

Paper PP01

The Evolving Role of Statistical Programmers in the Pharmaceutical Industry: Impacts of Artificial Intelligence

Cindy Marabotti, Sanofi, France

Abstract:

In the pharmaceutical industry, the role of the statistical programmer is critical to transforming clinical trial data into reliable insights thus facilitate quality evidence-based decision making. Indeed, statistical programmers are responsible for creating and providing statistical results at each interim decision-making point, as well as for the conclusion of each clinical trial. They are also responsible for electronic submission of data (and statistical results) to regulatory authorities.

With rapid changes in technologies (multilingual, visualization) and the regulatory landscape, this role is already undergoing significant changes. The emergence of artificial intelligence tools has the potential to further transform this job by automating some of his tasks. This should result in increased efficiency and speed. However, it is important to recognize that these tools also present potential limitations and risks.

In this poster, I will explore the impact of artificial intelligence on the role of a statistical programmer (will still this name?). I will discuss the potential benefits of these tools, such as improving data quality, accelerating timelines, improving productivity or well-being, as well as their opportunities, potential limitations, and risks (destroying jobs, carbon footprint). Then, I will highlight the importance and emergency of upskilling for statistical programmers and the role of managers in supporting their teams through this evolution. Despite these changes, the role of statistical programmer will continue to be critical to ensuring the accuracy and reliability of clinical trial analysis.

INTRODUCTION

In the pharmaceutical industry, the role of the statistical programmer is critical to transforming clinical trial data into reliable insights thus facilitate quality evidence-based decision making. Indeed, statistical programmers are responsible for creating and providing statistical results at each interim decision-making point, as well as for the conclusion of each clinical trial. They are also responsible for electronic submission of data (and statistical results) to regulatory authorities. With rapid changes in technologies (multilingual, visualization) and the regulatory landscape, this role is already undergoing significant changes.

In this paper/poster, I briefly expose the evolution of this job over the years and explain how the emergence of the Artificial intelligence rescans the cards and can help the statistical programmer in this new adventure to support the development of new drugs and treatments. Indeed, it remains crucial to support statistician and let offer this latter focuses on statistical innovations, clinical development strategy, results interpretation and decision making.

SOME BACKGROUND IN THE EVOLUTION OF THIS CLINICAL DEVELOPMENT AND IN THE JOB OF STATISTICAL PROGRAMMER

Over the years, we have already seen significant changes of this job. Indeed, Clinical trials are evolving rapidly, with more focus than ever on time and budget.

A few years ago, global standards were still in transition and few agencies (mostly FDA and PMDA) asked to comply with them (not to mention that exceptions could be obtained). We also had a very limited number of types of drugs: mainly biotechnology, and the drug development process followed a sequential model (phases I to II to III).

Nowadays:

- Data standards (CDISC) are required by multiple agencies (not just the FDA and PMDA), and exceptions are no longer required. To submit a dossier, more and more documents are to be provided in addition to the data (such as xml, reviewer guides, programs), with the constant objectives to prove the traceability and quality of the data
- There is an evolution of the requirements of health agencies: acceleration of regulatory requirements (such as on safety topic or on the use of RWE data or the Optimus project), reinforcement and clarification of requests (such as on GxP recommendations or the Accumulus project), reduction of submission timelines (RTOR and/or Orbis projects), openness to sending programs in languages other than SAS (such as R) and data visualization (such as R Shiny)
- In term of drugs, new technologies are available (such as antibodies: mono, bi-specific, multi; proteins; nanobodies; enzymes; genetic therapy; RNA; vaccines)
- New technologies are now available to support data collection (devices, wearables, RWE, decentralization); that lead to an increase of partnerships between pharma industries and across other companies (such as AWS)
- The drug development cycle no longer follows a sequential model but a more complex one, to match with innovative statistics approaches (Umbrella, Platform, Basket, adaptive design) and in order to provide better answers to scientific questions and accelerate decision making.
- Finally, new types of parameters to be analyzed (such as biomarkers) appear.

All these changes have required (and continue to require) that the statistical programmer adapt, gain in efficiency and find new alternatives (e.g. to spend less time behind his screen to code). These changes have allowed statistical programmers to have more responsibilities, for example by leading electronic data submission.

THE EMERGENCE OF THE ARTIFICIAL INTELLIGENCE

Who has never heard of GPT Chat or even tested/used it? Not a day goes by without hearing about it!

The impact of artificial intelligence on our daily tasks cannot be ignored! (Especially Natural Language Processing or more specifically the Large Language Model). Indeed, the advent of artificial intelligence (AI) has introduced new opportunities (and challenges) to help us in the race of clinical trial evolution (acceleration). The introduction of tools (such as generative AI in open source: pre-trained model that we can refine) has the potential to automate some of the tasks performed by statistical programmers. Does this mean the death of this job? I don't think so!

