© NOV 2023 | IRE Journals | Volume 7 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV7I5-1722574

Whether the Data Scientist Role Survives the Rise of Automated Analysis

OZOYA AMOS OHILEBO
Point Park University, Pittsburgh, Pennsylvania, United States

***Abstract-** In a familiar scenario, fear that the technology that made an occupation possible and popular will make the job superfluous is now re-emerging in a new arena: automated analysis. With more and more technical tasks being done by an automated machine learning system, commentators wonder whether or not the data scientist will exist! It is the aim of this paper to try to answer this question, and to suggest that the role may remain, but it must be redefined, and that the redefinition has a pattern which is well-studied in the history of automation. The paper builds on a body of research dating back to 2015 and up to 2023 covering the automation of employment, the economics of artificial intelligence, and how the role of data scientist has itself evolved, to explain what automated analysis replaces and what it does not. It argues that automation is for the components of data science, model building, model tuning, and routine analysis that are executable, well-defined, and repeatable; and that the framing of the problem, the exercise of judgment, and the integration of the analysis into organizational decision-making are left to humans. Since a job consists of a set of tasks, some of which can be automated, automation of some of the tasks results in a different mix of tasks in the job, not its disappearance, and that the human contribution to the tasks that cannot be automated should be redirected toward tasks that can be automated, thereby effectively making them more valuable, in line with the principle of complementarity. The end result of the paper is that the data scientist will not disappear, but will shift from a technical execution to a judgmental, problem-defining, and organizational-translating role, and that the work of the data scientist and the profession must embrace this change, not fear it.*

***Keywords:** data scientist, automated analysis, automation of employment, prediction and judgment, complementarity, role redefinition*

I. INTRODUCTION

Is the data scientist job category going to be replaced by automated data analysis is a particular example of the age-old question surrounding technology and work: Does technology kill the jobs it impacts? When it comes to data science, the question takes on a certain resonance: the technology automating the task is, after all, the data scientist's own creation. Data science driven by the practice of data science is the fruit of automated machine learning systems that can build and improve models with limited human effort (He et al., 2021). One might ask: With all these things automated, has the field become automated for practitioners, and will the role be around for another 10 years?

This paper makes two arguments: that the role has survived but is reinterpreted; that both arguments are important. The survival is true, too: automated analysis is not the end of the need for data scientists — automated analysis is not all that a data scientist does. The survival, however, is dependent upon re-identification, as the role remaining is not the same as the one that was there before. The parts of data science that are most impacted by automation are getting automated, and the human-impacted parts are getting human. The data scientist of the automated era is performing tasks and doing them in a different way than he or she did in the pre-automated era, and that's the kind of data scientist that's going to survive because he or she's willing to accept and adapt to that change rather than being a data scientist because the tasks being done are automated.

The argument is based on a series of research papers published between 2015 and 2023 in multiple disciplines. The economics of automation can be used to explain why automation of some tasks in a job does not destroy the job (Autor, 2015; Brynjolfsson & Mitchell, 2017; Frey & Osborne, 2017). Artificial intelligence economics helps to understand what types of human contributions will be more valuable in the future as machines gain more capabilities (Agrawal et al., 2019; Raisch & Krakowski, 2021). And commentary on the documents of the data science role shows that it has already evolved in the direction predicted (Davenport & Patil, 2022). The sections that follow combine these, creating a story of a role whose survival depends on change.

The Automation Anxiety and Its Long History

The fear of being replaced by an automated analysis tool by the data scientist is part of a long history of automation fears, and by situating it in that history, it becomes clear. After observing a century of predictions of technological unemployment returning for centuries, Autor (2015) began his analysis of workplace automation by noting that the displacement of human labor by machines has always been exaggerated. What he saw was that commentators became obsessed with the idea that automation is replacing work that computers can do, and neglected to consider how automation is enhancing human work, making human work more valuable, and creating new jobs. The fear of data science is one example of this common substitution mentality.

