

## Novartis feedback on Artificial Intelligence Act - 5 August 2021

Novartis welcomes the European Commission ambitious proposal for a comprehensive legal framework regarding the use of Artificial intelligence (AI). As a global healthcare company we are using this innovative science area and the latest technologies to discover and develop medicines and other treatments. For us and for many of our peers—and other companies in nearly all industries—artificial intelligence (AI) is transforming the way how we innovate and operate.

AI technologies offer tremendous potential for improving health care quality. Novartis has been using AI extensively in drug discovery for some time and, more recently, is working to apply AI to drug development in addition to other applications to improve our business processes. With respect to the development of medical products, we view the applicability of AI in two ways. One way includes the horizontal capabilities and tools that could help increase our understanding of unstructured data from various sources, such as images and texts for applications across the medical products we are developing. The other way is the application of AI algorithms and methodologies for product-specific uses across the life-cycle from early discovery, to development, manufacturing and deployment of our products and services.

At Novartis, we have defined and published our commitment to ethical and responsible use of AI in alignment with our code of ethics. This includes a commitment to deploy AI systems in a transparent and responsible way and to ensure that the use of AI systems has a clear purpose that is respectful of human rights, is accurate, truthful, not misleading, and appropriate for their intended context. We are in agreement that a well-designed AI regulation is supportive of the protection of fundamental rights, ensuring safety and attributing liability.

We have identified 3 key areas in the proposed AI Act that we would like to comment on in more detail:

### 1) Definition of scope

In terms of material scope, the proposed AI Regulation would apply to the placing on the market, putting into service and use of "*AI systems*". AI systems covered by the proposed AI Act are defined broadly to include:

- (i) software developed in accordance with the first Annex to the proposal – this Annex covers notably AI based on machine-learning approaches, logic and knowledge-based approaches and statistical approaches
- (ii) software that can, for a given set of human-defined objectives, generate outputs influencing the environments they interact with

Novartis believes that the scope of proposed regulation should be more precisely and narrowly focused on AI created through machine-learning approaches. In contrast, the application of more general terms such as "rules-based" and "statistical" AI could also be inferred to apply to applications which are not new, their underlying logic being completely transparent due to their explicit programming and/or curation. It is our concern that this regulation could otherwise potentially place undue burden on long-utilized techniques and approaches, which have been agreed among regulators and the healthcare industry to support decision-making today.

### 2) Risk assessment of AI systems

The current proposal introduces new oversight for "*high-risk*" AI systems which will require a case-by-case assessment from AI providers, based on other Annexes of the proposal and a series of criteria. Among the identified high risks are harm to health and safety that could result from application to human beings, the risk of negative impact on fundamental rights and the potential for discrimination. We are aligned with the need to address the aforementioned risks, but note that this proposal creates additional challenges for the development of AI-driven software, especially if a different risk-level is applied in

comparison to the Medical Device Regulation (MDR, 2017/745) and in-vitro Diagnostic Regulation (IVDR, 2017/746). We would welcome clarification on the proposed criteria to specify “malfunctioning” of an AI algorithm.

### 3) Harmonization of approaches

We understand the proposed AI Act will set forth obligations for manufacturers, importers and distributors that are in addition to obligations set forth in applicable regulations such as the Medical Device Regulation (MDR, 2017/745) or In-vitro Diagnostic Regulation (IVDR, 2017/746). Therefore, we would appreciate harmonized regulatory definitions for AI & related terms and aligned risk classification with the MDR and the IVDR and clarification on the responsibilities related to distributors who incorporate an algorithm vs. manufacturers of the algorithm.

Also we would appreciate more clarification on the range of non-medical device uses of AI in drug development and their classification for risk under the AI Act. For example, would AI uses to categorize patients based on radiology scans or genetic profiles fall into this category? This data could be captured from medical devices, in alignment with current GDPR, in the context of clinical trials and would optimize data analytics.

In terms of risk classification, would the following examples be classified as high or low risk?

- Machine Learning (ML) algorithm used to identify patients for inclusion in studies based on prognostic/ predictive features
- Cases where processing of patient data may constitute “biometric” identification in alignment with (EU) Nr. 910/2014.

Given the international interest in Artificial Intelligence, we also would appreciate a risk-based and globally harmonized approach to AI regulation including Member States and intra-European region alignment, with other countries/regions such as US Food and Drug Administration.

We welcome a transparent and predicted approach to the further development of the appendices.

We commend the EC for drafting this first of its kind legislative proposal which, together with related legislative initiatives, will contribute to positioning the EU as the leader of trustworthy AI and a key player in the digital space. As Novartis we look forward to further engaging with you to shape a system that is fit for the future and adapted to support and foster innovation.



