EXAMINING THE IMPACT OF AI AGENTS ON DATA SCIENTISTS' SKILLS

Papageorgiou, Giorgos[1]; Lekakos, George[1]

[1]Department of Management Science and Technology, Athens University of Economics and Business, Athens, Greece

Abstract

The growing adoption of AI agents marks a new inflection point for data science, as agentic systems increasingly execute multi-step analytical workflows across the data-science pipeline. This study examines the positioning of AI agent tools across the data-science pipeline and how their documented use reshapes the allocation of work between data scientists and AI systems. Treating six representative tools as witness cases, we reconstruct documented workflows to provide an existence proof of agentic coverage across all pipeline stages and synthesize recurring patterns in human–agent collaboration. The analysis demonstrates that AI agents assume responsibility for multi-step execution across the full pipeline, producing concrete analytical artefacts. In these workflows, the data scientist's contribution shifts toward directing agent execution and validating actions through inspection of intermediate artefacts. Building on an established competency framework, the paper articulates two agent-oriented skill extensions—Agent Steering and Agent Process Tracing—that capture these emerging responsibilities for data professionals.

Keywords: Artificial intelligence, AI Agents, Data Pipelines, Data Scientist skills

Introduction

Over the past two decades, organisations have transitioned from intuition-led decision-making to data-driven optimization, placing the data scientist at the center of organisational analytics. In practice, data scientists coordinate work across multiple stages—sourcing and preparing data, exploring and modeling it, and communicating results into decisions and deliverables. Today, this role faces another inflection point with the emergence of AI agents: goal-oriented systems that can plan, use tools, and iteratively refine multi-step workflows, producing concrete analytical artefacts such as structured datasets, code, models, or visualizations.

Unlike earlier forms of automation and single-turn AI assistance, AI agents increasingly support iterative, multi-step execution within specific stages of the data-science pipeline, often in close interaction with human users. Because these systems generate intermediate and final artefacts—

such as transformation steps, SQL queries, or narrative outputs—that shape the analytical process, the central issue is not only what AI can produce, but how analytical work is distributed and coordinated between humans and AI systems in practice.

Against this backdrop, this paper examines how AI agent tools are positioned across the data-science pipeline and how their use reshapes human–AI task allocation. Specifically, the study addresses two research questions: (RQ1) Where are AI agent tools present across the stages of the data-science pipeline, and what roles do they play at each stage? and (RQ2) Which human skills are foregrounded for data scientists as AI agents increasingly mediate data-science workflows? The paper makes two main contributions. First, it provides a pipeline-wide mapping of AI agent tools, showing how they are represented across multiple stages of work and what types of artefacts they produce. Second, by consolidating recurring human-executed tasks, the study derives two agent-oriented skill extensions—Agent Steering and Agent Process Tracing—that expand upon established competency frameworks to capture how human expertise is exercised in agent-mediated environments.

Literature Review

Data scientist role and baseline skill set

While the "data scientist" has emerged as a cornerstone of modern analytics, the role's definition remains fluid, traditionally positioned at the intersection of statistics, computer science, and domain expertise. Because boundaries between data scientists, engineers, and analysts are frequently contested (Fayyad and Hamutcu, 2022), this study treats "data scientist" as an umbrella term for professionals across the data lifecycle. To establish a skill baseline, we adopt Vinay's (2024) framework, which consolidates technical, analytical, and professional competencies. It captures proficiency in programming and data tooling, statistical and analytical reasoning, exploration analysis and visualization, domain understanding, communication, structured problem solving, ethical awareness, and continuous learning (Vinay, 2024). This baseline provides the reference point for evaluating how agentic paradigms reshape human practice.

Data science pipeline as an organizing frame

Data-science work is organized as a pipeline where data is transformed into actionable outputs. These pipelines are iterative and non-linear, adapted to specific contexts to support maintainability and reusability (Biswas et al., 2022). We adopt the 11 canonical stages identified by Biswas et al. (2022) as our organizing lens: data acquisition, preparation, storage, feature engineering, modeling, training, evaluation, prediction/inference, interpretation,

communication, and deployment. Following this framework, deployment is treated as an ongoing operational phase rather than a one-off release: once a solution is installed, its performance is monitored over time so it can be maintained and improved. Interpretation is defined as downstream post-processing of prediction/inference outputs to translate them into decision-relevant understanding; for instance, while numerical outputs alone may be insufficient, visualization can help stakeholders make decisions (Biswas et al., 2022). This framing is essential for accurately positioning agentic AI tools that support operational and interpretative tasks within the empirical mapping that follows.

