Type of article: Opinion

From 'Sexiest Job' to 'Most Responsible Role': The Evolution of Data Scientists

Abstract

This perspective article explores the evolving responsibilities of data scientists in a data-driven society where ethical, privacy, and governance standards are vital. Initially celebrated for its appeal and high earning potential, the data science profession has become one of the most responsible roles of the 21st century, requiring practitioners to balance their technical skills with a commitment to social impact and accountability. This article further examines responsible data scientists' essential qualifications and criteria, including ethical awareness, privacy safeguards, transparency, and continuous learning. Additionally, this article discusses hiring practices that prioritize these qualities and outlines strategies for fostering a data-driven culture grounded in responsibility and trust. In the current landscape, responsible data scientists are not merely analysts but also protect ethical data practices, which is crucial to building a transparent, fair, and sustainable digital world. The article also provides a framework and guidelines for identifying and recruiting responsible data scientists.

Keywords: *Responsible Data Science; Ethics; Privacy; Accountability; Governance*

1. Introduction

In the data-driven world, the role of a data scientist has evolved far beyond what Thomas H. Davenport and D.J. Patil described in their famous article, *"Data Scientist: The Sexiest Job of the 21st Century"* (Davenport and Patil 2012). While this pioneering article highlighted data science's allure and earning potential, its focus on fame, opportunity, and influence may have downplayed the critical ethical and societal responsibilities that data scientists carry today. Today, data science is increasingly recognized as one of the most responsible jobs in the modern world, a shift driven by mounting concerns around responsible artificial intelligence (AI), data ethics, privacy, and governance (Arrieta, Díaz-Rodríguez et al. 2020). A reflection on the current state of data science reveals that the title *"Data Scientist"* carries a weight far beyond the initial glamour of wealth and status, demanding a redefined understanding of what it means to be a responsible data scientist. As we advance, it is time to shift from embellishing the job title to recognizing the role as *"The Most Responsible Job of the 21st Century,"* where data scientists are the stewards of trust, ethics, governance, and privacy.

2. From "sexiest" to "responsible data scientist"

2.1. The responsibility revolution in data science

The influence of data scientists on our lives is profound, shaping decision-making in healthcare, finance, education, governance, etc. With this impact comes a tremendous responsibility to uphold ethical standards and transparency in every project. While the original article emphasized data scientists' roles in handling big data, it omitted imperious responsibility, particularly concerning AI ethics, data governance, and privacy, which are aspects that form the backbone of responsible data science today (Davenport and Patil 2012). The original article characterized data scientists as professionals equipped with massive datasets, extracting actionable insights from complex, unstructured information. This emphasis on innovation, however, overlooked a crucial aspect: the social and ethical implications of these insights. The landscape of data science has since evolved, including the role of a data scientist. Today's responsible AI and data governance form the bedrock of any robust data-driven solution. Ethical decision-making, privacy preservation, and compliance with data regulations are not mere add-ons but rather essential aspects of a data scientist's daily practice (Boppiniti 2023, Stahl 2023). As data scientists increasingly influence critical decisions in fields as diverse as healthcare, finance, criminal justice, and public policy, their work has significant societal implications. This

shift mandates a broader understanding of their responsibility, transforming data science from the *"sexiest"* to the *"most responsible"* job of the 21st century.

2.2. Ethics, privacy, and governance: the foundations of responsible data science

Data science has rapidly shifted to include policies on responsible AI, ethics, governance, and privacy. As data scientists, we must now consider the ethical ramifications of each insight or recommendation. For example, facial recognition technology has increased. What once appeared as a promising computer vision application raises alarms about privacy invasions and biases if implemented uncritically. Responsible data scientists, therefore, must champion a design approach that prioritizes fairness, accountability, and transparency in model development and application (Floridi and Taddeo 2016).