# **Novartis' commitment to the ethical and responsible use of Artificial Intelligence (AI) Systems.**

**A human-centered approach in using  
Artificial Intelligence to reimagine medicine.**

# Table of contents

|   |                |
|---|----------------|
| <b>Background</b>                       | <b>Page 3</b>  |
| <b>Introduction</b>                     | <b>Page 5</b>  |
| <b>Themes and respective principles</b> | <b>Page 8</b>  |
| Empower Humanity                        | Page 9         |
| Accountability                          | Page 10        |
| Mitigate Bias                           | Page 11        |
| Respect Privacy                         | Page 12        |
| Transparent and Explainable             | Page 13        |
| Safe and Secure                         | Page 14        |
| Environmental Sustainability            | Page 15        |
| Review, Learn and Adapt                 | Page 16        |
| <b>Glossary</b>                         | <b>Page 17</b> |
| <b>References</b>                       | <b>Page 19</b> |

# Background

The pharmaceutical value chain provides medicines companies with opportunities to gather meaningful data at every touchpoint, from early biomedical research, to clinical trials and medicines production, through to patient and healthcare community engagement. Leveraging Artificial Intelligence (AI) at scale enables the industry to unlock the power of this data to establish valuable insights. The ability to use these insights to inform and accelerate decision making is what makes AI a transformative technology for the pharmaceuticals industry.<sup>1</sup>

This paper outlines Novartis' commitment to leveraging AI responsibly and ethically, in line with our overall purpose of improving and extending people's lives.

As a leading global healthcare company, powered by advanced therapy platforms and data science, Novartis is undergoing a digital transformation to embed cutting-edge digital technologies and data science into all parts of its business.

**We have applied AI broadly across Novartis. With over 100 use cases already developed, we are transforming the way we:**

- Innovate across R&D to develop novel therapies and drugs
- Optimize business processes, operations and commercial activities
- Engage with patients, healthcare professionals and partners

**Since 2018 we have been committed to progressing our ambitious enterprise transformation, aimed at answering three big 'what if' questions:**

- What if we could bring medicines to patients two years faster by transforming how we innovate in R&D;
- What if, with an eye to reinvest in R&D, we could significantly reduce our costs by \$1-2 billion by revolutionizing the way we work – optimizing and automating processes to drive breakthrough innovation;
- What if we could reach twice as many patients twice as fast, by rethinking traditional approaches to customer engagement and creating more personalized experiences?



These technological developments come with both opportunities and challenges, leading to important questions which, as a leading pharmaceutical company, we need to address thoughtfully and affirmatively. With AI playing such a critical role in helping us to achieve our digital transformation goals, we recognize the need to define clear ethical principles around AI.

In full alignment with the principles and commitments within our [Code of Ethics](#)<sup>2</sup>, this paper highlights our commitment to and ambition for responsible and ethical use of AI across our business.

Specifically, on AI:

#### **Our commitment:**

To deploy AI systems in a transparent and responsible way. We will ensure that the use of AI systems has a clear purpose that is respectful of human rights, and is accurate, truthful, not misleading, and appropriate for their intended context.

#### **Why it matters:**

AI can help Novartis increase patient access, improve customer experience, drive automation, provide predictive analytics and detect potential misconduct. It also has the potential to be used to improve the speed and accuracy of diagnosis, treatment protocols, drug discovery, drug development, patient monitoring, and patient care, among other applications that will improve patients' lives and optimize the healthcare ecosystem.

#### **Our Methodology:**

We engaged a team of leading ethicists and data privacy, legal and AI specialists both from within Novartis and externally. We developed an inventory of our current practices in AI and designed the following principles in line with our wider corporate Code of Ethics. Commitment has undergone rigorous review with the Independent Bioethics Advisory Committee (IBAC)<sup>3</sup> and has been approved by Novartis' Trust and Reputation Committee, which is chaired by the CEO.



Artificial Intelligence, as discussed in this document, at the most fundamental level refers to intelligent agents that receive percepts from the environment and take actions that affect that environment<sup>4</sup>, often implemented as software programs<sup>5</sup>. In order to affect actions, AI systems aim at performing machine simulations of human intelligence processes such as learning, reasoning and self-correction.<sup>6</sup>

# Introduction

**Novartis is harnessing the power of data and digital in reimagining medicine, employing data science and Artificial Intelligence (AI) in three broad areas:**

## 1. The development of novel therapies and drugs

The use of AI is being explored in the pre-clinical phase to understand disease biology and drug candidates; in the clinical phase to help target populations and to design intervention studies; and in the development of digital therapeutics and devices to enable continuous monitoring.

### Generative Chemistry

We use generative chemistry to augment chemistry teams with well-annotated, high-quality ideas in a seamless fashion for our end users. We use Machine Learning to scan billions of molecules in our compound library and propose virtual molecules with a desired target profile, as defined by our drug discovery experts. It efficiently reports the multi-parametric ideation process every medicinal chemist undertakes daily. The output is a manageable set of optimized compound suggestions that can be readily synthesized. Discovery scientists can either directly choose from select compounds or be informed to come up with related, yet novel ideas.