AI agents and multi-agent systems

In the classical AI literature, an agent is defined as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" (Russell and Norvig, 2020). Because this study examines modern data-pipeline tools, we adopt a contemporary, LLM-era framing. In this context, AI agents are modular systems enabled by Large Language Models for task-specific automation, defined as autonomous software programs that perform specific tasks (Sapkota et al., 2026). Key characteristics include autonomy, task-specificity, and reactivity to real-time stimuli such as user requests, external API calls, or state changes in software environments. This framing also extends to agentic AI systems—explicitly framed as multi-agent systems—which consist of multiple AI agents collaborating to achieve complex goals (Sapkota et al., 2026). Accordingly, our review includes AI tools that operate as either single-agent or multi-agent systems.

Current context and research focus

Recent evidence suggests that AI agents are transitioning from experimental prototypes toward real-world deployment, with many organisations expecting to adopt them in the near term (World Economic Forum, 2025a). Simultaneously, employer-focused evidence indicates that AI-related capabilities are becoming an important labor-market signal; the Future of Jobs Report 2025 notes that two-thirds of surveyed employers plan to hire talent with specific AI skills (World Economic Forum, 2025b).

Yet two gaps remain. First, despite this momentum, the literature has not systematically mapped where AI agents are emerging across the data-science pipeline, leaving unclear whether agentic coverage is concentrated in specific stages or extends pipeline-wide. Second, established competency models such as Vinay's (2024) treat the data scientist as the executor of analytical work, while agentic systems increasingly take over that execution—raising the question of which human competencies are foregrounded when AI carries out multi-step workflows.

Research Design and Methodology

Methodological positioning and research design

This study adopts a documentation-based comparative case analysis to examine how AI agents are positioned across the data-science pipeline (RQ1) and which human tasks persist as agents mediate analytical workflows (RQ2). Both questions concern the design intent of agentic systems rather than their empirical performance or adoption in practice, making official documentation the appropriate evidence base. The study therefore follows Bowen's (2009) document analysis method and applies Merriam and Tisdell's (2016) multi-case logic, in which each tool's documented workflow is treated as a distinct case: first examined through within-case reconstruction and then integrated through cross-case synthesis to surface recurring patterns. The analysis proceeds through four steps:

Workflow Reconstruction: For each tool, evidence is drawn exclusively from sources authored by its developers—official product documentation, technical papers, and developer-authored technical articles—and synthesized to reconstruct the full end-to-end operational workflow. This ensures that the reconstructed workflow reflects the tool's prescribed operation rather than third-party interpretation.

Task Decomposition: These reconstructed workflows are decomposed into discrete, granular tasks to facilitate a high-resolution analysis of the work process.

Rule-based Task Attribution: Each identified task is assigned a performer (Human vs. AI) based directly on the operational language prescribed in the documentation: a task is coded as Human when the documentation explicitly prescribes user action (e.g., the user *submits*, *reviews*, or *provides input*) and as AI when it describes system execution (e.g., the system *generates*, *retrieves*, or *produces*). Because attribution follows documented operational language rather than researcher interpretation, the procedure is replicable: another researcher applying the same rules to the same documentation would reach the same coding.

Evidence Synthesis: Case evidence is aggregated to address the research questions. For RQ1, data is synthesized to map tool presence and functionality across the 11 stages identified by Biswas et al. (2022). For RQ2, human-executed tasks from all tables are consolidated into a cross-tool catalogue, grouped into clusters of similar tasks, and translated into corresponding skills following the JRC task-skill framework (Rodrigues et al., 2021), providing the empirical basis for the proposed skill extensions.

Case selection and evidence base

Cases were selected through stratified logic in which every canonical stage of the Biswas et al. (2022) data-science pipeline is covered by at least one witness case, providing existence proof of agentic coverage across the full pipeline (RQ1). Three inclusion criteria were applied: (i) the tool must exhibit agentic behaviour—either explicitly labelled as such or demonstrably LLM-enabled, autonomous, task-specific, and reactive, consistent with Sapkota, Roumeliotis and Karkee (2026); (ii) its end-to-end workflow must be sufficiently documented in authoritative first-party sources to permit reconstruction without speculative inference; and (iii) it must map clearly onto a stage of the Biswas et al. (2022) framework. Among eligible candidates per stage, the tool with the most complete and transparent documented workflow was retained, yielding the cohort of six witness cases analysed below.