Indeed, while AI has the potential to automate certain aspects of statistical programming and increase efficiency, it cannot replace the domain expertise and contextual understanding that statistical programmers bring. The generated content should still be validated by an expert. Even if the AI will not replace the statistical programmers, the role of statistical programmers must evolve with AI integration, requiring them to adapt/enhance some different set of skills and acquire new ones to

utilize these tools effectively. By embracing these skills and actively contributing to the data analysis process, statistical programmers can continue to play a valuable role.

WHICH TYPE OF BENEFITS CAN WE WIN BY ADAPTING THE AI IN OUR DAILY LIFE?

Of note: in this section, I will only discuss activities closely related to the job of statistical programmer and not everyday activities, where AI has already demonstrated added value (writing emails, documents, creating PowerPoint, etc.).

Artificial Intelligence (AI) brings several benefits to the job of a statistical programmer and in particular by automating some repetitive and non-added value tasks. By that, AI frees up time to focus on more strategic aspects of our work - and contribute to overall well-being. The future will probably provide us other areas where AI can help us.

Here are a few ways AI can positively impact our daily lives:

- Automation of repetitive tasks: AI can automate routine and repetitive tasks, such as data preprocessing, and generate basic statistical reports. This allows statistical programmers to focus on more complex and strategic aspects of their work (e.g., derivation of critical endpoints). AI frees up time, which can lead to reduced work-related stress and a better work-life balance. Last but not least, it will help us accelerate deadlines and allow us to obtain statistical results immediately (reduce the time between the DBL and the CSR).
- Faster information acquisition/digestion: AI can quickly process and analyze large amounts of data, providing statistical programmers with information within a shorter time frame.
- Extract information: Natural Language Processing (NLP) tools can help extract information from unstructured texts and offer the ability to use texts directly from end to end with less effort.

E.g. read the protocol and pre-fill metadata and programs or help create the pooled database by offering the best mapping.

- Data visualization and reporting: AI can generate interactive data visualizations and automated reports, supporting the decision making and allowing statistical programmers to communicate their results effectively to non-technical stakeholders.
- Error detection and quality control: AI can help detect errors, anomalies, and outliers in data sets, ensuring data quality and reducing search.
- Effective code generation:
 - AI-assisted tools can generate code for common/standard statistical tasks, saving time and reducing the risk of coding errors.
 - Propose code based on a question: AI-assisted tools can generate code based on questions, and when we have an urgent request, instead of starting from an empty page, AI can generate the first part of the program.

It is important to note that while AI offers many benefits, it will not replace human expertise (see the paragraph on the possible evolution of the statistical programmer). Indeed, statistical programmers will have to play a crucial role in designing analyses, interpreting results, and ensuring the validity and context of the statistical results proposed by the machine. The synergy between human judgment and AI capabilities will lead to stronger and more relevant outcomes.

WHAT ARE THE POTENTIAL LIMITATIONS OF THE AI?

Artificial Intelligence (AI) has made significant advancements, but it also has several limitations and challenges that need to be considered.

Here are some potential limitations of AI:

- **Data dependence:** AI models require large amounts of high-quality data for training. Without sufficient and representative data, the performance of AI systems can be compromised or lead to bias or fairness: they can induce biased results if the data on which they are formed are themselves erroneous. We must ask ourselves the question for each case: is it better to start from a blank page or not? Will AI have added value on this task or not?
- **Interpretability and Explainability:** Many AI models are considered "black boxes," making it challenging to understand how they arrive at their decisions, which is a concern in particular in the clinical development where the reproducibility and traceability are keys. AI can sometimes exhibit unpredictable behaviors that are hard to explain or understand.
- **Lack of Common-Sense Understanding:** AI models often lack true understanding of context and common sense, leading to incorrect responses in certain situations.
- **Ethical and Legal Concerns:** The use of AI raises complex ethical questions, including issues related to privacy, accountability, transparency. Health agencies, ethical commissions, countries, societies begin their reflection on good practices of AI.
- **High Computational Resource Requirement (Cost and Implementation Challenges):** Training and running complex AI models require significant computational resources. For pharmaceutical companies, AI is not our core business. Most Generative AI models required by pharmaceutical companies cannot be purchased 'off the shelf' since they require customization and integration of internal data to be of use. On the hand building in-house, generative AI for all our needs is likely cost-, resource-, and time-prohibitive.
- **Unemployment Concerns:** AI automation could lead to job displacement, especially for roles involving routine and repetitive tasks. (this is why the evolution of the role of statistical programmer is crucial)
- **Environmental Impact:** Training AI models can be energy-intensive, leading to concerns about their carbon footprint and contribution to environmental issues.