### MELLODDY<sup>7</sup> (Machine Learning Ledger Orchestration for Drug Discovery)

The MELLODDY project (Innovative Medicines Initiative consortium, of which we are part) has created an AI platform that learns from proprietary compound assay data (>one billion data points for 10 million small molecules) contributed by multiple pharmaceutical companies, while maintaining confidentiality through blockchain-based encryption. Companies maintain control over their own data and resulting Machine Learning models. The models learn correlations between chemical substructures and activities in biological assays of disease relevance and benefit from techniques such as 'transfer learning', the principle that prediction accuracy may be enhanced by learning from models in adjacent areas. MELLODDY will enable cheaper, faster and higher-throughput drug discovery by providing structure-activity information for legacy and current assays in our drug discovery pipeline.

**Novartis is harnessing the power of data and digital in reimagining medicine, employing data science and Artificial Intelligence (AI) in three broad areas:**

## **2. The optimization of business processes and operations**

The use of AI is being explored and may improve processes in clinical development, manufacturing, and supply chain by automating, optimizing and re-engineering processes. In the business services area, we use AI to ensure efficiencies, effectiveness and drive operational excellence and compliance.

### **Buying Engine: AI-powered marketplace for Novartis**

Designed to streamline and centralize purchasing decisions, Buying Engine aims to enable procurement efficiency across Novartis, by creating a 'one-stop-shop' algorithmic-based marketplace, starting with lab supplies, PPE and potential spare parts (indirect material). The goal of this system is to provide transparency and recommend optimal buying choices in near real-time, leveraging multiple techniques from knowledge-representation, recommender systems, optimization, and Machine Learning algorithms to achieve its goal.

### **AE Brain: Automating repetitive processes**

AE Brain improves the quality of our safety information and also reduces the burden of manual repetitive work. AE Brain processes messages to identify potential adverse events and technical complaints in these messages. The system ingests textual data from multiple sources and applies Natural Language Processing (NLP) technology to understand the contents of those text documents to identify adverse events. This system is integrated into the workflow of human experts as a decision support system.

### **Marketing Mix Models**

Marketing mix models (MMMs) are statistical models for measuring the effectiveness of various marketing activities such as promotion, media advertisement, etc. These models can be of many types, but multiple regression is the workhorse of most marketing mix modeling. Regression is based on a number of inputs (or independent variables) and how these relate to an outcome (or dependent variable) such as sales or profits or both.



**Novartis is harnessing the power of data and digital in reimagining medicine, employing data science and Artificial Intelligence (AI) in three broad areas:**

### **3. Engagement with patients, healthcare professionals and partners**

The use of AI is being explored to enhance engagement with stakeholders and participants in the healthcare systems with the objective of supporting patients and generating insights.

#### **Ai Nurse: empowering patients**

Novartis partnered with Tencent to develop a WeChat mini-app, called Ai Nurse, for patients diagnosed with heart failure. The patient engagement platform is designed to empower patients and their healthcare providers to be more aware of their condition and to take appropriate actions to improve their health and wellbeing. The app uses multiple AI-driven algorithms to transform voice to text and text to voice. Algorithms are used to anticipate disease progression, recommend activities and provide targeted coaching and education. All of this data is continuously assimilated and interpreted to assess a patient's improvement or worsening condition. Accordingly, nurses and physicians can remotely track patients, with full consent and privacy protections as discussed in this paper, and provide additional continuity of care recommendations.

For Novartis to improve and extend peoples' lives in a sustainable manner, we must collaborate with trusted partners in tech, academia and other areas. Hence, in collaboration with its AI partners, Novartis is committed to using AI systems responsibly and in full alignment to the commitments and principles articulated in our Code of Ethics:

1. Empower Humanity
2. Hold Ourselves Accountable
3. Mitigate Bias
4. Respect Privacy
5. Be Transparent and Explainable
6. Assure Safety and Security by Design
7. Prioritize Environmental Sustainability
8. Review, Learn and Adapt

# Themes and respective principles

**Novartis believes that any development, application or use of AI systems should be governed within the following ethical principles which are fully aligned to the respective Novartis Code of Ethics principles and commitments.**

