

Bridging Competency Gaps in Data Science: Evaluating the Role of Automation Frameworks Across the DASC-PM Lifecycle

Maike Holtkemper and Christian Beecks

University of Hagen, Chair of Data Science, Universitätsstrasse 11, 58097 Hagen, Germany

Keywords: Automation Frameworks, EDISON Data Science Framework, Data Science Process Models, DASC-PM.

Abstract: Successful data science projects require a balanced mix of competencies. However, a shortage of skilled professionals disrupts this balance, fragmenting expertise across the data science pipeline. This fragmentation causes inefficiencies, delays, and project failures. Automation frameworks can help to mitigate these issues by handling repetitive tasks and integrating specialized skills. These frameworks improve workflow efficiency across project phases but remain limited in critical areas like project initiation and deployment. This pre-study identifies tasks in each project phase using the DASC-PM model. The model structures the assessment of automation potential and maps tasks to the EDISON Data Science Framework (EDSF), determining which competencies automation can support. The findings indicate that automation enhances efficiency in early phases, such as Data Provision and Analysis, contrasting with challenges in Project Order and Deployment, where human expertise remains essential. Addressing these gaps can improve collaboration and create a more integrated data science workflow.

1 INTRODUCTION

In today's digital era, data drives decision-making across sectors as organizations analyze large datasets to gain insights, shape actions, and stay competitive (Robinson and Nolis, 2020). These insights remain difficult to realize, as many organizations still struggle to succeed in data science projects (Gökay et al., 2023). These projects require well-structured execution to reduce risks and improve outcomes (Haertel et al., 2022; Kutziyas et al., 2021).

However, the success of data science endeavors relies on both structured models and team competencies (Santana and Díaz-Fernández, 2023). These competencies, including e.g. data engineering, are essential for supporting the entire project lifecycle (Cuadrado-Gallego and Demchenko, 2020). However, growing competency gaps, driven by demand that outpaces the supply of skilled professionals, limit the transformation of raw data into insights and undermine project success (Mikalef and Krogstie, 2019).

Automation frameworks have been shown to automate repetitive tasks and reduce fragmentation across project phases (Wang et al., 2021; Macas et al., 2017). By reducing fragmentation across phases, these frameworks help bridge silos and create a collaborative environment throughout the data science lifecycle (Abdelaal et al., 2023). Tools such as AutoDS support data provision and analysis while link-

ing stages for smoother team transitions (Wang et al., 2021). These tools alleviate competency shortages by automating expertise-driven tasks, which in turn optimize workflows, reduce errors, and accelerate project completion (Abdelaal et al., 2023; Macas et al., 2017). These optimized workflows are enabled by frameworks that align with DASC-PM stages and show potential for integrated automation.

This pre-study investigates how automation frameworks support essential data science competencies, addressing the research question: *How do automation frameworks support the competencies needed for data science projects?* Using the EDSF (Demchenko et al., 2022), it examines how these tools complement or replace human expertise in data preparation, modeling, and evaluation. These competencies are mapped to DASC-PM tasks to identify gaps and ensure a coherent project approach. This mapping forms the main contribution by combining DASC-PM and EDSF to assess competency coverage, highlight unmet needs, and inform future tool development. As a pre-study, the paper outlines key challenges, proposes a structured approach, and lays the foundation for future research. The structure is as follows: Section 2 reviews the background, Section 3 details the methodology, Section 4 presents the findings, and Section 5 concludes with future directions.

2 BACKGROUND

Data’s role is expanding as organizations invest in AI and data projects to drive revenue, efficiency, and innovation (Santana and Díaz-Fernández, 2023). While terms like AI, data science, and machine learning differ, their core objectives remain consistent across industries (Kruhse-Lehtonen and Hofmann, 2020).

Many data science projects fail to meet expectations, with most never reaching production. VentureBeat (VentureBeat, 2019) reports an 87% failure rate, while NewVantage Partners (NewVantage Partners, 2019) finds 77% of companies struggle with AI adoption. Additionally, 70% report minimal AI impact (Ransbotham et al., 2019), and data scientists face challenges integrating models into operations. Davenport et al. (Davenport et al., 2020) note a widening gap between successful and failing organizations. Poor project management and technical hurdles further drive high failure rates (Joshi et al., 2021).