The responsibilities of a data scientist now demand a heightened awareness of privacy implications, such as those outlined in regulations such as the General Data Protection Regulation (GDPR) (EU 2023). Without privacy measures, sensitive data becomes a potential liability, risking user trust and legal consequences. Data governance frameworks ensure data are used, stored, and shared responsibly, promoting accuracy, accountability, and ethical stewardship (Janssen, Brous et al. 2020). This is essential, especially in sectors where decisions such as healthcare, finance, or criminal justice can profoundly affect individuals' lives.

Although “responsible” should have been integral to designing and developing data science solutions, it is never too late to realize the ignorance and start embedding it in our DNA. In recent years, a growing emphasis on responsible AI has highlighted the need for data scientists to integrate ethics into their practice (Gianni, Lehtinen et al. 2022). Unlike a decade ago, when data science was centered primarily around technical skills, today's data science landscape requires professionals to anticipate and mitigate unintended consequences from data-driven solutions. Bias in algorithms, data privacy concerns, and ethical accountability are no longer abstract considerations; they are substantial issues with real-world repercussions (Ferrara 2023). This awareness has led to ethical frameworks for AI and data science, with organizations such as the IEEE and the European Union establishing guidelines for AI ethics and data protection (Chatila and Havens 2019, EU 2024). These frameworks underscore that data scientists are not just analysts but also stewards of data charged with the ethical handling and responsible use of information. They are expected to uphold standards that protect user privacy, prevent bias, and

ensure transparency in algorithms and data science product development (Chatila and Havens 2019).

2.3. *The European Union (EU) AI Act and Other Frameworks*

The European Union AI Act 2024 emphasizes the risks associated with AI systems and how individuals and entities involved in AI system development, deployment, and management, including providers, deployers, product manufacturers, authorized representatives, importers, and distributors, should take responsibility for their actions (EU 2024). It also emphasizes various responsibilities and obligations to ensure responsible AI systems through articles (Figure 1) (EU 2024).

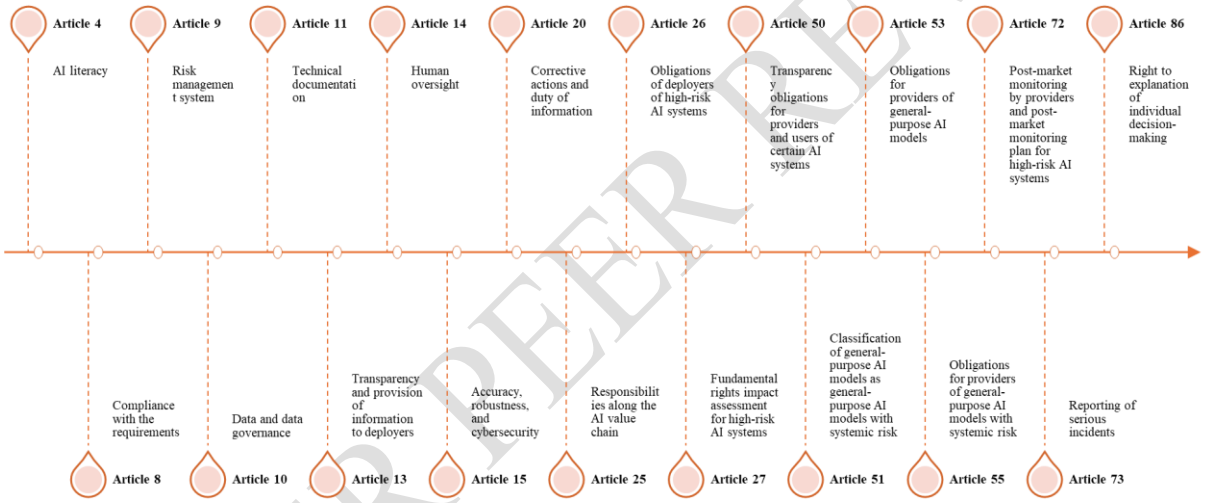

Figure 1. Articles covering responsibilities for making AI systems responsible.

Similarly, the Artificial Intelligence Risk Management Framework (AI RMF 1.0) also highlights the responsible development of AI systems to reduce risks to humanity by promoting responsible design, development, and deployment practices (Tabassi 2023).