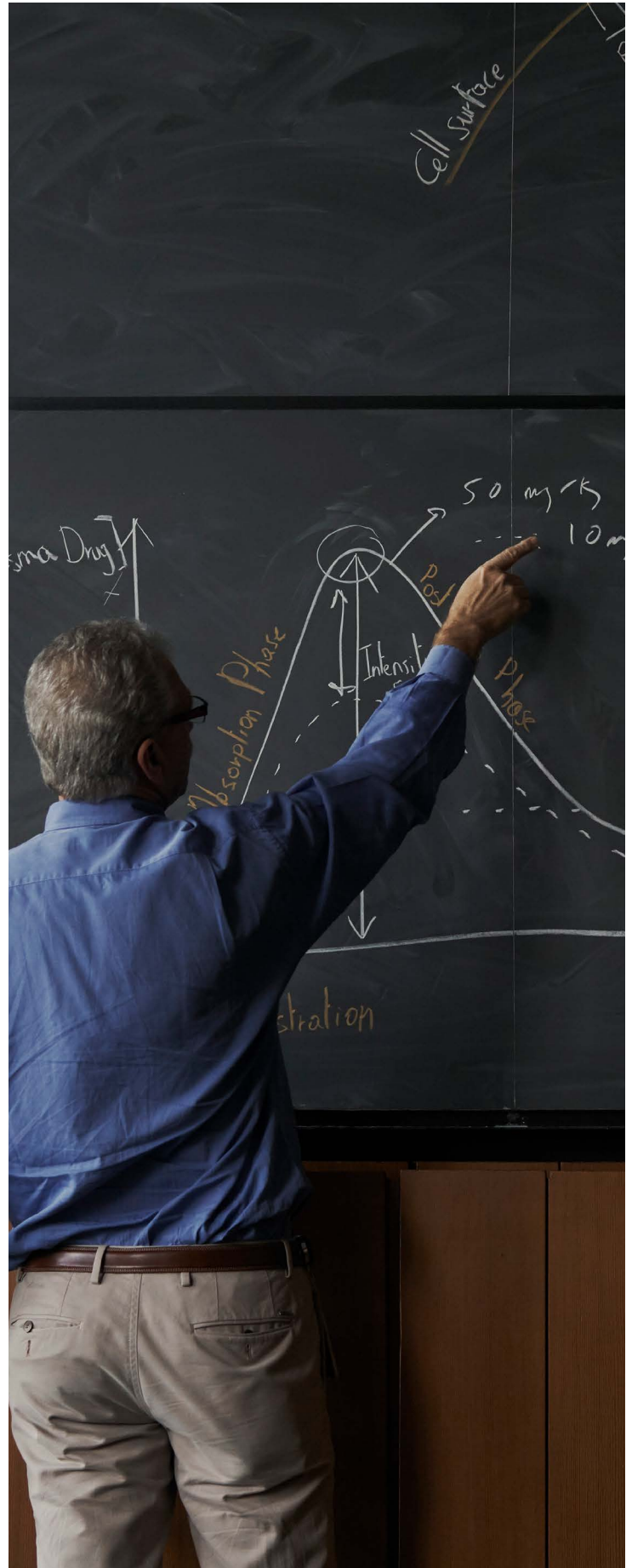


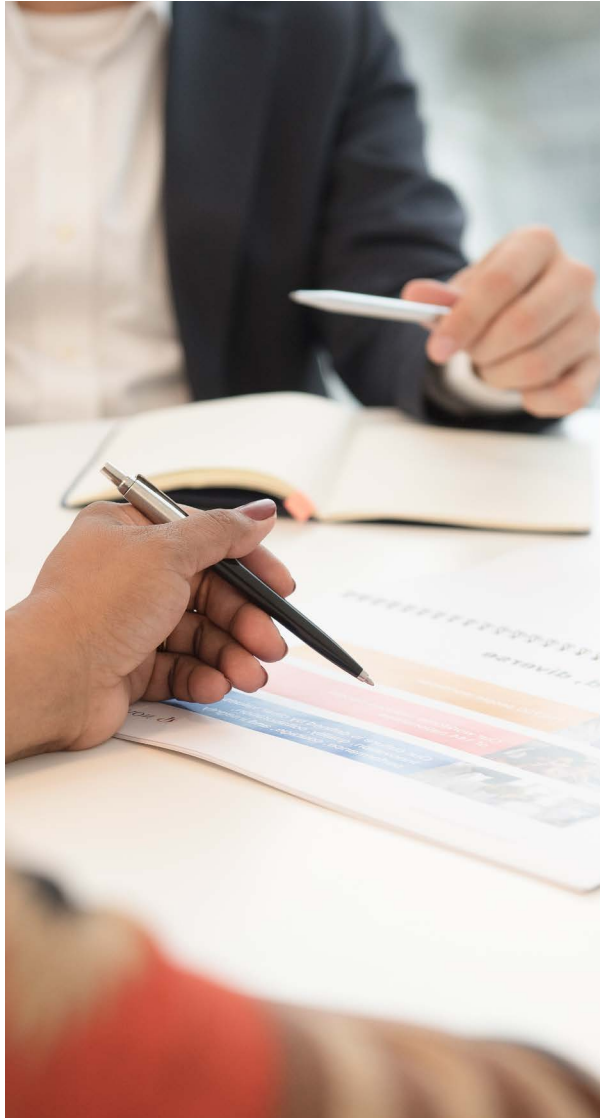
## PRINCIPLE 1:

### Empower Humanity<sup>8</sup>:

At Novartis, our values and culture are driven and defined by our purpose to reimagine medicines to improve and extend people's lives. Our everyday decision making is based on our ethical principles, as outlined in our Code of Ethics. These values and ethical principles form the basis from which we design, implement, and deploy AI. Novartis is committed to:

- Enforcing human-centric design in the deployment and use of AI systems;
- Building a mutually beneficial relationship between human knowledge, expertise and decision-making and the computational machinery which provides inferences and connections between data at scale.
- Respecting the rights and dignity of all people, and striving to prevent and mitigate identified adverse human rights impacts that may arise through our use of AI;
- Continuously assessing AI advances to ensure they proceed from within Novartis' context and are determined by Novartis, rather than influenced by external factors;
- Monitoring the impacts of AI to evolving human and societal values.