Empirical Analysis

Data acquisition with AD-AutoGPT

AD-AutoGPT is presented as an autonomous agent that leverages GPT-4 to execute a prompt-driven workflow for Alzheimer's disease infodemiology—i.e., the systematic collection and interpretation of online information signals (Dai et al., 2025). Starting from high-level user prompts (e.g., specifying topic, time window, and preferred sources), the system performs data acquisition by searching for Alzheimer's-related news via a Google API–based process and saving retrieved URLs and available metadata as a structured corpus (Dai et al., 2025). It then crawls the collected pages, summarizes article text, and produces structured information for downstream use, including spatiotemporal signals (locations derived via geoparsing and time derived from news metadata) and topic representations computed with LDA (Dai et al., 2025). Finally, the workflow supports interpretation of the corpus through analysis-oriented outputs and visualizations, such as location maps, time-based frequency plots (e.g., monthly counts), and topic-focused views (e.g., keyword summaries and intertopic distance/trend visualizations) (Dai et al., 2025). The tool records retrieved URLs and associated metadata as part of the structured corpus (Dai et al., 2025). This makes outputs traceable, enabling users to review the relevance and credibility of collected sources and, if necessary, issue follow-up prompts to refine or redirect the acquisition process.

#	Phase	Task Description	Performer
1	Goal Definition	Define specific instructions regarding data acquisition goal	Human
2	Source Discovery	Identify relevant online sources	AI
3	Data Extraction	Collect data & extract key information from text	AI

4	Data Structuring	Create structured, analysis-ready dataset	AI
5	Trend & Pattern Analysis	Detect temporal, spatial, and topic patterns	AI
6	Visualization Generation	Produce maps, timelines, and trend graphs	AI
7	Result Review	Assess outputs for relevance and accuracy	Human
8	Iterative Refinement	Adjust instructions and redirect the agent	Human

Table 1: AD-AutoGPT — Task Allocation.

Observation. The table shows the AI performing retrieval and structuring, while the human sets scope, reviews outputs, and redirects the workflow when needed.

Data preparation with Gemini in BigQuery

Gemini in BigQuery spans data preparation and storage by generating SQL-based transformation steps and materializing the results as persisted BigQuery tables in a cloud data-warehouse environment. In the documented workflow, the user may start with Gemini-suggested steps and/or provide their own natural-language prompt (via a description field) to specify an intended transformation; Gemini then generates the corresponding SQL expression and step description (Google Cloud, 2026a). The process is iterative: users can preview results, revise the prompt to regenerate the SQL expression, or edit the expression directly before applying the step, enabling refinement across successive transformations (Google Cloud, 2026b). The resulting artefacts are traceable and reusable within BigQuery, consisting of the generated SQL expressions and a destination table produced by running the sequence of steps (Google Cloud, 2026a). The user then assesses whether the output is satisfactory and either saves the preparation flow or continues iterating through further adjustments (Google Cloud, 2026a).

Google does not explicitly label Gemini in BigQuery data preparation as an agent. Accordingly, we treat it as exhibiting agent-like behaviour under our AI-agent framing, since its documented workflow supports task-specific automation from user prompts through iterative SQL generation and execution, with refinement based on intermediate results and user revisions—consistent with autonomy (within the tool), task-specificity, and reactivity (Sapkota et al., 2026).

#	Phase	Task Description	Performer
1	Start	Provide high-level natural language instruction	Human
2	Plan fixes	Generate multi-step plan of transformations	AI
3	Generate cleaning steps	Produce concrete SQL expressions	AI
4	Review	Inspect suggested steps	Human
5	Edit	Edit suggestions if needed	Human

6	Run & check results	Execute steps and produce cleaned table	AI
7	Finalize or continue	Decide whether further adjustments are needed	Human
8	Refinement loop	Update SQL when user provides new instructions	AI

Table 2: Gemini in BigQuery — Task Allocation.

Observation. The table indicates that the AI generates and applies transformations (as SQL), while the human defines intent, checks intermediate results, and iterates through revisions.

Multi-agent ML pipeline automation with MLZero

MLZero is presented as a multi-agent system designed to automate the core stages of the machine-learning pipeline, spanning feature engineering, modeling, training, evaluation and prediction. In the documented framework, inputs include the data and a brief task description or instruction, and the system decomposes the objective into coordinated subtasks handled by nine specialized agents distributed across perception, memory, and iterative coding modules (Fang et al., 2025). Operationally, the workflow begins with perception components that parse the task description and inspect available input files to establish the learning target and constraints, followed by a library-selection step that chooses an appropriate ML toolkit for the task. The system then enters an iterative coding loop in which code is generated, executed, evaluated, and revised using feedback from runtime logs and errors, supported by memory components that compress relevant documentation and retain prior failures to reduce repetition (Fang et al., 2025). The framework also allows optional user guidance during the iterative loop (i.e., optional per-iteration user input) when needed, while otherwise maintaining a high degree of automation. The documented outputs include executable training and inference scripts, run logs, trained artefacts, and task-specific prediction files, providing a traceable record of the automated pipeline execution (Fang et al., 2025).