This is not to say that the anxiety is unfounded, and the more careful literature on automation has not taken the displacement of particular tasks seriously. In a highly publicized study (Frey and Osborne 2017), they calculated the vulnerability of hundreds of jobs to computers and found that a large percentage of jobs were at serious risk of being automated. Their approach, which is based on the extent to which tasks can be automated algorithmically, has implications for data science, which has more routine and well-defined tasks that are actually the kind their framework identifies as automatable. The lesson of Frey and Osborne (2017) is not that whole occupations disappear, but that occupations differ with respect to the automatability of the tasks that make them up, thereby focusing on the task composition of data science, not the role as an undifferentiated whole.

The history as it has been suggested above then indicates a path between extremes of obsolescence and untouched continuity. It's not that the data scientist is being phased out and out and out, as the substitution-fixated anxiety would suggest, but it's not that the role is being replaced, either, as complacent dismissal of the anxiety might suggest. Rather, it handles the automatable work, adds to the human role in the tasks that still require human attention, and thus redefines the role as per Autor (2015). However, the question of what will survive can only be answered if one considers what the automated analysis actually does and what it doesn't do – not by discussing the abstract of whether the analysis role will disappear.

What Automated Analysis Actually Automates

To assess whether or not a role will survive, the precision needs to be known about the tasks that automated analysis is used to accomplish, and the literature provides that. He et al. (2021) gave an overview of the current state of the art for automating the modeling pipeline, which required precious human effort in the past, such as data preparation, feature engineering, model architecture selection, and hyperparameter tuning. Their stated goal is to create such effective systems that they would not need so much human expertise, which they now claim is the exclusive province of the specialist. The jobs being automated are the technical, procedural aspects of building the models, which consume a lot of a data scientist's hands-on time.

The interesting thing about these automated tasks is that they have a common property: They are well-specified tasks with clear goals and well-structured inputs. In fact, automated machine learning works best when the problem is stated, the data is collected, and the objective is specified — in other words, when all the elements are contained within a certain range, finding a good model becomes a technical optimization that can be done with automation. This aligns with Brynjolfsson and Mitchell's (2017) analysis of conditions for effective machine learning, which focused on having well-defined inputs, clear objectives, and ample relevant data. The automatable part of data science is this area, and it is a large part, much of the model-building that was traditionally thought to be the soul of the craft.

However, the core that can be automated is bounded, and the boundaries define the areas where the human contribution remains. Automated systems should be told what the problem is; they optimize towards something that is already presented; they don't select interesting goals; they act upon the data that has been supplied to them, and do not decide whether the data that they are given is a good representation of the problem in question. All these limitations correspond to human tasks: problem specification, selection of an

DOI: https://doi.org/10.64388/IREV7I5-1722574

objective, and satisfactoriness judgment of data. In general, Brynjolfsson and Mitchell (2017) noted that jobs are made up of various tasks, some of which would be ideal candidates for machine learning, and others would not; so the question is not how automation is going to kill off most jobs, but how it is going to reshape most of them. The data science story is no exception: the core of the field—the part that involves executable data—is automatable, while everything else—the part that involves framing and judgment—is not.

The Task-Level View and Why the Role Persists

Evaluating the survival of the role isn't easy, so it is important to look at the data scientist's job as a collection of individual jobs. Brynjolfsson and Mitchell (2017) explicitly explained that the impact of machine learning on employment is more complicated than a replacement story: most jobs involve both kinds of tasks for which machine learning is appropriate and inappropriate, and machine learning, in general, tends to transform rather than eliminate jobs. When applied to data science, this task-level view takes away the false dichotomy between the role that lives and the role that dies. The role doesn't just get swallowed up or disappear; the tasks that are automatable get automated, and the tasks that are non-automatable make up the core of the role.

The reason this reframing is important is that the things that automated analysis can't do are not the fringe bits of data science—they are the most important. The good framing of the problem is critical to whether the analysis is directed towards the important issue. To decide whether a result is trustworthy or not is to decide whether it is relevant to a decision or not. The analysis can only be effective if it is incorporated into the organization's decision-making process. These tasks were always there in data science and always done by skilled practitioners but sometimes forgotten 'under the technical execution'. As these tasks are automated, they are not simply what's left, but are actually what's always mattered most about the task: human judgment.