2.4. *The misalignment in data science training and the "data scientist tag"*

As data science gained popularity as one of the highest-paying jobs globally, institutions mushroomed, promising students' entry into this lucrative field. However, these programs prioritize technical skills over ethical considerations, governance, and responsible AI principles. Consequently, many may carry the "Data Scientist" title without fully embracing its

responsibilities. Although morality and responsibility are integral to culture and upbringing, teaching these values in nonfamily settings is also critical (Bentahila, Fontaine et al. 2021). Therefore, given the role's influence on decision-making and societal impact, certification or a degree in data science must transcend technical competency to include rigorous training in ethical principles, governance structures, and privacy safeguards (Yakowitz 2011). In addition to acquiring technical proficiency in programming, data manipulation, and statistical modeling, data scientists must be prepared to confront questions about bias, privacy, and transparency.

2.5. Ensuring integrity in data science solutions

Responsible data scientists acknowledge that their algorithms influence real people and should operate with an obligation to "not harm" that requires careful assessment, validation, and transparency in all stages of model development. The role is no longer about what a data scientist "can" do with big data; it is about what they "should" do. A famous example is social media's role in political discourse, with claims from all sides that Twitter's (now X) algorithms either amplify opponents or suppress their voices. This has led policymakers and researchers to call for greater transparency in how algorithms shape exposure to political content (Huszár, Ktena et al. 2022).

Moreover, transparency and interpretability in AI models are not just technical features but also ethical requirements. Providing clear explanations for model predictions enables accountability and builds trust with end-users. Transparency plays a crucial role in addressing a fundamental ethical concern in AI: enabling affected stakeholders to comprehend how decisions are made. This is particularly crucial in high-stakes applications such as credit scoring or medical diagnosis (Thunki, Reddy et al. 2021).

Thus, the following question arises: what does being a "responsible data scientist" mean?

Responsible data science is about making ethical judgments that prioritize the well-being of individuals and society. Technical skillfulness is no longer sufficient; data scientists must also be committed to ethical principles. For example, a responsible data scientist should ensure that algorithms do not inadvertently perpetuate discrimination or bias. They must proactively identify biases within datasets and commit to developing transparent models that stakeholders and the public can understand and trust. Additionally, they are expected not to introduce personal bias and maintain user privacy by anonymizing data, securing sensitive information, and ensuring that

data collection practices comply with regulations such as the GDPR and must adhere to the EU AI Act and other AI policies applicable in their regions or industries (EU 2023, EU 2024). In short, responsible data scientists serve as guardians of ethical AI, striving to balance innovation with accountability.

2.6. The future: data science as a steward of public trust

The field of data science holds immense potential to improve society, from diagnosing diseases earlier to creating more equitable educational systems. However, with this potential comes the responsibility of building trust and safeguarding the integrity of the data we work with (Li, Qi et al. 2023). For data scientists, the shift toward a responsibility-oriented mindset is paramount. Adopting a holistic approach that integrates ethics, privacy, and governance ensures that data scientists are proficient in technology and committed to society's well-being.

To embrace this paradigm shift, data science training programs must expand their curriculum to instill these principles, turning out professionals who are more than just technically adept. They should be advocates for ethical data use, be equipped to navigate complex moral landscapes and be capable of contributing meaningfully to an equitable future.

In a world increasingly reliant on data-driven decisions, responsible data scientists are the gatekeepers of our digital society. Therefore, the next generation of data scientists should be guided by a commitment to responsibility, integrity, and transparency, elevating the profession beyond "sexy" and into the realm of the genuinely consequential.

2.7. Are we cultivating a culture of responsible data science?

Despite the growing emphasis on ethics and responsibility, it remains an open question whether data science professionals embody these values. The proliferation of individuals with the "Data Scientist" title does not necessarily correlate with adherence to responsible practices. Without formal training in ethics and privacy, many self-taught or certificate-trained data scientists may lack the knowledge required to navigate ethical dilemmas effectively. This gap presents a challenge to the field: while data science is becoming increasingly democratized, it is imperative to ensure that professionals holding this title understand their responsibilities beyond the technical domain. Encouraging a culture of accountability among data scientists is essential for professional integrity and long-term viability.