#	Phase	Task Description	Performer
1	Problem framing	Define high-level ML objective & initial instructions	Human
2	Data access	Provide raw data folder	Human
3	Problem understanding	Analyse files/inputs & infer task/constraints	AI
4	Modeling setup	Select ML library	AI
5	Memory initialization	Initialize semantic & episodic memory	AI
6	Iterative coding	Provide/update per-iteration instructions	Human
7	Modeling	Generate training & inference code	AI
8	Execution	Execute code & analyse results	AI
9	Debugging	Summarize errors & retrieve relevant docs	AI
10	Iterative improvement	Refine and regenerate code	AI
11	Output generation	Produce predictions & artefacts	AI

12	Output validation	Inspect final outputs	Human
13	Re-orchestration (optional)	Restart workflow with updated instructions	Human

Table 3: MLZero — Task Allocation.

Observation. The table highlights automated planning/coding/execution by the system, while the human provides the objective and evaluates whether outputs satisfy the task.

Interpretation with Wren AI

Wren AI is described by its developers as an open-source GenBI agent and can be positioned primarily within the interpretation stage, where users query a connected database through natural-language questions (WrenAI, 2025e). In the documented workflow, the user submits a question in plain language and the system generates the underlying SQL and executes it against the connected data source to return results (WrenAI, 2025a). A central component supporting this behaviour is Wren's Modeling Definition Language (MDL), which provides a semantic layer for defining data models and relationships so that generated queries align with the intended organisational semantics (WrenAI, 2025d). Wren also exposes the executed SQL (e.g., via a 'View SQL' option), allowing users to inspect it (WrenAI, 2025a). After execution, Wren presents the answer as an integrated response that includes a concise textual summary and a preview of the resulting data (WrenAI, 2025c). The tool also supports visualization by automatically generating an appropriate chart for the returned results, which can be accessed in a dedicated chart view/tab (WrenAI, 2025b). In the chart view, users can request a regenerated chart (e.g. "Increase the font size"), which re-runs the visualization pipeline (WrenAI, 2025b).

#	Phase	Task Description	Performer
1	NL query interpretation	Submit question	Human
2	NL query interpretation	Interpret question & infer intent (LLM+MDL)	AI
3	SQL generation/execution	Generate & execute SQL statement	AI
4	Chart selection	Select, configure & render chart	AI
5	Insight generation	Generate narrative summary / report output	AI
6	Exploration	Inspect SQL	Human
7	Exploration	Revise question	Human
8	Exploration	Suggest follow-up questions	AI
9	Visualization customization	Request regenerated chart	Human
10	Visualization customization	Re-run visualization pipeline	AI
11	Final output	Review final answer	Human

Table 4: Wren AI — Task Allocation.

Observation. The table reflects AI-led query/visual generation from questions, with the human primarily steering the exploration through follow-up prompts and validation.

Presentation with PowerPoint Copilot

PowerPoint Copilot can be positioned in the communication and presentation stage, where it supports generating and refining slide decks from user intent expressed in natural language and from user-provided files (Microsoft, 2025a). In the documented workflow, the user provides a prompt (and may upload or reference supporting documents), after which Copilot extracts key points from the input and organizes them into an initial slide structure (Microsoft, 2025a). Copilot then generates slide text, applies layouts and basic styling within PowerPoint, and can draft speaker notes to accompany the slides (Microsoft, 2025b). The workflow is iterative: the user reviews the draft for accuracy and fit with the intended message and issues follow-up rewrite requests or edits to refine slide wording and structure as the deck evolves (Microsoft, 2025c).

Microsoft does not explicitly label PowerPoint Copilot as an "agent." We therefore classify it as agent-like under our AI-agent framing, as its documented LLM-enabled, prompt-driven workflow supports task-scoped slide generation and iterative refinement in response to user feedback—aligning with the agent features adopted in our literature framing.

#	Phase	Task Description	Performer
1	Prompting	Provide prompt with specific data & upload files	Human
2	Content Extraction	Extract key topics and structure	AI
3	Slide Structuring	Convert ideas into slide outline	AI
4	Slide Generation	Generate text content	AI
5	Design Automation	Apply layouts/styling and suggest images/icons	AI
6	Speaker Notes Drafting	Draft speaker notes	AI
7	Human Review	Verify accuracy	Human
8	Iterative Revision	Request rewrites	Human

Table 5: Microsoft PowerPoint Copilot — Task Allocation.

Observation. The table shows the AI generating and revising slide artefacts, with the human directing content goals and approving or correcting intermediate drafts.