## PRINCIPLE 2: Accountability<sup>9</sup>:

As an accountable organization, Novartis is committed to establishing robust governance over the design and use of AI. Such rigorous governance includes appropriate leadership and oversight, risk and impact assessments, appropriate policies and procedures, transparency, training and awareness, monitoring and verification, response and enforcement. Therefore, Novartis is committed to:

- Maintaining human accountability in decision-making processes of designing, delivering and operating AI systems;
- Providing autonomy to associates in the controlling, creation, training, deployment and operation of AI systems;
- Performing business and regulatory impact assessments of AI systems within the Novartis value chain before integration and deployment;
- Applying Novartis Information Technology (IT) and Operation Technology (OT) controls and processes to plan, implement and continuously monitor AI systems, in alignment to the commitments in the Code of Ethics;
- Proactively monitoring and mitigating potential negative AI consequences;
- Enabling the auditability of the AI systems via validation and verification functionalities and keeping an audit trail in line with best practice.

We are building capabilities to elevate the practice of data science and AI across the enterprise and sparking a mindset shift so associates feel empowered by data science and AI, not threatened by it. In collaboration with the Novartis Learning organization, we have established the 'Data Science Academy' that brings online and in-person education to Novartis associates. We are working towards annual accreditation for both data scientists and executives who have to enable data science.



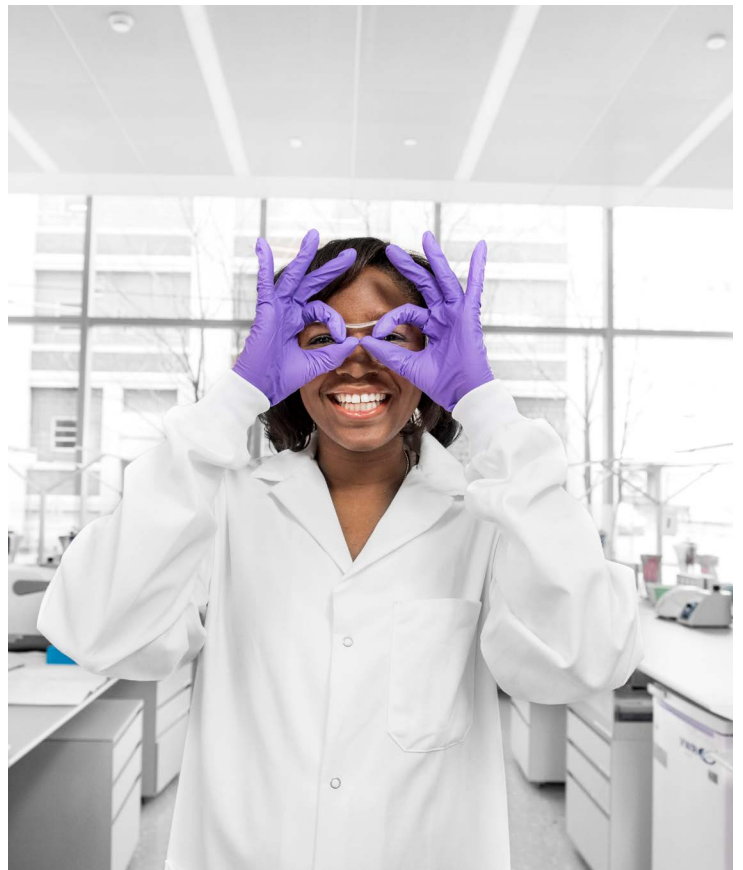
## PRINCIPLE 3:

### Mitigate Bias<sup>11,12,13.</sup>

Data and algorithms used in AI systems need to meet Novartis' strong commitment to fairness and non-discrimination detailed, inter alia in our Code of Ethics; particularly where AI systems are used in sensitive areas that closely touch critical decisions regarding drug development, socio-economic benefits, hiring and matters that relate to human behavior. We are committed to mitigating the risk of bias throughout the process, from data gathering, model creation and application of the model.

To that end, we will strive to:

- Design, develop, test, train and operate AI algorithms based on inclusive and representative data to eliminate possible biases and known discriminatory aspects such as race, gender, ethnicity, sexual orientation, political or religious beliefs;
- Use data samples that are representative of the studied and analyzed population to eliminate or prevent unconscious bias;
- Perform a risk impact assessment on the AI systems before their use in production to eliminate the risk of bias or discrimination;
- Develop and use AI systems in ways that reflect the social and cultural diversity of Novartis;
- In the short-term, assess, acquire or develop tools and establish techniques to assess statistical bias in data-sets from external sources – mitigating bias in all data sourced from outside of Novartis;
- Ensure the responsible use of AI when applied to the real world, as outlined in our 'Empower Humanity' Principle.





