

One AIaaS Does not Fit All:

Challenges concerning Service Configuration of cloud-based Artificial Intelligence Services

Background

Artificial Intelligence (AI) is undoubtedly one of the most actively debated technologies, providing auspicious opportunities to contribute to individuals' well-being, the success and innovativeness of organizations, and societies' prosperity and advancement. However, one major challenge for organizations is the complex and demanding process of adopting and integrating AI, which is rather considered a journey and not a destination. Most organizations still fail to adopt AI and harness its full potential due to the scarcity of AI experts; a lack of organizations' abilities and budgets to set up and maintain the extensive IT resources needed; and limited knowledge on how to deploy and configure the AI-based systems effectively, among others.

To foster AI diffusion and application, cloud providers such as Amazon, Google, IBM, Microsoft, Salesforce, or SAP have started to offer machine learning, deep learning, analytics, and inference as a service, bringing the discussions about provisioning AI capabilities from the cloud into practice. Also, start-ups and small and medium-sized enterprises (SME) are following the trend and providing unique cloud-based AI services tailored to SMEs' needs in various industries. These services became known as *Artificial Intelligence as a Service* (AIaaS), meaning cloud-based systems providing on-demand services to organizations and individuals to deploy, develop, train, and manage AI models.

While the emergence of AIaaS offers manifold opportunities (e.g., complexity abstraction, automation, cost savings etc.), AIaaS on the one hand inherits a variety of issues and challenges relating to AI and cloud computing in general. On the other hand, AIaaS not only intensifies inherited issues (e.g., AIaaS being perceived as a black-box) but also leads to various novel socio-technical challenges and issues that may severely impede its value contributions if not handled appropriately. One very important but yet insufficiently researched issue concerns trade-offs when provisioning AIaaS, that are, the improvement of one AIaaS characteristic interferes with another characteristic. For example, AIaaS providers enter a trade-off between accuracy and fairness vs. generalizability of pre-trained models offered in the cloud. As the most decisive factor for general model applicability is the model's underlying dataset, providers aim to conceptualize a set of generic training data. In practice, there is no one-model-fits-all solution, thus, there is no single model or algorithm that can handle all dataset varieties, and therefore, using a dataset with general data potentially leads to low prediction accuracy and discrimination. An illustrative example of the generalization challenge and resulting discrimination is an AIaaS offering a pre-trained model to predict a person's weight, which receives characteristics such as age, gender, and height as input. The problem is that people from North America may have a significantly different distribution of body weight to the mentioned characteristics than, for instance, people in Asia, and hence the training data could be unrepresentative, which will not only adversely affect the models' ability to handle unseen test data but may also lead to biases. Related trade-offs exist, such as an easy to use AIaaS can not achieve a high degree of customizability.

To date, we lack knowledge about these trade-offs and how they impact service configuration and service usage. The goal of this research project is to reveal prevalent trade-offs when provisioning AIaaS and identify respective service configuration archetypes (e.g., an AIaaS archetype improves generalizability but lacks accuracy). The research project takes a socio-technical perspective and is not purely technical but rather considers both organizational and technical trade-offs. Results of the research project should be jointly published as a research article, and more importantly, also summarized and synthesized into a compelling teaching case for students to bring AIaaS into upcoming lectures.

The ideal candidate has basic knowledge in AI and cloud services. Knowledge on AIaaS and a solid understanding of AI technologies is nice to have, but not necessary. Students should have high degrees of self-efficacy to manage this research project and openness to discuss trade-off challenges with AIaaS providers and users. The work allows you to gain deep knowledge and experience in rapidly growing fields, such as cloud services, AI particularly machine learning, and service configuration and usage, and also offers you a unique gateway to learn and apply scientific research methods.

Possible prospective tasks of the intern

- Briefly review related research papers in the field of AIaaS
- Perform interviews with local AIaaS providers and users
- Derive and discuss trade-offs of AIaaS provisioning
- May come up with service configuration archetypes based on the trade-offs
- Summarize the results in a compelling teaching case and research articles

General information about the workgroup, the university, and the region

- Robert W. Gregory is Associate Professor of Business Technology at University of Miami Herbert Business School. He holds a diplom, a combined bachelor's and master's degree, in Management Information Systems from the University of Cologne, Germany, a master's degree in International Management from the Community of European Management Schools (CEMS), and a Ph.D. equivalent, Dr. rer. pol., in Business Administration from Goethe University Frankfurt, Germany. Prof. Gregory's research program focuses on novel management and information systems phenomena related to the diffusion and innovation with digital technologies and the associated transformation of individuals, organizations, and markets. His research has appeared in premier scientific journals, including MIS Quarterly, Information Systems Research, and Academy of Management Review. His teaching covers digital innovation and disruption, digital transformation, and product and project management across undergraduate, master's, MBA, and executive levels and spans multiple countries and cultures. He has worked and consulted with leading companies such as Deutsche Bank, DBS Bank, T-Systems, Volkswagen AG, Microsoft, Ricoh, Xerox, and more.
- The University of Miami is a private research university with more than 10,000+ full-time, degree-seeking undergraduates and 6,000 full-time, degree-seeking graduates from around the world. Here, students study what they love, collaborate across disciplines, and tackle real-world challenges. Our vibrant community is made up of students and scholars from all over the world that challenge each other to push beyond their limits and achieve more.

Introductory literature

- Lins, S., Pandl, K.D., Teigeler, H. et al. Artificial Intelligence as a Service. *Bus Inf Syst Eng* 63, 441–456 (2021). <https://doi.org/10.1007/s12599-021-00708-w>
- Pandl, K. D., Teigeler, H., Lins, S., Thiebes, S., Sunyaev, A. Drivers and inhibitors for organizations' intention to adopt artificial intelligence as a service. *Proceedings of the 54th Annual Hawaii International Conference on System Sciences, HICSS 2021, 4-8 January 2021*
- Geske, F., Hofmann, P., Lämmermann, L., Schlatt, V., Urbach, N. Gateways to Artificial Intelligence: Developing a Taxonomy for AI Service Platforms. *Proceedings of the 29th European Conference on Information Systems, ECIS 2021*.
- Thiebes, S., Lins, S. & Sunyaev, A. Trustworthy artificial intelligence. *Electron Markets* 31, 447–464 (2021). <https://doi.org/10.1007/s12525-020-00441-4>