Post-deployment troubleshooting with Gemini Cloud Assist Investigations

Gemini Cloud Assist investigations can be positioned within post-deployment operations (incident investigation and troubleshooting), which fall within the deployment stage as defined by Biswas et al. (2022), and is described as a root-cause analysis AI agent for diagnosing issues in cloud infrastructure and applications (Google Cloud, 2025). In the documented workflow,

the user starts an investigation in the Google Cloud console by describing the issue (and, when applicable, adding context such as a time window or affected resources) (Google Cloud, 2026c). The system then examines the operational data already available in the cloud environment—such as logs, configuration information, and metrics—and surfaces ranked Observations that summarize what it found, each linked to the underlying evidence so users can verify the basis of the finding (Google Cloud, 2026c). Finally, it synthesizes these observations into probable root-cause hypotheses and recommended next steps, and it supports iterative troubleshooting by allowing the investigation context to be revised and rerun as new information becomes available (Google Cloud, 2026c).

#	Phase	Task Description	Performer
1	Investigation setup	Describe the issue and define time window / scope	Human
2	Signal aggregation	Collect and correlate logs, configurations, and metrics	AI
3	Sense-making	Generate ranked observations from collected signals	AI
4	Diagnosis	Propose hypotheses and recommended fixes	AI
5	Evaluation	Assess hypotheses, inspect evidence	Human
6	Iteration or Resolution	Refine and re-run investigation if needed	Human

Table 6: Gemini Cloud Assist Investigations — Task Allocation.

Observation. The table indicates AI-led evidence synthesis into observations and recommended actions, with the human providing incident context and confirming next steps.

Results

Pipeline coverage of AI agent tools

This section synthesizes findings across cases to address RQ1, which concerns the presence and positioning of AI agent tools across stages of the data-science pipeline. Across the six cases examined, AI agent tools were identified in every major stage of the pipeline, spanning upstream data acquisition and preparation, core machine-learning workflows, interpretation, communication, and post-deployment operations. Although each tool targets specific pipeline stages, the cases collectively demonstrate that AI agent tools span the entire data-science pipeline. Figure 1 summarizes this mapping by situating each tool within the pipeline stage(s) it supports. The figure illustrates both the breadth of coverage and the functional diversity of AI agent tools across the end-to-end workflow, establishing the context for the cross-tool task

consolidation developed next (ML pipeline includes feature engineering, modeling, training, evaluation and prediction).

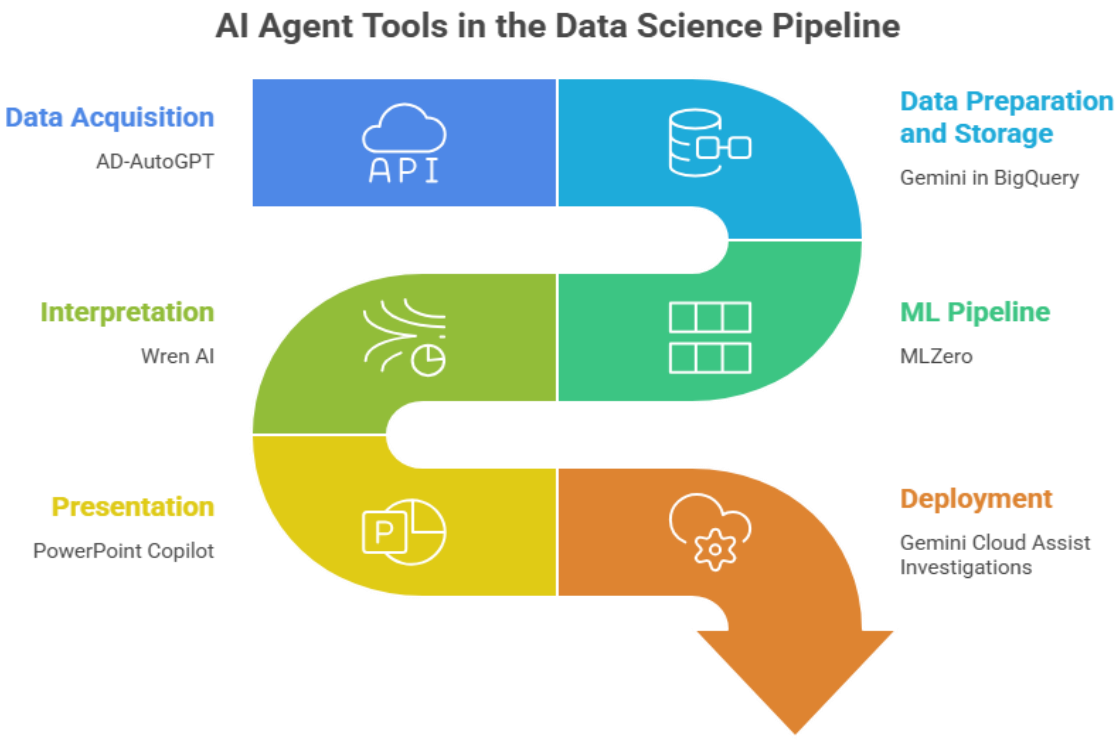

Figure 1: Mapping AI tools in the Data Science Pipeline, ML pipeline includes 5 stages. Visual elements generated with AI tool; conceptual design by the author.