Research highlights a disconnect between technical processes and organizational practices in data science projects, increasing risks such as poor project management, competency gaps, and data quality issues (Saltz, 2021; Martinez et al., 2021). Boina (Boina et al., 2023) discusses the integration of Data Engineering and Intelligent Process Automation (IPA) to enhance business efficiency and innovation. Reddy et al. (Reddy et al., 2024) identify a competency gap as a key barrier to effective data science adoption and alignment with organizational goals. Li et al. (Li et al., 2021) identify critical skills and domain knowledge gaps in U.S. manufacturing by analyzing job postings and professional profiles, highlighting the need for targeted workforce training. Aljohani et al. (Aljohani et al., 2022) reveal a persistent mismatch between university curricula and job market demands through a large-scale analysis.

To address these complexities, it is essential to clearly define the competencies required. Gartner (James and Duncan, 2023) predicts that by 2026, leading data science teams will need increasingly diverse skill sets, resulting in significant changes to team structures. In response, competence frameworks have become increasingly important for defining and cultivating the skills needed in data science initiatives (Salminen et al., 2024; Brauner et al., 2025).

2.1 The EDISON Data Science Framework (EDSF)

The EDSF, developed during the EDISON project (2015–2017), defines key competencies for data scientists (Cuadrado-Gallego and Demchenko, 2020). It

provides a structured curriculum and knowledge base to support skill development (European Commission, 2017). A core component, the Competence Framework for Data Science (CF-DS), links essential competencies to relevant knowledge and skills, ensuring a standardized training model across Europe (European Commission, 2017).

EDSF aligns with the European e-Competence Framework (European Committee for Standardization, 2014), defining competence as the ability to apply knowledge, skills, and attitudes to achieve results (European Committee for Standardization, 2014, p.5). CF-DS categorizes competencies into five areas: Data Analytics (DSDA), Data Engineering (DSENG), Data Management (DSDM), Research Methods and Project Management (DSRMP), and Domain-Specific Knowledge (DSDK). Missing competencies can hinder project success, as expertise is often distributed across teams. Automation frameworks can help bridge these gaps, enhancing efficiency and outcomes (Potanin et al., 2024). The full EDISON CF-DS framework is available [here](#).

Table 1: Excerpt of the EDISON CF-DS.

Cat.	Sub-Category	Sub-Sub-Category
Data Analytics - DSDA	DSDA01: Use a variety of data analytics techniques	Machine Learning, Data mining, Prescriptive analytics, Predictive analytics, Data life cycle
	DSDA02: Apply designated quantitative techniques	Statistics, Time series analysis, Optimization, Simulation, Deploy models for analysis and prediction
	DSDA03: Identify, extract, combine available heterogeneous data	Modern data sources (audio, video, image, ...), Verify data quality

2.2 Leveraging Automation to Address Competency Gaps

The demand for skilled data scientists continues to rise, yet a shortage of qualified professionals persists (Demchenko and José, 2021). Both academia and industry seek innovative solutions to address this competency gap. As data volumes grow, developing machine learning models and extracting insights becomes more complex, increasing the manual effort required for data processing and analysis (Elshawi et al.,

2019).

Automation frameworks help by handling tasks traditionally requiring specialized expertise. Research shows that data processing (69%) and collection (64%) are well-suited for automation (Manyika et al., 2017). These tools reduce repetitive work, accelerate decision-making, and enhance efficiency (Abbaszadegan and Grau, 2015). AutoDS (Wang et al., 2021) and AutoCure (Abdelaal et al., 2023) support various stages of the data science process.

Automation also improves software testing, with frameworks like Robot reducing execution time by over 80% (Alok Chakravarthy and Padma, 2023). While automation saves time and reduces errors, its complexity varies across tasks. Some tasks can be fully automated, while others require human oversight. As Automated Data Science (AutoDS) evolves, Human-Computer Interaction (HCI) research highlights a shift in perception—data scientists now see automation as a collaborator rather than a competitor (Wang et al., 2021; Wang et al., 2019).

2.3 Data Science Process Model (DASC-PM)

A survey identifying meta-requirements led a group of data science experts from academia and industry to develop the DASC-PM (Schulz et al., 2022). This model structures data science projects into a five-stage process, integrating scientific practices, application domains, IT infrastructures, and their impacts (Schulz et al., 2022). The five phases—Project Order, Data Provision, Analysis, Deployment, and Application—operate within three overarching areas: Domain, Scientificity, and IT Infrastructure.

The Project Order phase defines domain-specific problems and selects use cases, requiring diverse competencies. Data Provision covers data acquisition, storage, and management for analysis. In Analysis, the team applies or develops methodologies, ensuring validation. Deployment implements analytical results, while Application monitors model usage and gathers insights for improvements.