3. Building a framework for responsibility in data science

The quote by Hosmer and Lemeshow, *"It is the analyst, not the computer, who is ultimately responsible for the review and evaluation of the statistical model,"* remains highly relevant today, emphasizing that human judgment, not just tools, is essential in the analysis process (Hosmer Jr, Lemeshow et al. 2013). As data science matures, the future lies in cultivating a workforce that embraces responsibility as a core value. This requires a collective effort from academic institutions, professional organizations, and individual data scientists. Although the sense of ethics and responsibility are part of the upbringing of individuals, educational programs must integrate comprehensive ethics training to prepare students for the moral complexities of their future careers. Second, professional bodies should advocate for industry-wide standards that promote responsible AI, establishing accountability mechanisms to guide ethical conduct. Finally, individual data scientists must recognize that their role transcends data manipulation; they are entrusted with shaping the future of AI in ways that honor societal values and protect the public good.

3.1. How do we find responsible data scientists?

Finding responsible data scientists is not easy; therefore, there could be several ways to consider when recruiting them. Some of the critical aspects or criteria required to hire responsible data scientists are depicted in Table 1.

Table 1. Where and how to find responsible data scientists

Sl. No.	Practice	Description
1	Clear role requirements	Hiring authorities or teams should specify the ethical and governance responsibilities expected of the role, such as handling sensitive data, compliance initiatives, or responsible AI practices.
2	Beyond technical skills	Hiring authorities or teams should evaluate candidates' understanding of ethical implications in data science, including bias mitigation, fairness, and privacy, to ensure responsible and trustworthy practices.
3	Transparent communication	Hiring authorities or teams should seek candidates who communicate decisions, including ethical implications, risks, and limitations, to maintain accountability and foster a data-informed culture.

4	Experience with responsible tools and technologies	Hiring authorities or teams should identify candidates familiar with tools for responsible data handling, such as anonymization techniques or differential privacy, reflecting a commitment to ethical practices.
5	Practical assessments and case studies	Hiring authorities or teams should include ethical or privacy-sensitive scenarios in assessments to gauge candidates' ability to balance insights with responsible decision-making.
6	Domain knowledge	Hiring authorities or teams should prioritize candidates who understand the nuances of the domain, specifically in life sciences and healthcare and associated regulatory and governance standards, to support privacy and data protection.
7	Referrals for ethically minded candidates	Hiring authorities or teams should encourage referrals of data scientists known for ethical decision-making, as they likely share the organization's values around privacy and governance.
8	Data science conferences with a responsible AI focus	Hiring authorities or teams should participate in responsible AI, ethics, and data privacy events to meet candidates who value and understand responsible data practices.
9	Responsible data science specialists	Hiring authorities or teams should target responsible data science specialists for specialized needs and may hire freelancers with expertise in governance or privacy to enhance the team's commitment to responsible practices.
10	Responsible data science culture	It is imperative to nurture a culture that prioritizes data ethics, responsible AI, and privacy, helping data scientists thrive in an environment that values ethical data use, and top leadership can play a significant role in achieving this.

3.2. Who are not responsible data scientists?

Similarly, it is imperative to understand who are not responsible data scientists so that screening potential candidates becomes more manageable (Table 2).

Table 2. A summary of those who are not responsible for data scientists

Sl. No.	**Characteristics**	**Description**
1	Motivated solely by name and fame	Candidates primarily driven by prestige and financial gain rather than a commitment to ethical, accurate, and impactful data science are not ideal for responsible data scientist roles.
2	Hired primarily through quotas without skills	Individuals brought into roles based on reservation quotas or diversity targets without necessary background or skills, potentially impacting project quality and ethical standards, may