To fully leverage the benefits of AI while addressing these limitations, it's important to approach AI development and implementation with careful consideration of ethical, societal, and practical implications. Collaboration between all stakeholders (developers, regulatory agencies, business, etc.) is essential to navigate these challenges, limit them and create AI systems that are beneficial and responsible.

WHAT ARE THE POSSIBLE EVOLUTIONS OF THE STATISTICAL PROGRAMMER?

As AI takes on some of the tasks of a statistical programmer, it frees up time and resources for the programmer to focus on more creative, strategic, and high-value activities. Here are some activities that the statistical programmer can engage in while AI handles a portion of their work:

1. **Design requirements:** The statistical programmer can dedicate more time to designing and proposing new/innovative and sophisticated visual statistical analysis. They can explore propositions based on new questions, and test approaches to adopt for the best outcomes.
2. **Validation of the AI results and Communication:** As mentioned in the limitations of AI, it may generate complex results, but the statistical programmer continue to be responsible of validating, assessing and making a judgement on the quality of results produced by AI models. He can interpret, contextualize, and communicate them clearly and concisely to non-technical stakeholders. During the validation of an electronic submission component, he can also play an active role in assessing and managing regulatory aspects related to the use of AI in the clinical development domain.

3. **Generation and tuning AI:** For some statistical programmers, they can fine-tune pre-trained models and they generate the results.
4. **Model Optimization and Customization:** While AI can automatically adjust certain model parameters, the programmer can handle more advanced optimization and customization of models for specific scenarios (such as complex algorithm for solid tumors from end to end, e.g., from scan provided to investigator then to the determination of the best overall response then to the creation of the Kaplan Meier curve).
5. **Take more tasks from the statistician:** With automation saving time, the statistical programmer can dedicate time to support the statistician on their operational tasks.
6. **Training and Mentoring:** Experienced statistical programmers can allocate time to train and mentor less experienced team members, sharing their knowledge and skills.

In essence, the role of the statistical programmer will evolve into more strategic, creative, and high-level activities where human expertise is crucial for making informed decisions, communicating with other stakeholders, and ensuring the effectiveness and ethics of statistical analysis.

HOW DO WE PREPARE OURSELVES FOR THE EVOLVING STATISTICAL PROGRAMMER ROLE?

As seen below, generative AI comes with a plethora of opportunities to transform our ways of working and enhance our practices. That said, this disruptive technology would require significant adaptation. A lot of work would be focused on upskilling, engaging, and educating all employees.

Our role in this evolution of the business, as managers, will be to:

Supporting change and culture shift; assess and seize any relevant opportunities using thoughtful risk-taking approach: what are the main challenges of AI? How do we think we can maximize its benefits while minimizing risk?

Helping a team evolve their jobs in the context of artificial intelligence, especially AI generative technologies, requires a combination of strategic planning, skill development, open communication, and a supportive work environment. Here are some steps a manager can take to facilitate this evolution:

1. **Education and Awareness:** Start by educating the team about AI, its capabilities, and its potential impact on their roles. Foster an understanding of how AI generative technologies work and what they can contribute to the team's tasks and projects.
2. **Identify Opportunities:** Work with the team to identify areas where AI generative technologies can enhance their work. This could involve automating repetitive tasks, aiding in creative brainstorming, generating content, or improving decision-making through data analysis.
3. **Skill Development:** Provide opportunities for team members to acquire the necessary skills to work alongside AI. (see section relative to the up-skilling).
4. **Collaborative Approach:** Emphasize that AI should be viewed as an assistant rather than a replacement. Encourage the team to work together with AI tools to achieve better results. Highlight how AI can amplify their abilities and allow them to focus on more complex and strategic aspects of their roles.
5. **Experimentation and Pilot Projects:** Initiate small-scale pilot projects to test the integration of AI generative technologies within the team's workflow. This allows team members to gain hands-on experience, assess the benefits and challenges, and fine-tune their approach.

6. **Feedback Loop:** Establish a feedback loop where team members can share their experiences and suggestions regarding the use of AI. This will help identify areas for improvement and adjustments in the integration process.
7. **Continuous Learning:** AI technologies evolve rapidly. Encourage a culture of continuous learning by providing resources, workshops, and access to the latest AI advancements. This keeps the team updated and adaptable to new opportunities.
8. **Ethical Considerations:** Address ethical concerns associated with AI generative technologies. Keep inform on the evolution of the regulatory agencies, countries on the AI in particular on data privacy, traceability, reproductivity, validation, to ensure responsible and ethical use of AI.
9. **Support and Recognition:** Recognize and celebrate the team's successes in effectively integrating AI into their workflow. Provide support and guidance when challenges arise and acknowledge their efforts in adapting to new technologies.
10. **Adaptability and Flexibility:** Understand that the implementation of AI can bring about changes in roles and responsibilities. Be open to adjusting job descriptions and expectations as the team's tasks evolve due to the integration of AI.
11. **Clear Communication:** Maintain transparent communication about the goals, progress, and impact of AI initiatives. Address any concerns or uncertainties the team might have.
12. **Long-Term Vision:** Develop a long-term vision for how AI can contribute to the team's growth and success. Align this vision with the organization's strategic goals.