Cross-tool consolidation of human tasks

We aggregated all human-executed tasks across the six AI agent tools into a single cross-tool catalogue. The tasks clearly converged into two dominant clusters (Table 7): Directive tasks initiate or redirect the agent's trajectory (e.g., setting goals, specifying constraints, refining instructions), whereas evaluative tasks involve inspecting intermediate artefacts and judging output reliability and acceptability.

Tool	Human task	Category
AD-AutoGPT	Define specific instructions regarding data acquisition goal	Directive
	Assess outputs for relevance and accuracy	Evaluative
	Adjust instructions and redirect the agent	Directive
Gemini in BigQuery	Provide high-level natural language instruction	Directive
	Inspect suggested steps	Evaluative
	Edit suggestions if needed	Directive
	Decide whether further adjustments are needed	Directive
Wren AI	Submit question	Directive
	Inspect SQL	Evaluative

	Revise question	Directive
	Request regenerated chart	Directive
	Review final answer	Evaluative
MLZero	Define high-level ML objective & initial instructions	Directive
	Provide raw data folder	Directive
	Provide/update per-iteration instructions	Directive
	Inspect final outputs	Evaluative
	Restart workflow with updated instructions	Directive
PowerPoint Copilot	Provide prompt with specific data & upload files	Directive
	Verify accuracy	Evaluative
	Request rewrites	Directive
Gemini Cloud Assist Investigations	Describe the issue and define time window / scope	Directive
	Assess hypotheses, inspect evidence	Evaluative
	Refine and re-run investigation if needed	Directive

Table 7. Cross-tool consolidation of human tasks.

From task clusters to skills

Translating these clusters into skill implications follows the European Commission Joint Research Centre task–skill framework. The framework defines a task as the smallest unit of work in an economic process and a skill as the ability to perform a task well, with clusters of similar tasks constituting task domains that correspond to skill domains (Rodrigues et al., 2021). Applied here, the Directive cluster forms a task domain centered on goal-setting and redirection, while the Evaluative cluster forms a task domain centered on artefact-based inspection and judgment; each corresponds to a distinct skill domain, yielding the two agent-oriented skills proposed below (see Figure 2):

Agent Steering (directive cluster). In research on agentic information systems, *steerability* is typically discussed as a property of the AI system—i.e., the extent to which an agent can be directed to align with user goals and values (Saffarizadeh et al., 2024). Building on this notion while shifting the locus of analysis to the user, we define *Agent Steering* as the human capability required to exploit steerable agents in practice: *the capacity of data professionals to actively guide and redirect an AI agent's workflow over time toward an intended outcome.* Agent Steering involves two interrelated abilities*: (i) articulating goals and prompts that are sufficiently clear and structured to guide the agent's operations; and (ii) formulating revised instructions that redirect the agent accordingly.*

Agent Process Tracing (evaluative cluster). The evaluative cluster is grounded in the logic of *process tracing*, which examines sequences and diagnostic evidence to understand how

outcomes are produced (Collier, 2011). Transferred to AI agent workflows, this becomes the ability to reconstruct and evaluate the agent's execution by reading its artefacts as evidence. We therefore define *Agent Process Tracing* as *the data professional's capacity to reconstruct and critically assess the path an AI agent has taken through a workflow by inspecting its intermediate and final artefacts*. This includes interpreting how intermediate outputs (e.g., steps, SQL, code, tables, visualizations, or narrative deliverables) relate to one another, inferring how the task was operationalized, and judging whether the resulting trajectory and outputs are methodologically appropriate and reliable.

In practice, the two skills function as a recurrent interaction cycle: steering initiates or redirects the workflow; process tracing evaluates the produced artefacts; and evaluation informs subsequent steering decisions. Building on Vinay's (2024) competency framework as a baseline, we propose these two skills as an additional, agent-oriented competency pillar. They draw on established capabilities (e.g., problem formulation, technical literacy, analytical judgment) but are distinguished by a different locus of action: governing and validating autonomous agent execution rather than directly performing most operational steps.

