# *Guiding Principles for AI at Sanofi*

**GRI Standards:**

2-23: Policy Commitments

**EXECUTIVE SUMMARY**

These are the Guiding Principles for AI at Sanofi for Artificial Intelligence (AI) applications used in our products and services, evidence generation or affecting our workforce. This guidance puts forward our concrete asks from sponsors and project managers at Sanofi regarding how we expect them to implement AI in any projects in scope. Technology, practice, laws, and regulations and hence our guidance is evolving. This guidance is non-binding and provides orientation for responsible implementation on a best effort basis. We intend to adapt and formalize risk-based controls further in future as the body of good AI practice solidifies.

To complement the Guiding Principles approved in June 2021, Sanofi has engaged in several additional activities around AI and data ethics.

# TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1. DEFINITION

Artificial Intelligence (AI) is a technology which has progressed significantly with big data capabilities over recent years and attracts increasing interest for its promising potential applications. It also raises concerns about responsible use, reliability, and loss of human control. There are numerous definitions of AI, as it is an evolving technology. Our definition of AI centers on common elements of these definitions and machine learning applications, which use training and/or operational data to achieve specific objectives. Such a narrow definition of AI excludes general AI and reflects current practice in the business field, as broader, more general AI technology is not yet available.

***By AI we refer to systems, that use data to learn and adapt their action or output accordingly.***

AI application may be dynamically adapting based on data processed or locked. Examples of such AI applications can include Conversational Chatbots or Bots, which screen transactions or data points for extraordinary patterns indicating errors or fraud, or quality issues. AI functionality can also interpret images and be embedded in diagnostic devices or applications, spotting patterns indicative of risks or pathologies and possibly also recommending a related course action or offering conclusion.

## 1.2. OUR INTENT

Our products and services endeavor to improve lives and we aim to observe the highest standards in research, development, production, and commercialization. Our goal is to inform patients, consumers, and healthcare professionals (HCPs) transparently and objectively about our products and services, providing robust evidence, so they can make informed decisions while respecting their independent rights and decisions. Further, we will provide appropriate information for users to assess the benefits and risks involved with the use of our products and services. The patients and their safety are of primary importance. We want to work with all parties in the healthcare system and will strive to make our products available to all who may benefit from them.

Based on this intent, we have articulated a set of Guiding Principles for the responsible use of AI technology to reap the benefits of this technology and mitigate potential risks from misuse of this technology.

## 1.3. SCOPE

These Guiding Principles apply to all Sanofi affiliates, where AI is implemented in any of our products, or is used to provide services, evidence to external parties or used in processes affecting our workforce or affecting patients.

We intend to advocate these Guiding Principles also to our collaborating business partners.

# 2. *Collection, Storage and Use of Human Biological Samples for Research Policy - Lay Summary*

## 2.1. QUALITY AND CONFORMANCE WITH GOOD PRACTICE STANDARDS

***We aim for quality and conformance with good practice standards***

We acknowledge our responsibility for the quality of our products, evidence produced and services to perform as specified. We will continually adapt our internal governance frameworks to provide for adherence to these Guiding Principles and all applicable laws and regulations. We will include education, training and advice, practice sharing, data protection, consistency in implementation standards, complaints, continuous improvement, and oversight. We will also adapt these Guiding Principles as necessary as technology and its applications will evolve.

## 2.2. QUALITY AND CONFORMANCE WITH GOOD PRACTICE STANDARDS

***We will assess the reliability of data.***

We aim to use reliable data. AI applications are most robust if data are representative. We will build and combine different datasets to identify new relevant factors or applications and will aim to make our applications more robust for the target population and intended use. We will assess differences between training and real-world data, seek to identify any drift and evaluate their impact. We will train and operate AI application with quantitatively and qualitatively sufficient data, available and accessible in a secured and controlled environment.

## 2.3. ROBUSTNESS OF OUR AI APPLICATIONS

***We will assess the robustness of our AI applications.***

Our AI applications will be tested for stability before we release them. We will develop, maintain, improve and operate our applications, which deal with patient data or regulated processes under a Quality Management System (QMS), which will reflect current good industry practice. We will assess the risk in AI applications and submit them to a regular review and evaluation on the safety and robustness of their results. Applications which are critical to patient health and safety will be subject to a continuous monitoring of complaints, input data and their results.

We acknowledge, that an AI application generating scientific hypotheses can’t be validated to produce valid hypothesis only, but those hypotheses should be validated in prospective trials.

In certain cases, part of training data may correlate with the desired AI output without any causal or fundamental relationship. The trained output hence may become dependent on these factors. Where applicable, we will reduce such nuisance input factors with no link to the intended output or claim to the extent reasonably possible. For clinical applications, we will evidence the clinical association between the application’s input and the clinical condition, as well as the correct and reliable generation of output for the intended use.

## 2.4. MITIGATION OF BIAS OF OUR AI APPLICATIONS

***We will strive to mitigate bias in our AI applications.***

If training data is not representative, then AI applications may be less reliable in certain groups or situations. We will seek to identify potential bias, including for vulnerable groups or minorities, and to the extent available or obtainable, we will use respective identifying properties. Drift or bias will be assessed and evaluated, and appropriate corrective action will be taken (including additional training or

withdrawal as necessary). We will strive to prevent unfairly biased output for subgroups, which could e.g. originate from non-representative or biased training data or insufficient inclusion of minorities.