Domain expertise guides objective setting, data interpretation, and ethical considerations. Scientificity ensures methodological rigor and structured management. IT infrastructure supports all phases, assessed for needs and scalability. DASC-PM addresses gaps in existing models, providing a structured, evolving framework (Schulz et al., 2022) (see Table 2).

This study examines how automation frameworks support competencies essential for data science projects. As tasks grow more complex and skilled professionals remain scarce, automation helps bridge

skill gaps. Using the CF-DS, the study maps DASC-PM tasks to assess automation's role in complementing human expertise and identifying competency gaps that impact project success.

3 METHODOLOGY

3.1 Literature Review

In line with Webster and Watson's approach (Webster and Watson, 2002), a literature review was conducted to identify relevant studies on automation frameworks in data science. As this is a pre-study rather than a complete research work, no forward or backward search was performed; instead, a keyword search was used to gain initial insights. A search for "automation framework" and "data science" was conducted in ACM Digital Library and IEEE Xplore, focusing on publications from the last five years. Through this keyword search, 127 automation frameworks were found, whereas 38 remained after the abstract evaluation and 22 after the full-text evaluation. Topics like simulation platforms (Aryai et al., 2023) and educational frameworks such as AutoDomainMine (Varde, 2022) were noted but not explored in depth, as the study focuses on automation frameworks for data science projects.

3.2 Quality Appraisal and Qualitative Content Analysis

After the initial literature review, a quality assessment was conducted on the selected 22 articles to ensure their reliability and relevance. The evaluation methodology follows Kitchenham's guidelines (Kitchenham, 2004). Each article was assessed using predefined criteria, classifying them as low, medium, or high quality based on the standards outlined by Nidhra et al. (Nidhra et al., 2013).

The quality assessment criteria included the four questions:

- Does the research align with the objectives of this study? (general alignment)
- Is the study focused on an automation framework? (automation framework)
- Does the automation framework address tasks within the DASC-PM process model? (task-related)
- Are the findings relevant to the aims of this study? (usefulness of the results)

Table 2: Automation Frameworks covering DASC-PM Phases.

ID	AutoFM	Reference	Year	Goal	Phase
1	DQA	(Shrivastava et al., 2019)	2019	Data quality	2
2	Sweeper	(Thawanthaleunglit and Sripanidkulchai, 2019)	2019	Data quality	2-4
3	AutoDS	(Wang et al., 2021)	2021	ML config.	2-5
4	VizSmith	(Bavishi et al., 2021)	2021	Visualization	1-2,4
5	GrumPy	(Mota, JR., Joselito et al., 2021)	2021	Data analysis	2
6	DVF	(Lwakatare et al., 2021)	2021	Data validation	2
7	OneBM	(Lam et al., 2021)	2021	Feature Engin.	2-4
8	AutoPrep	(Bilal et al., 2022)	2022	Data processing	2
9	QuickViz	(Pitroda, 2022)	2022	EDA	2
10	ADE	(Galhotra and Khurana, 2022)	2022	Data labeling	2-3
11	NLP	(Mavrogiorgos et al., 2022)	2022	Data quality	2
12 a	Datadiff	(Petricek et al., 2023)	2023	Merging tables	2
12 b	CleverCSV	(Petricek et al., 2023)	2023	Parsing tables	2
12 c	Ptype	(Petricek et al., 2023)	2023	Column types	2
12 d	ColNet	(Petricek et al., 2023)	2023	Annotating data	3
13	AutoCure	(Abdelaal et al., 2023)	2023	Data quality	2-3
14	AI	(Patel et al., 2023)	2023	EDA	1-2,5

Each article was scored based on its compliance with predefined criteria, with weights assigned accordingly. Articles scoring with 4 points were rated as high quality, those below 1 as low quality, and those between 1 and 3 as medium quality. Out of the 22 articles reviewed, 14 were rated as high quality, and 8 were classified as medium quality.

A qualitative content analysis was conducted on the 14 high-quality papers, identifying 17 automation frameworks for systematic analysis. The analysis followed Kuckartz's methodology (Kuckartz and Rädiker, 2022), focusing on how these frameworks align with the DASC-PM models. A deductive coding approach categorized text passages into the six phases and tasks of CRISP-DM (Table 2) and the five phases of DASC-PM. Two researchers independently coded the documents, following the guidelines in (Kuckartz and Rädiker, 2022). MAXQDA2022 (Version 22.8.0) was used for the analysis.