The stability of the role, therefore, is a stability through recomposition. A data scientist whose job was just to perform the tasks that have been automated would be out of a job, and people who thought that their job was only about those tasks might be devalued. However, the role as defined is still alive; that is the part of human endeavours in making sound decisions based on data, which is not being performed by task automation. Instead of asking "Does the role exist?," ask "Do practitioners and the profession redefine the role around the role's enduring activities?," and the answer is in their power.

Prediction Automates, Judgment Remains

The economic discussion of AI clarifies the narrative of what remains and highlights the distinct value of human interventions in making AI more valuable. In recent times, advances in the field of artificial intelligence have expanded the capability of prediction, which is the generation of missing information from available information, while judgment, which evaluates the payoffs of different actions, is the task left for the human being to convert prediction into decision (Agrawal et al. 2019). Prediction and judgement go hand-in-hand, so that if prediction is cheapened by automation, the value of that which is scarce and human: judgement. This is an exact economic reason why the data scientist role will not just exist but evolve into a judgment role as automation increases.

There is a close connection to data science. In this context, much of what is generated by automated analysis is prediction – a model that predicts, classifies, or estimates. Automation of model-building is a natural extension of the automation of prediction production. However, this prediction is not the choice, and the transformation from prediction to choice demands the judgment that Agrawal et al. (2019) ascribe to the human side of the divide. The data scientist who is interpreting automated predictions and making judgments on their use, evaluating importance and consequences, making decisions when to believe and when to ignore model results, or when to override, is doing just the complementary function whose value increases when prediction is automated. The role lives by moving to this judgment, and the automation of prediction makes it more valuable, not less.

Raisch and Krakowski (2021) agreed with this, by demonstrating that automation and augmentation are not discrete, but that automation can be augmentation. The automated predictive core of data science does not take the place of human labor; it merely changes what humans do — it increases the surrounding judgment to which it is complementary, in a perpetual interaction between machine and human performance. This interdependence implies that the higher the level of automation of the predictive execution, the more the human role will be subject to the judgment complementing it, and, therefore, advanced automation and a judgment-based role are not incompatible. That's part of the reason the data scientist position is still alive with automated analysis: because automated analysis is making it more about judgment.

The Role Redefined in Practice

This idea that the role continues to exist in a new guise is not just an assumption, but something that is evident in the evolution of the data science role in practice. Ten years later, Davenport and Patil (2022) revisited the idea that the data science job had become even more attractive and sought to determine whether it was still as popular and necessary. They noted that the role had become more institutionalized, that the job was no longer just about the technical aspects, but that other aspects of the job, such as the management of the analysis within an organization, including communicating and ethics, had become more significant and that much of the purely technical work had been absorbed or supplemented by tools and other work roles that had been adjacent. This is exactly the shift to judgment, framing, and integration into organizations that this argument predicts will be the redefinition of how the technical is executed.

Their insights that groups are looking for more diverse teams of data scientists than the individual technical master are particularly salient. In times when the importance of data science was limited to technical execution, the model was to hire highly talented people who could do all the work. As technical execution is automated and more accessible, the model moves toward teams where technical execution is automated along with human framing, judgment, and integration that the automation does not give. The job exists, but the practice shifts to human-centered work and the person data scientist becomes more and more attractive for human activities that can't be automated by analysis.

This realistic reframing is also a response to a natural question: If automated analysis works so well, do organizations need more or fewer data scientists? The solution is that automation of the technical core does not diminish, but transforms, the need for the human contribution, and can even increase it by putting analysis within reach in more contexts. From the general case, as Autor (2015) pointed out, automation of certain tasks can also increase the need for complementary tasks that must be performed by humans, and can lead to a net increase in overall employment in an occupation, as the added automation capability opens up new opportunities that need to be framed and judged by humans. This is the trend of complementarity — and not full substitution — that Davenport and Patil (2022) observed as data scientists continued to be in high demand despite automation of technical skills.

Complementarity and the Conditions of Survival

While the need for the data scientist is solidly based, it is not guaranteed, and it is necessary to set a condition if one is to be frank and helpful. The role does continue to exist if and only if the human tasks in the role remain truly complementary to the automated tasks and truly cannot be automated. The argument for survival would have been less if framing, judgment, and integration were themselves easily automatable. They are not automatable because they involve such human capacities as understanding of the organizational context that cannot be captured in a data set; exercising accountable judgment about consequences; integrating analysis into human decision processes, all of which are the reasons mentioned throughout this paper that make them persistently beyond the capabilities of machines.