We will be transparent and caution or limit the use of our applications for subgroups, for which results are not sufficiently robust. We will strive to make our applications available to all who can benefit from them, yet we acknowledge that we may not have sufficient data to provide robust results for all populations. In these cases, we will take reasonable steps to include them.

## 2.5. TRANSPARENCY AND EXPLAINABILITY OF AI APPLICATIONS

***We will be transparent and will strive to explain AI applications.***

We will notify the use of AI to the user. We will assess the performance and quality of our applications and we will assess the plausibility of results. We will strive to make such information also available to the user as suitable and reasonably practical and to the extent it will not compromise our intellectual property.

We will suitably explain to our users the intended use of our AI applications and how they work. We will determine and inform users about the scope of use for our AI applications. We will also explain how we intend to control, maintain, correct and improve these applications.

## 2.6. PROTECTION OF PERSONAL DATA

***We will aim to protect personal data.***

We will aim to protect personal data. We will work with anonymized or pseudonymized data where possible. We will prevent or protect individuals from re-identification.

We will collect necessary personal data for a legitimate need and based on the individual's consent. Secondary use can also include use of their dataset in AI development or evaluation, or likewise in big data exploration or studies. We will respect, when primary or secondary use consent is withheld or withdrawn and will prospectively ensure that their data will no longer be used for such purpose.

We will retain personal data, while we will be required to keep records or while we will use it for research or development or evaluating potential bias in applications.

## 2.7. RESPECT OF INDIVIDUAL SELF-DETERMINATION

***We will respect the self-determination of individuals.***

We will refrain from any intervention, prescriptive output, nudging or autonomous action, which has not been expressly consented by the user based on fair and open information.

We will take reasonable steps to learn why users may dismiss the results of our AI applications to improve them. We will accept their decisions.

## 2.8. HUMAN INTERVENTION

***We will offer human intervention.***

For all autonomous AI applications which are critical to the interest of the affected individual, we will offer a mean to complain and trigger direct interaction with a human representative of the company.

# 3. Questions and Answers

**Are Advanced Analytics included in scope?**

Advanced analytics uses AI tools to identify factors which correlate to develop new hypothesis or identify new research targets or pathways. Only in a subsequent and mandatory phase, the hypothesis is evaluated by humans and validated typically in a randomized controlled trial, i.e. a prospective study. Individuals are not exposed to autonomously generated exploratory results. In the confirmatory trial step, any AI used must be subject to these guiding principles the applicable Quality and Regulatory procedures.

**Is this Guidance binding?**

This guidance puts forward our concrete asks from sponsors and project managers at Sanofi how we expect them to implement AI in any projects in scope. Technology, practice, laws, and regulations and hence our guidance is evolving. This guidance provides orientation for responsible implementation on a best effort basis. We intend to adapt and formalize risk-based controls further in future as the body of good AI practice solidifies.

**What are the differences among Anonymous, Pseudonymized and De-identified data?**

Pseudonymization refers to personal data which is protected with a separate key necessary to access identifying parts of the data. Anonymization, per EU regulation, refers to personal data, which cannot be linked to back to an identifiable individual at all (irreversible). US regulation refers to De-identified data, which has identifying parts of personal data removed and prohibit reidentification through combination of data while this may still be possible. Data protection laws and regulation do not protect anonymous data.

**Are Sanofi AI applications free from bias?**

We seek to design training data to reasonably minimize bias, notably for minorities and vulnerable groups to avoid discrimination. Real world operational use may evolve and differ from the trained population and hence bias may develop. There will always be residual bias and identification requires subgroup analysis and comparison with operational, real-world data. Bias also exists in traditional environments outside AI.

**How do you ensure AI applications are robust?**

We aim to test AI applications not only with historic data prior to release, but also prospectively to validate that the application provides correct output in practice. Operational data may differ or shift from training data. We will also verify the quality of training data.

We undertake to regularly re-test applications with real-world/operational data in appropriate intervals to assure continued performance.

# 4. Additional Activities around AI and Data Ethics

To complement the Guiding Principles approved in June 2021, Sanofi has engaged in additional activities around AI and data ethics:

- A dedicated Risk on "AI and Data Ethics" has been created on our Global Risk Map, owned by the Ethics & Business Integrity (EBI) Department — this is overseen in partnership with the Chief Data Officer and other key functions.
- A Guidance on the Ethics of Innovation in Healthcare and eHealth has been issued by Digital in partnership with EBI.
- Preparation is ongoing to implement the recently published IFPMA Principles on Data Ethics within Sanofi — the IFPMA Commission on Future Health Technologies, which produced these guidelines for the pharmaceutical industry, is chaired by Sanofi's Chief EBI Officer.
- The enclosed Sanofi Guiding Principles are aligned with IFPMA Artificial Intelligence Principles issued in June 2022.

- The French *Association Nationale des Sociétés par Actions* (where all CAC40 companies are represented) has launched a reflection on Digital and AI governance, and Sanofi is participating through its Chief EBI Officer and Head of Corporate Legal Functions.