The EDISON CF-DS was applied to evaluate the competencies required for using automation frameworks. This framework categorizes competencies into five key areas: Data Analytics (DSDA), Data Engineering (DSENG), Data Management (DSDM), Research Methods and Project Management (DSRM), and Domain Knowledge (DSDK). The EDISON CF-DS was used in the second coding phase.

4 FINDINGS

4.1 Analysis of Automation Frameworks

The 17 frameworks found in the literature were aligned with the various phases and tasks of the DASC-PM process model (see Table 3). The analysis revealed that only two frameworks covered partially the Project Order phase. However, 16 frameworks were relevant to Data Provision, 10 to Analysis, 0 to Deployment, and 2 to Application.

The Project Order phase, which includes tasks such as assessing the suitability of the use case, methods, and objectives, is minimally supported by automation tools. For example, tools like VizSmith and AutoDS cover tasks like evaluating the suitability of methods and objectives, but other tasks, such as assessing the data basis, remain largely unsupported.

The Data Provision phase sees more extensive support from automation tools. Tasks such as data anonymization, aggregation, cleansing, and filtering are well-covered by tools like AutoDS, VizSmith, and QuickViz, which automate data preparation. However, tasks related to data protection and metadata management are not widely covered by the tools.

In the Analysis phase, tools like AutoDS, Sweeper, and VizSmith provide significant coverage for tasks such as identifying suitable analytical methods, selecting the best parameter configuration, and evaluating results. These tasks are crucial for ensur-

Table 3: Covered DASC-PM Areas, Tasks, and Subtasks.

Phase	Task	Subtask	DQA	Sweeper	AutoDS	VizSmith	GrumPy	DVF	OneBM	AutoPrep	QuickViz	ADE	NLP	Datadiff	CleverCSV	Ptype	ColNet	AutoCure	AI
Project Order	Sustain. Check	Suitability of the use case	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-
		Suitability of the obj.	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	x
Data Provision	Data Preparation	Data aggregation	-	x	x	x	-	-	x	x	-	x	-	x	-	-	-	-	x
		Data annotation	-	x	x	x	-	-	x	x	-	x	-	x	-	-	-	-	x
		Data cleansing	x	x	x	-	x	x	-	x	-	-	x	x	-	x	-	x	x
		Data filtering	x	x	x	-	-	x	-	x	-	-	x	x	-	-	-	x	-
		Data structuring	x	x	x	-	-	x	-	x	-	-	-	x	-	-	-	x	x
		Data transformation	-	-	x	-	-	-	x	x	-	-	-	-	x	-	-	-	-
		Dimensional reduction	-	x	x	x	-	-	x	x	-	x	-	x	-	-	-	-	x
		Format adjustment	-	-	x	-	-	-	x	x	-	-	-	-	x	-	-	-	-
	Data Mgmt	Data protection	x	x	x	-	x	x	-	x	-	-	x	x	-	x	-	-	x
		Storing raw data	-	-	x	x	x	x	-	-	-	x	x	-	x	-	-	x	-
		Data access	x	x	x	x	x	x	-	x	-	x	x	x	x	-	-	x	x
	Explorative Data Analysis	Data validation	-	x	x	x	x	x	-	-	x	x	-	-	x	-	-	x	x
		Data visualization	x	x	x	x	x	x	-	-	x	-	-	-	-	x	-	-	x
		Ident. central attr.	x	x	x	x	x	x	-	-	x	-	-	-	-	x	-	-	x
		Understanding content	x	x	x	x	x	x	-	-	x	-	-	-	-	x	-	-	x
		Statistical analysis	-	x	x	x	x	x	-	-	x	x	-	-	x	-	-	x	x
		Examining the necessity of data transformations	-	x	x	x	x	x	-	-	x	x	-	-	x	-	-	x	x
		Exam. missing values	-	x	x	x	x	x	-	-	x	x	-	-	x	-	-	x	x
Analysis	Anal. Methods	Determining potentially suitable procedures	-	-	x	-	-	-	x	-	-	x	-	-	-	-	-	x	-
		Selection	-	-	x	-	-	-	x	-	-	x	-	-	-	-	-	x	-
	Appl. Anal. Methods	Setting up a development environment	-	x	x	-	-	-	x	-	-	x	-	-	-	-	-	-	-
		Constructing the progress	-	x	x	-	-	-	x	-	-	x	-	-	-	-	x	-	-
		Reducing dimensions	-	x	x	-	-	-	x	-	-	-	-	-	-	-	x	-	-
	Dev. Anal. Methods	Designing the procedure	-	x	x	-	-	-	x	-	-	-	-	-	-	-	x	-	-
		Benchmarking	-	x	x	x	-	-	x	-	-	-	-	-	-	-	x	-	-
	Evaluation	Comparing procedures	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Evaluating results	-	x	x	x	-	-	x	-	-	-	-	-	-	-	x	-	-
		Performance tests	-	-	x	-	-	-	x	-	-	-	-	-	-	-	x	-	-
Deploy.	Ensuring Applic.	Ensuring constant applicability of the model	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	x
Application	Monitoring	Gathering domain-specific findings for iterative developments	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x