		not be a good fit for responsible data scientist positions.
3	Lack of ethical awareness	Candidates failing to prioritize ethical considerations leads to biased analyses or decisions with potentially harmful consequences and are not fit for responsible data scientist roles.
4	Disregard for privacy	Professionals who overlook data privacy principles, mishandle sensitive information, and risk individual privacy and organizational compliance may not be ideal for the responsible data scientist role.
5	Limited transparency	Data scientists who do not openly communicate model limitations, potential biases, or data-related decisions, reducing accountability and stakeholder trust, therefore, are not a good fit for responsible data scientist roles.
6	Ignorance of governance policies	It is essential to be aware of governance policies relevant to data and AI systems; therefore, those who lack awareness or disregard data governance practices risk data security, compliance, and reliability.
7	Prioritizing speed over accuracy	Data scientists prioritize rapid results over the accuracy, robustness, and quality of insights, leading to unreliable and potentially damaging outcomes, and are not ideal for responsible data scientist roles.
8	Insensitive to bias in data or models	Those data science professionals who fail to identify or address biases within data or models, resulting in unfair or discriminatory conclusions, are not suitable for this responsibility.
9	Unaccountable for model impact	Although it would be a topic for debate, it is essential that data scientists should not avoid taking responsibility for the social, ethical, and real-world consequences of models, especially if outcomes are adverse.
10	Focus solely on technical skills	While hiring data scientists, recruiters should avoid hiring those who emphasize technical skills without considering their work's broader social, ethical, and regulatory implications.
11	Lack of continuous learning on ethics and privacy	Continuous learning is critical in becoming relevant and aware of the social, environmental, and policy impacts of the developed data science solutions; therefore, recruiters should discourage candidates who do not stay updated on evolving ethical, legal, and governance standards essential for responsible data science.

Hence, data science may have initially been celebrated as the "sexiest" job of the 21st century, but it is now clear that its value lies in its capacity to responsibly address society's most pressing challenges. By prioritizing ethics, governance, privacy, and accountability, the data

science community can transform the field into one of the most responsible and impactful professions of our time. Let us redefine success in data science not by fame or financial gain but by our commitment to ethical practices and our dedication to the betterment of society. However, challenges are associated with identifying, nurturing, and recruiting responsible data scientists.

4. Challenges and barriers to implementing responsible data science

Implementing responsible data science faces challenges even in well-resourced settings. These include the complexity of balancing innovation with ethical considerations, navigating regulatory landscapes, and ensuring stakeholder alignment on privacy and governance principles. The rapid pace of technological advancements often outpaces the development of frameworks for accountability and fairness, leaving gaps in addressing algorithmic bias and transparency (Binns 2018). Additionally, resistance to change within organizations and a lack of awareness about ethical implications hinder the adoption of responsible practices.

These challenges are magnified in low-resource settings due to limited infrastructure, resources, and expertise. These settings often lack the necessary technological tools and reliable data infrastructure, hindering ethical and privacy-preserving practices (Jejeniwa, Mhlongo et al. 2024). Moreover, access to skilled professionals who understand and apply governance, transparency, and ethical standards is often scarce, leaving organizations vulnerable to biased or incomplete data analyses. Another significant barrier is the financial constraint that limits investment in advanced technologies, such as privacy-preserving tools or bias-detection algorithms, often seen as expensive luxuries rather than necessities (Binns 2018). Additionally, the absence of robust legal and regulatory frameworks in many low-resource settings makes it challenging to enforce data governance or hold organizations accountable for unethical practices. Cultural and contextual factors further complicate responsible data science implementation. For instance, different societal norms around privacy and data ownership can lead to misaligned priorities. At the same time, a lack of stakeholder awareness about ethical data practices can result in resistance or non-compliance (Milan and Velden 2016). Addressing these challenges requires tailored strategies considering the socioeconomic and cultural contexts of low-resource settings.

5. Conclusion

In conclusion, the role of a data scientist goes far beyond a "sexy" job title—it is a profession of immense responsibility that profoundly influences society. While technical expertise is vital, the true value of a data scientist lies in their commitment to ethics, accountability, and societal well-being. Prioritizing responsible AI, data governance, privacy, and transparency ensures that data science becomes a force for solving real-world challenges. Organizations must focus on hiring data scientists who integrate these principles into their work, fostering a culture of ethics and accountability. By doing so, we can shape a future where data science drives innovation while upholding its responsibility to society.