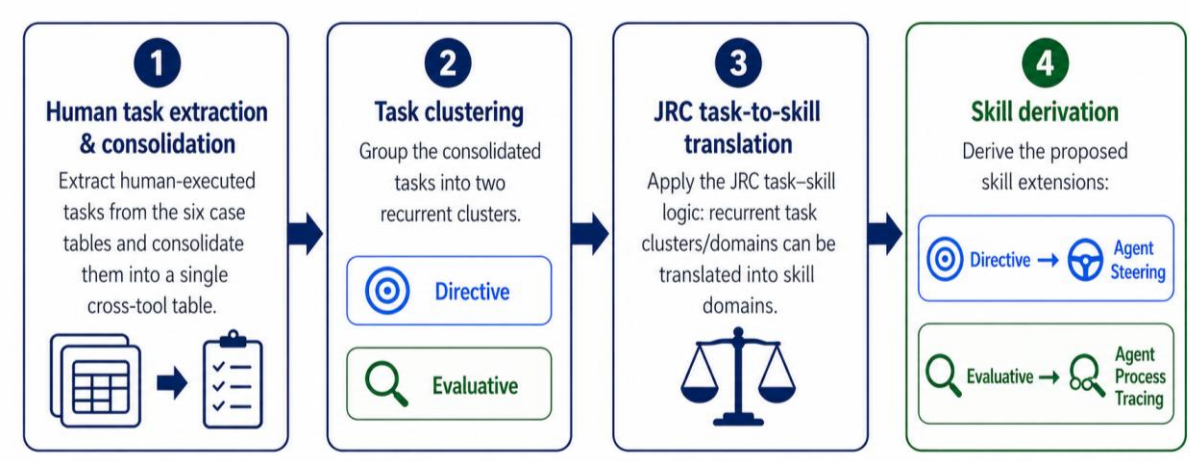

Figure 2: Empirical pathway from human tasks (Tables 1–6) to the proposed skill extensions, structured by the JRC task–skill framework. Visual elements generated with AI tool; conceptual design by the author.

Conclusion and Discussion

The emergence of AI agents marks a fundamental inflection point for data science, shifting the boundaries of automation from single-task execution to the autonomous management of complex workflows. Addressing RQ1, this study utilizes the six analysed tools as witness cases to provide existence proof of agentic coverage across all data pipeline stages, demonstrating that agentic workflows now span every stage of the data-science pipeline. From upstream acquisition to deployment, these cases illustrate how agents are assuming the operational

burden of producing analytical artefacts—code, models, and visualizations—thereby liberating professionals from traditional manual execution. However, this transition is not a story of human displacement, but one of profound role reconfiguration (RQ2). The data scientist's contribution is shifting from direct execution to strategic orchestration and critical validation. By synthesizing our findings, we propose two skill extensions that expand the established Vinay (2024) framework: Agent Steering and Agent Process Tracing. Steering represents the capacity to strategically guide agents toward intended goals, while Process Tracing enables the reconstruction and auditing of the agent's trajectory to ensure methodological integrity.

For practitioners, effective work with AI agents is now characterised by an evidence-based steer–trace–revise cycle: professionals define high-level objectives, monitor intermediate execution through the inspection of artefacts, and validate progress before accepting or refining outputs. For tool designers, these findings highlight urgent priorities for observability—exposing the agent's internal trajectory and action history—and controllability—providing explicit intervention points such as stepwise execution or manual overrides. Such design elements are essential for aligning agentic tools with the accountability and transparency demands of professional data work.

Limitations and Future Research

Two limitations should be noted. First, the rapid evolution of the agent ecosystem means that the analysis should be read as a contemporary snapshot, with documented workflows likely to expand and shift as new tools emerge. Second, the design-level focus of the study does not assess how well agents actually execute the workflows their documentation prescribes; execution quality and reliability under real-world conditions remain outside the scope. By mapping where AI agents are positioned across the data-science pipeline and identifying the skill extensions that emerge in agent-mediated work, the study provides a conceptual foundation for empirical follow-up.

Future research should test these skill propositions through primary data collection with practicing data scientists—using semi-structured interviews and questionnaires—to validate Agent Steering and Agent Process Tracing in practice and examine their impact on workflow quality.

References

Biswas, S., Wardat, M. and Rajan, H. (2022), "The Art and Practice of Data Science Pipelines: A Comprehensive Study of Data Science Pipelines in Theory, In-the-Small, and In-the-Large",

Proceedings of the 44th International Conference on Software Engineering (ICSE 2022), pp. 1-13.

Bowen, G. (2009), "Document Analysis as a Qualitative Research Method", *Qualitative Research Journal*, Vol. 9, No. 2, pp. 27-40.

Collier, D. (2011), "Understanding Process Tracing", *PS: Political Science & Politics*, Vol. 44, No. 4, pp. 823-830.