ing the analytical models align with the data and objectives. However, there is a lack of support for more detailed tasks such as testing and establishing the procedure, indicating a focus on operational rather than methodological development and testing.

The Deployment phase, which involves preparing results for recipients, creating technical environments, and ensuring the system's viability, shows substantial tool coverage. Tools like AutoPrep, OneBM, and QuickViz assist in automating the creation of environments and the transfer of results. However, tasks related to technical infrastructure, such as testing software licenses and identifying hardware stacks, are only partially covered by the automation tools.

Finally, the Application phase, which includes monitoring, ensuring constant applicability of the model, and gathering domain-specific findings for iterative improvements, is somewhat well-supported by tools like AutoDS and QuickViz. These tools aid in ongoing model validation and updates, ensuring long-term applicability. However, gaps remain in tasks related to iterative improvement, suggesting the need for more work in domain-specific analysis for continuous adaptation of models.

The tasks that are not covered by any automation framework are highlighted below:

- Project Order:
 - Sustainability Check: Suitability of the methods, Assessing the data base, Considering past projects; Ensuring Realizability: not covered.
- Data Provision:
 - Data Preparation: Data anonymization, integration, Creating data preparation plans, Logging the data preparation, Process automation, Schema integration;
 - Data Management: Backing up prepared data, Metadata management;
- Analysis:
 - Identifying Suitable Analytical Methods: Identifying requirements, Determining the problem class, Researching comparable problems;
 - Applying Analytical Methods: Ensuring validity, Considering multiple analytical methods, Selecting the best parameter configuration, Weighing time against benefit, Ensuring replicability and transparency, Establishing criteria;
 - Developing Analytical Methods: Determining differences with relevant existing procedures, Establishing the procedure, Testing the procedure, Implementation;
 - Tool Selection: not covered;

- Evaluation: Determining the evaluation criteria, Estimating added value, Reviewing realizability;

- Deployment:

- Technical and Methodical Provision: Preparing the results for the recipients, Building the product environment, Transferring the results, Context creation, Automating processes, Dealing with IT resources, Technically testing the system used;
- Ensuring Technical Realizability: Considering time criticalities, Considering durations, Dealing with the connected data sources, Identifying the hardware and software stacks, Identifying technical conditions and opportunities, Testing software licenses, Legal framework conditions, Create memory access concept, Ensure operations and support;
- Ensuring Applicability: Identify target recipients, Establishing UI/UX design, Ensure memory access, Involve users, Create a documentation concept, Create a training concept, Regularly checking the quality of analysis results

The mapping of DASC-PM tasks to EDISON competencies reveals that the DASC-PM framework spans a broad range of competencies, integrating data science analytics, engineering, and business analytics. Table 4 shows the overall tasks per phase, where 1) Project Order, 2) Data Provision, 3) Analysis, 4) Deployment, 5) Application. Data Preparation and Data Management tasks align heavily with DSDA (Data Science Analytics) and DSENG (Data Science Engineering) competencies, focusing on data handling, process automation, and technical implementation. The Analysis phase, in turn, highlights a strong focus on DSDA competencies for analytical methods. Deployment and Application tasks emphasize DSENG for technical realization and DSBA for ensuring business applicability and impact.

Overall, the framework demonstrates a comprehensive approach to data science projects, combining technical, analytical, and business perspectives. The importance of data quality, accessibility, and business relevance throughout the project lifecycle is underscored by the integration of DSDM (Data Management) and DSBA competencies.

5 CONCLUSION

This study analyzed how automation frameworks support tasks across the DASC-PM process model. The analysis revealed substantial variation in coverage.

Table 4: Mapping DASC-PM Tasks to EDISON Competencies.

