V., Lyytinen, K., Haefliger, S., and de Vaujany, F.-X. 2023. *Implementation of Weakly Structured Systems: Moving from Local Practices to Common Organizational Rules*, presented at the The International Conference on Information Systems (ICIS), , December 8. (https://doi.org/10.13140/RG.2.2.33369.49766).

Fomin, V. V., and Mosakas, K. 2023. "Singularity and Control," in *Future Law, Ethics, and Smart Technologies*, J.-S. Gordon (ed.), Brill, pp. 126–137. (https://doi.org/10.13140/RG.2.2.27856.94726).

Gashi, M., Vuković, M., Jekic, N., Thalmann, S., Holzinger, A., Jean-Quartier, C., and Jeanquartier, F. 2022. "State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification," *BioMedInformatics* (2:1), MDPI, pp. 139–158.

Gremsl, T. 2022. *Ethik 4.0? der "Faktor Mensch" in der digitalen Transformation*, Gesellschaft - Ethik - Religion, Paderborn: Brill Schöningh.

Grimm, P., and Hammele, N. 2019. "Künstliche Intelligenz: Was Bedeutet Sie Für Die Autonomie Des Menschen?," in *Digitale Ethik: Leben in Vernetzten Welten*, pp. 153–170.

Heschl, J. 2016. "Internal Control - Key to Delivering Stakeholder Value," *ISACA NOW BLOG*, , March 10. (https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2016/internal-control-key-to-delivering-stakeholder-value).

Hutter, F., Kotthoff, L., and Vanschoren, J. 2019. *Automated Machine Learning: Methods, Systems, Challenges*, Springer Nature.

Kalman, R. E. 1960a. "Contributions to the Theory of Optimal Control," *Bol. Soc. Mat. Mexicana* (5:2), pp. 102–119.

Kalman, R. E. 1960b. "On the General Theory of Control Systems," in *Proceedings First International Conference on Automatic Control, Moscow, USSR*, pp. 481–492.

Kinder-Kurlanda, K. 2019. *Keynote Speech at the 18th Annual STS Conference Graz*, , May 6. (https://sts-conference.isds.tugraz.at/event/2/attachments/1/22/BoA2019_-_STS_Conference_Graz.pdf).

Kirchschläger, P. G. 2021. *Digital Transformation and Ethics: Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence*, (1st edition.), Edition Sigma, Baden-Baden: Nomos.

Kloker, A., Fleiß, J., Koeth, C., Kloiber, T., Ratheiser, P., and Thalmann, S. 2022. "Caution or Trust in AI? How to Design XAI in Sensitive Use Cases?," in *AMCIS 2022 Proceedings*, p. 16. (https://aisel.aisnet.org/amcis2022/sig_dsa/sig_dsa/16).

Königstorfer, F., and Thalmann, S. 2020. "Applications of Artificial Intelligence in Commercial Banks–A Research Agenda for Behavioral Finance," *Journal of Behavioral and Experimental Finance* (27), Elsevier, p. 100352.

Königstorfer, F., and Thalmann, S. 2022. "AI Documentation: A Path to Accountability," *Journal of Responsible Technology* (11), Elsevier, p. 100043.

Koenigstorfer, F.; Haberl, A.; Kowald, D.; Ross-Hellauer, T., Thalmann, S. (2024) Open Science Practices in Information Systems Research, ai documentation, open science, reproducibility. Proceedings of the 57th Hawaii International Conference on System Sciences, 682

Luo, Y., Liang, P., Wang, C., Shahin, M., and Zhan, J. 2021. "Characteristics and Challenges of Low-Code Development: The Practitioners' Perspective," in *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pp. 1–11.

Lyytinen, K., Nickerson, J. V., and King, J. L. 2020. "Metahuman Systems = Humans + Machines That Learn," *Journal of Information Technology* (36:4), SAGE Publications Ltd, pp. 427–445. (https://doi.org/10.1177/0268396220915917).

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., and Fernández-Leal, Á. 2023. "Human-in-the-Loop Machine Learning: A State of the Art," *Artificial Intelligence Review* (56:4), pp. 3005–3054. (https://doi.org/10.1007/s10462-022-10246-w).

Nimmi, P. M., Vilone, G., and Jagathyraj, V. P. 2022. "Impact of AI Technologies on Organisational Learning: Proposing an Organisation Cognition Schema," *Development and Learning in Organizations: An International Journal* (36:5), pp. 7–9. (https://doi.org/10.1108/DLO-08-2021-0148).

Osoba, O. A., and Welser IV, W. 2017. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, Rand Corporation.

Polzer, A. K., and Thalmann, S. 2022. "The Impact of AutoML on the AI Development Process," in *Proceedings of the 2022 Pre-ICIS SIGDSA Symposium*. (https://aisel.aisnet.org/sigdsa2022/13).

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., and Tomsett, R. 2022. "Deciding Fast and Slow: The Role of Cognitive Biases in AI-Assisted Decision-Making," *Proceedings of the ACM on Human-Computer Interaction* (6:CSCW1), pp. 1–22. (https://doi.org/10.1145/3512930).

Saeed, W., and Omlin, C. 2023. "Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities," *Knowledge-Based Systems* (263), p. 110273. (https://doi.org/10.1016/j.knosys.2023.110273).

Schneider, J., Abraham, R., Meske, C., and Vom Brocke, J. 2023. "Artificial Intelligence Governance For Businesses," *Information Systems Management* (40:3), pp. 229–249. (https://doi.org/10.1080/10580530.2022.2085825).

Soklaski, R., Goodwin, J., Brown, O., Yee, M., and Matterer, J. 2022. *Tools and Practices for Responsible AI Engineering*, arXiv. (http://arxiv.org/abs/2201.05647).

Xanthopoulos, I., Tsamardinos, I., Christophides, V., Simon, E., and Salinger, A. 2020. "Putting the Human Back in the AutoML Loop.," in *EDBT/ICDT Workshops*.

Zuboff, S. 1985. "Automate/Informate: The Two Faces of Intelligent Technology," *Organizational Dynamics* (14:2), pp. 5–18. (https://doi.org/10.1016/0090-2616(85)90033-6).

Zuboff, S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*, New York, NY, USA: Basic Books, Inc.

Zuboff, S. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology* (30:1), pp. 75–89. (https://doi.org/10.1057/jit.2015.5).