Dai, H., Li, Y., Liu, Z., Zhao, L. and Wu, Z. (2025), "AD-AutoGPT: An Autonomous GPT for Alzheimer's Disease Infodemiology", *PLOS Global Public Health*, Vol. 5, No. 5.

Fang, H., Han, B., Erickson, N. and Zha, X. (2025), "MLZero: A Multi-Agent System for End-to-end Machine Learning Automation", *Proceedings of NeurIPS 2025*, available at: https://neurips.cc/virtual/2025/loc/san-diego/poster/119504

Fayyad, U. and Hamutcu, H. (2022), "From Unicorn Data Scientist to Key Roles in Data Science: Standardizing Roles", *Harvard Data Science Review*, Vol. 4, No. 3.

Google Cloud (2025), "How Gemini Cloud Assist investigations works", *Google Cloud Blog*, available at: https://cloud.google.com/blog/products/management-tools/gemini-cloud-assist-investigations-performs-root-cause-analysis (accessed January 2026).

Google Cloud (2026a), "Introduction to BigQuery data preparation", *Google Cloud Documentation*, available at: https://docs.cloud.google.com/bigquery/docs/data-prep-introduction (accessed January 2026).

Google Cloud (2026b), "Prepare data with Gemini", *BigQuery Documentation*, available at: https://docs.cloud.google.com/bigquery/docs/data-prep-get-suggestions (accessed January 2026).

Google Cloud (2026c), "Troubleshoot issues with Cloud Assist Investigations", *Google Cloud Documentation*, available at: https://docs.cloud.google.com/cloud-assist/investigations (accessed January 2026).

Merriam, S.B. and Tisdell, E.J. (2016), *Qualitative Research: A Guide to Design and Implementation*, 4th Edition, San Francisco, CA: Jossey-Bass.

Microsoft (2025a), "Create a new presentation with Copilot in PowerPoint", available at: https://support.microsoft.com/en-us/office/create-a-new-presentation-with-copilot-in-powerpoint-3222ee03-f5a4-4d27-8642-9c387ab4854d (accessed January 2026).

Microsoft (2025b), "Keep your presentation on-brand with Copilot", available at: https://support.microsoft.com/en-us/topic/keep-your-presentation-on-brand-with-copilot-046c23d5-012e-49e0-8579-fe49302959fc (accessed January 2026).

Microsoft (2025c), "Prepare your presentation with Microsoft 365 Copilot", available at: https://support.microsoft.com/en-us/office/prepare-your-presentation-with-microsoft-365-copilot-7f06429e-c0c2-4819-8119-b519ad599796 (accessed January 2026).

Rodrigues, M., Fernández-Macías, E. and Sostero, M. (2021), *A Unified Conceptual Framework of Tasks, Skills and Competences*, Seville: European Commission, Joint Research Centre, available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC121897

Russell, S. and Norvig, P. (2020), *Artificial Intelligence: A Modern Approach*, 4th Edition, Hoboken, NJ: Pearson.

Saffarizadeh, K., Keil, M. and Maruping, L. (2024), "Relationship Between Trust in the Artificial Intelligence Creator and Trust in Artificial Intelligence Systems: The Crucial Role of Artificial Intelligence Alignment and Steerability", *Journal of Management Information Systems*, Vol. 41, No. 3, pp. 645-681.

Sapkota, R., Roumeliotis, K. and Karkee, M. (2026), "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges", *Information Fusion*, Vol. 126, Part B.

Vinay, S.B. (2024), "Data Scientist Competencies and Skills Assessment: A Comprehensive Framework", *International Journal of Data Scientist (IJDST)*, Vol. 1, No. 1, pp. 1-14.

World Economic Forum (2025a), *AI Agents in Action: Foundations for Evaluation and Governance*, World Economic Forum, available at: https://www.weforum.org/publications/ai-agents-in-action-foundations-for-evaluation-and-governance/ (accessed January 2026).

World Economic Forum (2025b), *The Future of Jobs Report 2025*, World Economic Forum, available at: https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/ (accessed January 2026).

WrenAI (2025a), *Asking Questions*, available at: https://docs.getwren.ai/cloud/guide/home/ask (accessed January 2026).

WrenAI (2025b), *Generate Chart*, available at: https://docs.getwren.ai/oss/guide/home/chart (accessed January 2026).

WrenAI (2025c), *Know Your Answers*, available at: https://docs.getwren.ai/cloud/guide/home/answer (accessed January 2026).

WrenAI (2025d), *What is Modeling Definition Language (MDL)*, available at: https://docs.getwren.ai/oss/engine/concept/what_is_mdl (accessed January 2026).

WrenAI (2025e), *WrenAI Documentation*, available at: https://github.com/Canner/WrenAI (accessed January 2026).