Survival also depends on the profession's response; hence, the redefinition is not automatic. The field that bases its definition of competence on rote-learnable technical tasks, which trains and rewards its members mainly for those tasks, and which refuses to give way to the judgment and integration of the role, will be affected by the displacement of its members by automation, even if the role continues as a redefined one. No one can say the role has been taken up by someone else, unless that someone moves with it. As with everything else, a profession that focuses exclusively on executing or a profession that relies

© NOV 2023 | IRE Journals | Volume 7 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV7I5-1722574

entirely on automation does not do well, whereas a profession that focuses on augmenting the human contribution does well, as both Raisch and Krakowski (2021) warn. This preservation of the role is linked to the survival of the practitioners but is different, and the survival of the practitioners relies on their willingness to accept the redefinition the role will undergo.

Last, the survival depends on retaining a sufficient level of technical expertise to be able to carry out the human tasks with proficiency. If a person is not familiar with the automated analysis that he/she is framing, judging, and overseeing, then he/she cannot be competent in performing the framing, judging, and overseeing. The technical knowledge is not eliminated from the job but rather transferred from an end to a basis for judgment. Shrestha et al. (2021) highlighted that their outputs need to be interpreted by humans, and that humans must have sufficient technical knowledge to interpret outputs from automated analytical systems to be able to extract value from them. A data scientist is not a non-technical role; it is a role that has been established around a technical base and that requires judgment and integration.

The Economics of Displacement and Reinstatement

The claim that the data scientist role survives through recomposition rather than elimination is grounded in a broader economic account of how automation affects labor. Acemoglu and Restrepo (2019) framed the effect of automation in terms of two opposing forces: a displacement effect, by which automation removes labor from tasks it previously performed, and a reinstatement effect, by which the creation of new tasks restores labor demand in activities where humans hold a comparative advantage. On this account, automation does not simply subtract jobs; it shifts the task content of production, and the net effect on any occupation depends on whether the reinstatement of new tasks keeps pace with the displacement of old ones. Applied to data science, this framework predicts exactly the recompositing argued for here: automated analysis displaces the executable tasks while new tasks of framing, judgment, and governance are reinstated as the human core of the role.

The importance of analyzing automation at the level of tasks rather than whole occupations is underscored by work that revisited earlier, more alarming estimates. Arntz et al. (2017) showed that occupation-level estimates of automation risk substantially overstate the danger, because they treat entire occupations as automatable when in fact occupations comprise bundles of tasks, only some of which can be automated. When the analysis is conducted at the task level, the share of work at high risk of automation falls sharply. This finding directly supports the survival argument advanced here: the data scientist role is not a monolith that automation either eliminates or spares, but a bundle of tasks whose automatable components are automated while its non-automatable components persist and define the recomposed role.

History offers a further reason to doubt that automating the technical core of data science eliminates the occupation. Bessen (2019), examining two centuries of automation in manufacturing, showed that productivity-enhancing technology often increased rather than decreased employment in an industry, because more efficient production expanded demand, and that employment declined only later when demand became satiated. He also emphasized that even where automation reshapes an industry, it typically requires workers to acquire new skills and move into new tasks rather than rendering them simply redundant. For data science, this suggests both that the automation of analytical work may expand rather than contract the demand for data professionals by making analysis feasible in more contexts, and that the workers who thrive will be those who make the transition to the new, non-automated tasks.

New Tasks and the Division of Labor Between Humans and Machines

A task-level account of which parts of data work automation claims, and in what order, further clarifies what survives. Huang and Rust (2018) proposed that the intelligences required for work can be distinguished into mechanical, analytical, intuitive, and empathetic, and argued that AI tends to acquire these in order, mastering mechanical and analytical tasks before intuitive and empathetic ones. They further argued that firms should decide the division of labor between humans and machines at the level of tasks rather than jobs, replacing humans in the tasks machines perform well and redirecting human effort toward the tasks that require higher intelligence. For data science, this implies that the analytical execution most susceptible to automation is taken first, while the intuitive judgment and contextual reasoning that surround it remain human, which is precisely the recompositing of the role toward its higher-intelligence tasks.