Remember that successful integration of AI generative technologies involves a balance between technology and the human element. By fostering a collaborative and supportive environment, managers can help their teams adapt and thrive in the evolving landscape of AI.

UP-SKILLING

Upskilling and training are critical to ensuring that we can adapt to the changes and take advantage of new opportunities. It is essential to stay informed about emerging technologies and regulatory guidelines and to develop new skills accordingly.

To adapt their role in the AI context, statistical programmers can focus on developing the following key skills (for some: we have already started):

Technical Skills:

- *Programming proficiency:* strong programming skills are essential. They should continue to improve their proficiency in programming languages. R and Python are the first two. (Others are emerging such as Julia or falcon).
 - Python provides comprehensive libraries for AI and machine learning, while R is well suited for statistical modeling and visualization.
- *Data visualization:* it is essential to present complex data and analysis results in a clear and concise manner. Statistical programmers must develop data visualization skills using tools such as R Shiny. The knowledge of this technology is important; indeed, the objective is not to replace a static output in visual output.
- *Data Science and Machine Learning* – Statistical programmers should understand the principles behind machine learning and become familiar with data science concepts. Statistical programmers can

leverage AI and automation to streamline repetitive tasks and improve efficiency. This way, they will have more time to focus on more complex and value-added activities, such as developing innovative solutions to address analytical challenges or own new tasks (previously done by statisticians or new emerging tasks)

- Deep Learning and Neural Networks: awareness is enough.
- Natural Language Processing (NLP) - Large Language Model (LLM), fine-tuning): a good comprehension as NLP can be useful for semi-automation of tasks.
- Big Data technologies (can be a +)

Competencies:

- *Collaboration and communication*: They need to develop stronger collaborative model to understand and translate the requirements of statisticians into actionable programming tasks and to find the best solution (among all solutions) to deliver results faster. In addition, they must effectively communicate their findings and knowledge to a non-technical audience, other stakeholders (like data science expert) and digital team.

- *Continuous learning and adaptability mindset*: technologies and AI are rapidly evolving. Statistical programmers must adopt a continuous learning mindset and stay informed about emerging trends, technologies, and methodologies. Adapting to new tools and techniques ensures they can effectively leverage AI capabilities and help advance statistical analysis.

- *Problem solving and critical thinking*: AI brings new challenges. Strong problem solving and critical thinking skills are critical for statistical programmers to tackle complex problems and deliver innovative solutions.

- Finally, develop a mindset of finding opportunities in all challenges/uncertainties.

With the increasing integration of AI and machine learning, the statistical programmer could be interested in combining several skills in programming, statistics, artificial intelligence, data analysis and humans. The current strength of the statistical programmer is his specific area of expertise: he has knowledge of clinical development and regulatory context.

CONCLUSION

The statistical programmer role is undergoing significant evolution, and we must be proactive in adapting to these changes. It is essential to stay informed, upskill ourselves and prioritize effective communication and time management.

Evaluating the impact of AI on a job requires a holistic approach that considers both quantitative and qualitative aspects. It's essential to involve everyone in the evaluation process and maintain open communication to address any challenges or concerns that arise during the transition.

We are at the beginning of the AI area, a lot of challenges, questions are there.

Lastly, humans must be in the loop and use their judgement. Users must be able to identify when a response is erroneous as using outputs without verifying the contents can result in potentially dangerous misinformation. Generative AI is not a replacement for critical thinking. Statistical programmer must be trained in how to engineer prompts, identify bias in content generate, and in the traceability and accountability of the use of generative AI-created content. Up-skilling and partnerships will be needed to successfully adopt generative AI.

My wish: human will develop an AI eco-responsible, without destroying ourselves.

ACKNOWLEDGMENTS

I would like to thank my Sanofi colleagues Mei Wu, Charlotte Cheinin, Valérie Gouet and Jérôme Repincay for their insights.

DISCLAIMER

The opinions and thoughts written in this paper/poster are subject to my own independent views and are not subject to the opinions of the organization that I represent.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Cindy Marabotti

Sanofi

Email: cindy.marabotti@sanofi.com