Phase	Task	EDISON Competencies
1	Sustainability Check	DSDA01,DSDA03,DSDA04,DSDA05,DSDM05
	Ensuring Realizability	DSENG01,DSENG03-05,DSRMP02,DSRMP06,DSDA02,DSDA04-05,DSBA01,DSBA03
2	Data Preparation	DSENG01,DSENG03,DSENG05-06,DSDA01-0,DSDM03-06,DSBA04
	Data Management: Data protection	DSENG04-06,DSDM03,DSDM06,DSDA03
	EDA: Data validation	DSDA01-04,DSDA06,DSBA03
3	Identifying Analytical Methods	DSDA01-03,DSBA03
	Applying Analytical Methods	DSENG01,DSENG03,DSDA01-02,DSDA04-05,DSBA04
	Developing Analytical Methods	DSDA01-03,DSDA05
	Tool Selection	DSENG01-02,DSDA01,DSBA03
	Evaluation	DSDA02-05,DSBA03-04
4	Technical and Methodical Provision	DSDA04-6,DSBA03,DSENG01-04
	Ensuring Technical Realizability	DSDA01,DSDA03-04,DSBA04,DSENG01-02, DSENG04-06
	Ensuring Applicability: Identify target recipients	DSBA01-03,DSENG06, DSDA01,DSDA06,DSBA05-06
5	Monitoring	DSDA02,DSDA04,DSDA06,DSBA04

Tools like AutoDS, VizSmith, and QuickViz offer strong support for tasks in the Data Provision phase. These tasks receive far less support in the Project Order and Application phases. Key responsibilities such as risk assessment, infrastructure testing, and post-deployment monitoring remain largely unsupported. This lack of coverage became evident through a mapping of DASC-PM tasks against data science competencies using the EDISON framework. The mapping highlighted where automation can help mitigate existing competency gaps.

These gaps inform several actionable recommendations. The recommendations include conducting empirical studies to explore practical and organizational barriers to automation adoption. These studies should also examine context-specific constraints that affect tool effectiveness. Benchmarking efforts should assess how well existing tools support underserved DASC-PM tasks, including infrastructure setup, risk analysis, and post-deployment monitoring. Integration guidelines can help teams embed automation into workflows while managing compliance, scalability, and security. These workflows should align with both organizational processes and project requirements. Competency-to-tool mappings can support this alignment by helping teams select tools that match their existing competencies.

Automation tools that bridge these gaps can

strengthen collaboration and integration across all DASC-PM phases. This integration can reduce fragmentation, align competencies, and increase the success of data science projects. This study serves as a pre-study and provides a structured foundation for future research, tool development, and implementation efforts aimed at more holistic and effective automation.

REFERENCES

- Abbaszadegan, A. and Grau, D. (2015). Assessing the influence of automated data analytics on cost and schedule performance. *Procedia Engineering*, 123:3–6.
- Abdelaal, M., Koparde, R., and Schoening, H. (2023). Autocure: Automated tabular data curation technique for ml pipelines. In Bordawekar, R., Shmueli, O., Amsterdamer, Y., Firmani, D., and Kipf, A., editors, *Proceedings of the Sixth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management*, pages 1–11, New York, NY, USA. ACM.
- Aljohani, N. R., Aslam, A., Khadidos, A. O., and Hassan, S.-U. (2022). Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. *Journal of Innovation & Knowledge*, 7(3):100190.
- Alok Chakravarthy, N. and Padma, U. (2023). A comprehensive study of automation using a webapp tool