## PRINCIPLE 4:

### Respect Privacy<sup>14</sup>:

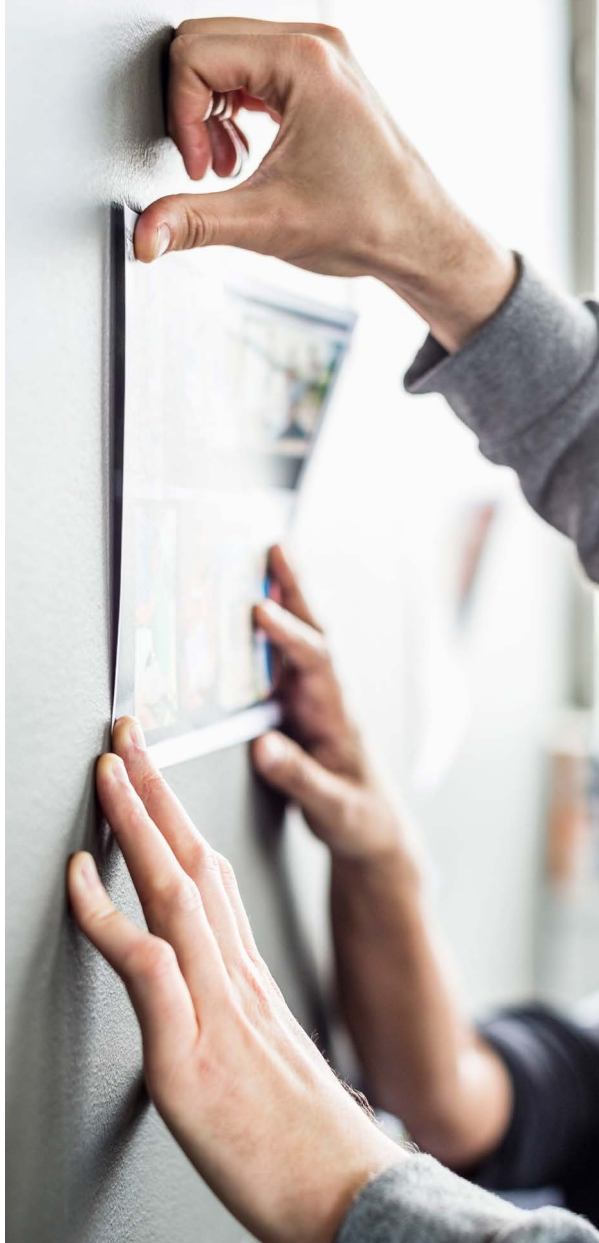
In some instances, AI systems are 'trained' on and use personal information. Outputs of AI systems may also impact the privacy of individuals.

- Novartis has established and implemented Global Data Privacy Principles<sup>15</sup> that govern the use of personal information. These Principles apply without exception to the design and use of any AI system. The Principles are:
  - Transparency: We are transparent about what personal information we process, how and why we collect it, use it, and who we share it with. We explain this in clear and simple language.
  - Legitimate and Meaningful Collection: We connect all collection and use of personal information to specific business purposes related to how we operate, innovate or engage.
  - Responsible and Sustainable Processing: We use personal information only in ways compatible with the purposes for which it was collected. We facilitate Individuals to exercise their rights with regards to their personal information.
  - Security: We protect personal information by using reasonable safeguards to prevent its loss, unauthorized access, use, alteration or unauthorized disclosure.
  - Integrity and Quality: We take appropriate steps to keep personal information accurate and up to date.
  - Minimal Retention: We keep personal information only for as long we can legitimately use it.



Designing and training viable AI models requires large samples of real-world data. In order to preserve transparency and explainability, while correcting for bias, the data sets used in the process may need to be preserved, creating tension with Privacy Principles, in particular the principle of minimal retention. Therefore, whenever possible, we are committed to finding alternatives to the use of personal information when designing AI. Such alternatives may be synthetic or anonymized data, or documentations of our approaches that do not require the retention of the training data sets.

For Novartis' full and detailed approach to Respecting Privacy, including details of implementation and tangible use cases of these principles in our business, please refer to the Novartis' Global Data Privacy Policy.<sup>16</sup>



## PRINCIPLE 5: Transparent and Explainable<sup>17</sup>:

Novartis strives to create transparency around the design and use of AI systems to explain how such systems work through:

- Short term: Openly disclosing / informing end-users when they are interacting with an AI system;
- Mid-term: Enabling the auditability and traceability of the decision pathways taken by AI systems using IT tools and infrastructure;
- Mid-term: Transparently communicating and explaining the limitations, purpose, decisions and capability of AI systems as new visualization models are developed;
- Ensuring the use of AI systems has a clear purpose that is accurate, truthful, not misleading, and appropriate for their intended context; aligning with the principles of Beneficial AI.<sup>18</sup>

## PRINCIPLE 6: Safe and Secure<sup>19</sup>:

AI systems need to be safe, performing as intended, secure and resistant to compromise via unauthorized parties. Hence, in the design, implementation and use of AI systems, Novartis commits to the following:

- Technically robust systems that translate in-depth human understanding to stable operations based on a review of impact assessments and the specific context of the use-case;
- If the AI systems are deployed in relation to products and manufacturing environments, we are committed to reporting adverse events within 24 hours of discovery to the Novartis Safety Department and quality complaints to Quality Assurance, and then transparently communicating the risks of our medicines and devices to regulatory authorities;
- In relation to confidentiality, Integrity and Availability of Novartis Information, we hold ourselves accountable for the information and technology that we handle, with an obligation to safeguard our patients' and partners' information.