Evidence on how organizations actually deploy AI supports this measured, task-level picture rather than the wholesale-replacement fear. Davenport and Ronanki (2018), surveying real-world corporate adoption of AI, found that most deployments automated specific tasks or assisted human workers rather than replacing entire roles, and that the more transformative ambitions frequently outran what the technology could reliably deliver. Their account suggests that automated analysis, in practice, takes over particular tasks within data work while leaving the professional to perform and govern the rest, which is consistent with a role that persists in redefined form rather than one that disappears.

Finally, automation does not only subtract tasks; it generates new ones, including roles that did not previously exist. Wilson et al. (2017) argued that artificial intelligence is creating entire categories of new work, among them roles dedicated to training, explaining, and sustaining AI systems, that did not exist before the technology's spread. Faraj et al. (2018) complemented this by showing that learning algorithms reshape occupational boundaries, dissolving some roles while giving rise to others organized around the mediation of algorithmic work. For the data scientist, these accounts imply that the automation of analytical execution is accompanied by the emergence of new tasks and roles, many of them concerned with directing and governing the automated analysis, so that the occupation is not merely preserved but partly reconstituted around work that the rise of automated analysis itself creates.

CONCLUSION

Even as automated analysis has emerged, the role of the data scientist remains, but in a different form than it used to be. Automated analysis focuses on the executable and well-specified aspects of data science, the model construction, tuning, and regular analytics that automated machine learning systems are increasingly able to do, while reserving problem formulation, judgment, and decision making for humans. Since a job is a collection of tasks and not a single task, automating some of these tasks doesn't remove the job, and rather, it recomposes it (Brynjolfsson & Mitchell, 2017), moving the human contribution to tasks that machines cannot do.

Not marginal but central tasks are retained, and according to the logic of complementarity, such retained tasks are becoming more valuable as the tasks that can be automated become automated. The economics of AI sees value in judgment as the transition from automated prediction to action or decision (Agrawal et al., 2019), and the interdependency between automation and augmentation implies that automation of the predictive core reinstates the role of the human in judgment (Raisch & Krakowski, 2021). The history of automation implies that what the machines can't do becomes more valuable, and perhaps even more demanded as the tasks they can do get automated (Autor, 2015; Frey & Osborne, 2017). The data science role has been observed to shift from prediction and redefining towards judgment, framing, and integration into the organization as predicted (Davenport & Patil, 2022).

So, does the data scientist role exist? The answer is = yes, but it has been transformed. Its life depends on the human tasks it still performs being truly resistant to automation, which seems to be the case, and on the profession redefining competence about the human tasks, and not the automated ones; and on practitioners continuing to have the technical base of competent judgment, and not automation. The data scientist of the era of automated analysis is not a mere executor of technical analysis, but a creator, evaluator, and user of analysis which is being increasingly automated by machines. The role that vanishes is the narrow one that is performed in an automatable way. The one that remains is the one that automated analysis teaches us to care for more, not less, because of human judgement.

REFERENCES

[1] Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of*

Economic Perspectives, 33(2), 3–30. https://doi.org/10.1257/jep.33.2.3

[2] Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy, 47*, 1–6. https://doi.org/10.1016/j.infoecopol.2019.05.001

[3] Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters, 159*, 157–160. https://doi.org/10.1016/j.econlet.2017.07.001

[4] Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

[5] Bessen, J. (2019). Automation and jobs: When technology boosts employment. *Economic Policy, 34*(100), 589–626. https://doi.org/10.1093/epolic/eiaa001

[6] Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science, 358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

[7] Davenport, T. H., & Patil, D. J. (2022, July–August). Is data scientist still the sexiest job of the 21st century? *Harvard Business Review*. https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century

[8] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

[9] Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

[10] Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

[11] He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems, 212*, Article 106622. https://doi.org/10.1016/j.knosys.2020.106622

[12] Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459

[13] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

[14] Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

[15] Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research, 123*, 588–603. https://doi.org/10.1016/j.jbusres.2020.09.068

[16] Wilson, H. J., Daugherty, P. R., & Morini-Bianzino, N. (2017). The jobs that artificial intelligence will create. *MIT Sloan Management Review, 58*(4), 14–16.