- for robot framework. In Hemanth, J., Pelusi, D., and Chen, J. I.-Z., editors, *Intelligent Cyber Physical Systems and Internet of Things*, volume 3 of *Engineering Cyber-Physical Systems and Critical Infrastructures*, pages 577–586. Springer International Publishing, Cham.
- Aryai, V., Mahdavi, N., West, S., and Henze, G. (2023). An automated data-driven platform for buildings simulation. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 61–68, New York, NY, USA. ACM.
- Bavishi, R., Laddad, S., Yoshida, H., Prasad, M. R., and Sen, K. (2021). Vizsmith: Automated visualization synthesis by mining data-science notebooks. In IEEE/ACM, editor, *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 129–141. IEEE.
- Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., and Kadir, R. A. (2022). Auto-prep: Efficient and automated data preprocessing pipeline. *IEEE Access*, 10:107764–107784.
- Boina, R., Achanta, A., and Mandvikar, S. (2023). Integrating data engineering with intelligent process automation for business efficiency. *International Journal of Science and Research (IJSR)*, 12(11):1736–1740.
- Brauner, S., Murawski, M., and Bick, M. (2025). The development of a competence framework for artificial intelligence professionals using probabilistic topic modelling. *Journal of Enterprise Information Management*, 38(1):197–218.
- Cuadrado-Gallego, J. J. and Demchenko, Y., editors (2020). *The Data Science Framework: A View from the EDISON Project*. Springer eBook Collection. Springer International Publishing and Imprint Springer, Cham, 1st ed. 2020 edition.
- Davenport, T. H., Mittal, N., and Saif, I. (2020). What separates analytical leaders from laggards? *MIT Sloan management review*.
- Demchenko, Y., Belloum, A., et al. (2022). *Data Science Competence Framework (CF-DS)*. EDISON Community Initiative.
- Demchenko, Y. and José, C. G. J. (2021). Edison data science framework (edsf): addressing demand for data science and analytics competences for the data driven digital economy. In IEEE, editor, *IEEE Global Engineering Education Conference (EDUCON)*, pages 1682–1687.
- Elshaw, R., Maher, M., and Sakr, S. (2019). Automated machine learning: State-of-the-art and open challenges.
- European Commission (2017). Education for data intensive science to open new science frontiers.
- European Committee for Standardization (2014). *European e-Competence Framework*.
- Galhotra, S. and Khurana, U. (2022). Automated relational data explanation using external semantic knowledge. *Proceedings of the VLDB Endowment*, 15(12):3562–3565.
- Gökay, G. T., Nazliel, K., Şener, U., Gökalp, E., Gökalp, M. O., Gençal, N., Dağdaş, G., and Eren, P. E. (2023). What drives success in data science projects: A taxonomy of antecedents. In García Márquez, F. P., Jamil, A., Eken, S., and Hameed, A. A., editors, *Computational Intelligence, Data Analytics and Applications*, volume 643 of *Lecture Notes in Networks and Systems*, pages 448–462. Springer International Publishing, Cham.
- Haertel, C., Pohl, M., Nahhas, A., Staegemann, D., and Turowski, K. (2022). Toward a lifecycle for data science: A literature review of data science process models. *Pacific Asia Conference on Information Systems 2022*.
- James, S. and Duncan, A. D. (2023). Over 100 data and analytics predictions through 2028. *Gartner Research*, pages 1–24.
- Joshi, M. P., Su, N., Austin, R. D., and Sundaram, A. K. (2021). Why so many data science projects fail to deliver. 62(3):84–90.
- Kitchenham, B. (2004). Procedures for performing systematic reviews.
- Kruhse-Lehtonen, U. and Hofmann, D. (2020). How to define and execute your data and ai strategy. *Harvard Data Science Review*.
- Kuckartz, U. and Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden*. Grundlagentexte Methoden. Beltz Juventa, Weinheim and Basel, 5. auflage edition.
- Kutzius, D., Dukino, C., and Kett, H. (2021). Towards a continuous process model for data science projects. In Leitner, C., Ganz, W., Satterfield, D., and Bassano, C., editors, *Advances in the Human Side of Service Engineering*, volume 266 of *Lecture Notes in Networks and Systems*, pages 204–210. Springer International Publishing, Cham.
- Lam, H. T., Buesser, B., Min, H., Minh, T. N., Wistuba, M., Khurana, U., Bramble, G., Salonidis, T., Wang, D., and Samulowitz, H. (2021). Automated data science for relational data. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 2689–2692. IEEE.
- Li, G., Yuan, C., Kamarthi, S., Moghaddam, M., and Jin, X. (2021). Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis. *Journal of Manufacturing Systems*, 60:692–706.
- Lwakatare, L. E., Rånge, E., Crnkovic, I., and Bosch, J. (2021). On the experiences of adopting automated data validation in an industrial machine learning project. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 248–257. IEEE.
- Macas, M., Lagla, L., Fuertes, W., Guerrero, G., and Toulkeridis, T. (2017). Data mining model in the discovery of trends and patterns of intruder attacks on the data network as a public-sector innovation. In Terán, L. and Meier, A., editors, *2017 Fourth International Conference on eDemocracy & eGovernment (ICEDEG)*, pages 55–62. IEEE, Piscataway, NJ.

- Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., and Dewhurst, M. (2017). A future that works: Automation, employment, and productivity.
- Martinez, I., Viles, E., and Olaizola, I. G. (2021). A survey study of success factors in data science projects. 1:2313–2318.
- Mavrogiorgos, K., Mavrogiorgou, A., Kiourtis, A., Zafeiropoulos, N., Kleftakis, S., and Kyriazis, D. (2022). Automated rule-based data cleaning using nlp. In *2022 32nd Conference of Open Innovations Association (FRUCT)*, pages 162–168. IEEE.
- Mikalef, P. and Krogtstie, J. (2019). Investigating the data science skill gap: An empirical analysis. In *2019 IEEE Global Engineering Education Conference (EDUCON)*, pages 1275–1284. IEEE.
- Mota, JR., Joselito, Santana, R., and Machado, I. (2021). Grumpy: an automated approach to simplify issue data analysis for newcomers. In *Brazilian Symposium on Software Engineering*, pages 33–38, New York, NY, USA. ACM.
- NewVantage Partners (2019). Big data and ai executive survey.
- Nidhra, S., Yanamadala, M., Afzal, W., and Torkar, R. (2013). Knowledge transfer challenges and mitigation strategies in global software development—a systematic literature review and industrial validation. *International Journal of Information Management*, 33(2):333–355.
- Patel, H., Guttula, S., Gupta, N., Hans, S., Mittal, R. S., and N, L. (2023). A data-centric ai framework for automating exploratory data analysis and data quality tasks. *Journal of Data and Information Quality*, 15(4):1–26.
- Petricek, T., den van Burg, G. J. J., Nazábal, A., Ceritli, T., Jiménez-Ruiz, E., and Williams, C. K. I. (2023). Ai assistants: A framework for semi-automated data wrangling. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):9295–9306.
- Pitroda, H. (2022). A proposal of an interactive web application tool quickviz: To automate exploratory data analysis. In *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, pages 1–8. IEEE.
- Potanin, M., Holtkemper, M., and Beecks, C. (2024). Exploring the integration of data science competencies in modern automation frameworks: Insights for workforce empowerment. In *Intelligent Systems Conference*, pages 232–240.
- Ransbotham, S., Khodabandeh, S., Fehling, R., LaFountain, B., and Kiron, D. (2019). Winning with ai. *MIT Sloan management review*.
- Reddy, R. C., Mishra, D., Goyal, D. P., and Rana, N. P. (2024). A conceptual framework of barriers to data science implementation: a practitioners’ guideline. *Benchmarking: An International Journal*, 31(10):3459–3496.
- Robinson, E. and Nolis, J. (2020). *Build a career in data science*. Manning, Shelter Island.
- Salminen, K., Hautamäki, P., and Jähi, M. (2024). Aligning industry needs and education: Unlocking the potential of ai via skills. In *2024 Portland International Conference on Management of Engineering and Technology (PICMET)*, pages 1–10. IEEE.
- Saltz, J. S. (2021). Crisp-dm for data science: strengths, weaknesses and potential next steps. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2337–2344.
- Santana, M. and Díaz-Fernández, M. (2023). Competencies for the artificial intelligence age: visualisation of the state of the art and future perspectives. *Review of Managerial Science*, 17(6):1971–2004.
- Schulz, M., Neuhaus, U., Kaufmann, J., Kühnel, S., Alekozai, E. M., Rohde, H., Hoseini, S., Theuerkauf, R., Badura, D., Kerzel, U., et al. (2022). Dasc-pm v1.1 a process model for data science projects.
- Shrivastava, S., Patel, D., Bhamidipaty, A., Gifford, W. M., Siegel, S. A., Ganapavarapu, V. S., and Kalagnanam, J. R. (2019). Dqa: Scalable, automated and interactive data quality advisor. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2913–2922. IEEE.
- Thawanthaleunglit, N. and Sripanidkulchai, K. (2019). Sweeper. In *Proceedings of the 2019 3rd International Conference on Software and e-Business*, pages 17–23, New York, NY, USA. ACM.
- Varde, A. S. (2022). Computational estimation by scientific data mining with classical methods to automate learning strategies of scientists. *ACM Transactions on Knowledge Discovery from Data*, 16(5):1–52.
- VentureBeat (2019). Why do 87% of data science projects never make it into production?
- Wang, D., Andres, J., Weisz, J. D., Oduor, E., and Dugan, C. (2021). Autods: Towards human-centered automation of data science. In Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., and Drucker, S., editors, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, New York, NY, USA. ACM.
- Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., and Gray, A. (2019). Human-ai collaboration in data science: Exploring data scientists’ perceptions of automated ai. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24.
- Webster, J. and Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, (26):13–23.