Competing Interests Disclaimer

The author has declared that they have no known competing financial interests, non-financial interests, OR personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author wishes to acknowledge all the data science practitioners who do their job responsibly. The author thanks his colleagues and friends for reinforcing the belief that responsible data scientists are essential and that ethical considerations should be a prerequisite in hiring any data scientist. The author also declares that the views presented in this paper are personal and do not reflect the view of the institutions he is associated with.

Disclaimer (Artificial Intelligence)

Author(s) hereby declares that generative AI technologies, such as Large Language Models, etc., have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1. The author acknowledges that GPT 4o was used to reduce unintentional bias in the arguments in this article and create tables (1-2) through prompt engineering: texts were provided to GPT 4o.

References

Arrieta, A. B., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina and R. Benjamins (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information fusion **58**: 82-115.

Bentahila, L., R. Fontaine and V. Pennequin (2021). "Universality and cultural diversity in moral reasoning and judgment." Frontiers in Psychology **12**: 764360.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Conference on fairness, accountability and transparency, PMLR.

Boppiniti, S. T. (2023). "Data Ethics in AI: Addressing Challenges in Machine Learning and Data Governance for Responsible Data Science." International Scientific Journal for Research **5**(5).

Chatila, R. and J. C. Havens (2019). "The IEEE global initiative on ethics of autonomous and intelligent systems." Robotics and well-being: 11-16.

Davenport, T. H. and D. Patil (2012). "Data scientist: The Sexiest Job of the 21st Century." Harvard business review **90**(5): 70-76.

EU. (2023). "Complete guide to GDPR compliance." from https://gdpr.eu/.

EU (2024). EU AI Act 2024. 2024/1689. T. E. Union, Official Journal of the European Union.

Ferrara, E. (2023). "Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies." Sci **6**(1): 3.

Floridi, L. and M. Taddeo (2016). What is data ethics?, The Royal Society. **374:** 20160360.

Gianni, R., S. Lehtinen and M. Nieminen (2022). "Governance of responsible AI: From ethical guidelines to cooperative policies." Frontiers in Computer Science **4**: 873437.

Hosmer Jr, D. W., S. Lemeshow and R. X. Sturdivant (2013). Applied logistic regression, John Wiley & Sons.

Huszár, F., S. I. Ktena, C. O'Brien, L. Belli, A. Schlaikjer and M. Hardt (2022). "Algorithmic amplification of politics on Twitter." Proceedings of the National Academy of Sciences **119**(1): e2025334119.

Janssen, M., P. Brous, E. Estevez, L. S. Barbosa and T. Janowski (2020). "Data governance: Organizing data for trustworthy Artificial Intelligence." Government information quarterly **37**(3): 101493.

Jejeniwa, T. O., N. Z. Mhlongo and T. O. Jejeniwa (2024). "AI solutions for developmental economics: opportunities and challenges in financial inclusion and poverty alleviation." International Journal of Advanced Economics **6**(4): 108-123.

Li, B., P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi and B. Zhou (2023). "Trustworthy AI: From principles to practices." ACM Computing Surveys **55**(9): 1-46.

Milan, S. and L. v. d. Velden (2016). "The alternative epistemologies of data activism." Digital culture & society **2**(2): 57-74.

Stahl, B. C. (2023). "Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems." Scientific Reports **13**(1): 7586.

Tabassi, E. (2023). "Artificial intelligence risk management framework (AI RMF 1.0)."

Thunki, P., S. R. B. Reddy, M. Raparthi, S. Maruthi, S. B. Dodda and P. Ravichandran (2021). "Explainable AI in Data Science-Enhancing Model Interpretability and Transparency." African Journal of Artificial Intelligence and Sustainable Development **1**(1): 1-8.

Yakowitz, J. (2011). "Tragedy of the data commons." Harv. JL & Tech. **25**: 1.