### Clinical Trial Procedures

The Novartis clinical trial procedures are aligned with the CONSORT<sup>20</sup> 2010 and SPIRIT<sup>21</sup> 2013 statements, which are evidence-based guidelines to ensure transparent evaluation of new interventions in clinical trials - in study design, methodology and reporting.

Since publication, both statements have had AI extensions, developed through international multi-stakeholder consensus, to ensure the safe use of AI in clinical trials. Novartis follows both AI extensions to guide the safe and ethical use of AI in clinical trials, and ensure transparency in the reporting of AI-specific information.

### Nerve Live

Data and analytics platform, Nerve Live, harnesses past and present operational data, providing access to decades of drug development “experience” buried across multiple sources. The platform enables the systematic application of machine learning and predictive analytics to generate “intelligence”: new insights across multiple functional areas. To action the insights and create “value,” we crafted skillfully designed end-user applications for domain experts to plan, track, predict, compare and monitor domain activities, optimize costs, and maximize quality.<sup>22</sup>



## PRINCIPLE 7:

### Environmental Sustainability<sup>23</sup>:

AI systems need to be designed sustainably, inter alia, assessing the resource usage and energy consumption to limit the risks to the environment. To address the environmental footprint of AI systems (e.g. assessing the resource usage and energy consumption), the Environmental Sustainability principle within the Code of Ethics would apply. This principle lays out that Novartis is committed to minimizing the environmental impact of our activities and products over their lifecycle.

Novartis is aiming for carbon neutrality across the supply chain by 2030. In AI, this means addressing three broad areas:

- Short term: Partnering with like-minded sustainable technology platforms. Novartis will introduce sustainability as a key component in procurement of the computational infrastructure required for AI solutions and services;
- Mid-term: Ensuring optimal use of algorithms with internal implementation of AI, by training data scientists to be selective about the algorithms they want to train upfront, before committing them to the computational power required to deploy deep learning;
- Long-term: Reviewing internal operations, such as Novartis Technical Operations (NTO) to assess how AI can be used to reduce carbon footprint.



## PRINCIPLE 8:

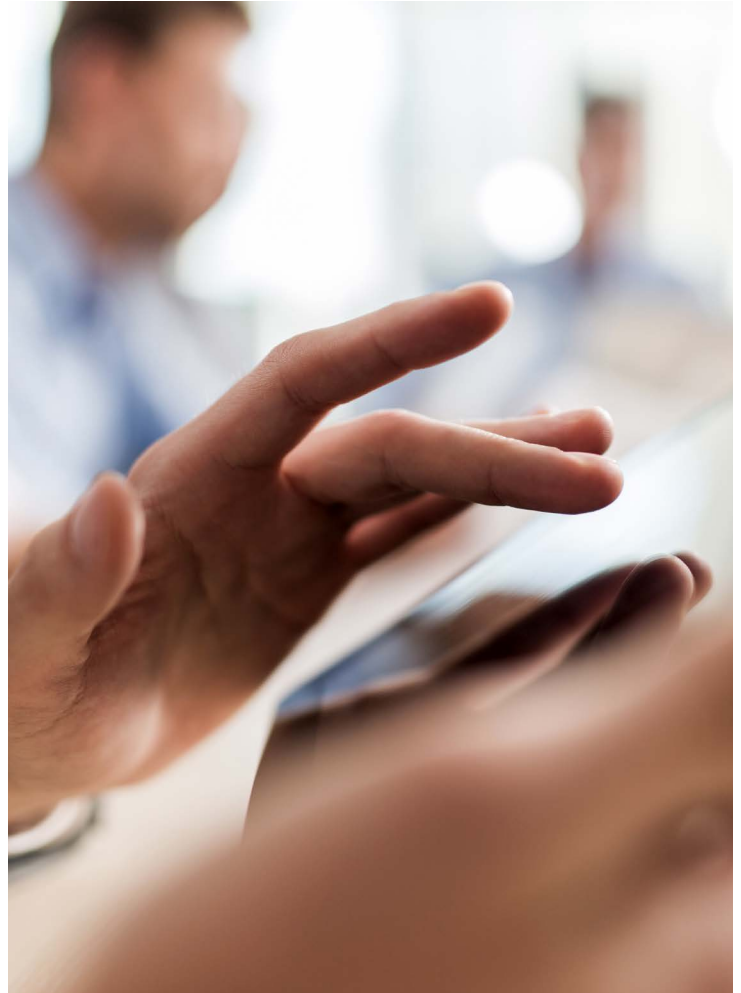
### Review, Learn and Adapt<sup>24</sup>:

AI systems need to support and enable professional standards. As such, Novartis is committed to:

- Ensuring and maintaining professionalism and accountability in the creation and deployment of AI systems; ensuring that associates have the necessary depth of understanding of the ethical implications;
- Implementing and using AI systems that augment, complement and empower human capabilities and skills to improve speed, quality and maximize impact in a positive way;
- Enhancing the offering of our Data Science Academy to educate data scientists as well the broader group of Novartis associates on the use of AI;
- Empowering, educating and training associates in the short-term to have the right ethical professional awareness (knowledge, experience and required skills) as they use or operate AI systems, to ensure that ethical commitments (as laid out in the Code of Ethics) are not compromised; moving in the mid-term to a system of certification;

#### Disclaimer:

Novartis recognizes that AI is evolving rapidly in our industry and in society overall. We also recognize that we are at the beginning of the journey to embed these principles for responsible and ethical use of AI in our governance structures, our operations, and our businesses. Novartis is committed to becoming a leading and responsible voice in helping shaping and governing AI. We believe that our leadership, and willingness to adapt and learn as AI evolves, will also build critical trust with our patients, our associates, our partners, and other stakeholders as the benefits of AI are realized in the years ahead.





# Glossary

**This glossary is meant to help in the understanding of the terms used in this paper.**

## Artificial Intelligence or AI systems

Artificial Intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behavior by analyzing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as Machine Learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).

## AI systems life cycle

An AI system's life cycle encompasses its development (including research, design, data provision, and limited trials), deployment (including implementation) and use phase.

## Beneficial AI

Ensuring that AI is always grounded in the principle of improving and extending human life. "Instead of building systems that optimize arbitrary objectives, we need to learn how to build systems that will, in fact, be provably beneficial for us."<sup>25</sup>

## Bias

Bias is an inclination of prejudice towards or against a person, object, group or position.

## Communication

AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system. This entails that AI systems must be identifiable as such. Beyond this, the AI system's capabilities and limitations should be communicated to AI practitioners or end-users in a manner appropriate to the use case at hand. This could encompass communication of the AI system's level of accuracy, as well as its limitations.

## Explainability

Explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Moreover, trade-offs might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability). Whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process. Such explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator or researcher). In addition, explanations of the degree to which an AI system influences and shapes the organizational decision-making process, design choices of the system, and the rationale for deploying it, should be available (hence ensuring business model transparency).

## Good Machine Learning Practices (GMLP)

GMLP are those Artificial Intelligence or Machine Learning practices (e.g. data management, feature extraction, training and evaluation) that are akin to good software engineering practices or quality system practices.

### Human-centric AI

The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.

### Machine Learning (ML)

The scientific study of algorithms that build a mathematical model of sample data to make predictions or decisions without being explicitly programmed to perform the task<sup>xi</sup>. ML is often considered to be a branch of AI.<sup>26</sup>

### Natural Language Processing (NLP)

A subfield of AI concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. NLP draws from many disciplines including computer science and computational linguistics.<sup>27</sup>

### Prevent Misconduct

Novartis intend to use AI in a legal and responsible way to detect and prevent potential misconduct within its business practices

### Time spans

Short term refers to 1-2 years; mid-term to 3-5 years; long-term to 5+ years.

### Synthetic Data

Microdata records created to improve data utility while preventing disclosure of confidential respondent information. Synthetic data is created by statistically modeling original data and then using those models to generate new data values that reproduce the original data's statistical properties. Users are unable to identify the information of the entities that provided the original data.

### Recommender Systems

A recommender system is one that uses active information-filtering techniques to exploit past user behavior to suggest information tailored to an end user's goals.<sup>29</sup>

### Trustworthy

Trustworthy AI has three components, which should be met throughout the system's entire life cycle: (1) it should be lawful, complying with all applicable laws and regulations (2) it should be ethical, ensuring adherence to ethical principles and values and (3) it should be robust, from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm.

### Transparency

Transparency is closely linked with the principle of explicability and encompasses transparency of element relevant to an AI system: the data, the system and the business models. The Berkman Klein definition is as follows: "Principles under this theme articulate requirements that AI systems be designed and implemented to allow for oversight, including through translation of their operations into intelligible outputs and the provision of information about where, when, and how they are being used."<sup>30</sup>

### Traceability

The data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency. This also applies to the decisions made by the AI system. This enables identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes. Traceability facilitates auditability as well as explainability.

# References

1. <https://50years.ifpma.org/present-to-future/technology/>
2. <https://www.novartis.com/sites/www.novartis.com/files/code-of-ethics-english.pdf>
3. <http://www.ge2p2.org/> - IBAC- provides analysis and recommendations on Novartis guidelines and policies for the ethical conduct of clinical research, and on selected ethical challenges which may arise in clinical trials, development programs, managed access programs and other areas across the Novartis enterprise. IBAC is comprised of bioethicists, clinicians, healthcare practitioners, patient advocates and other domain knowledge experts appropriate to the problem at hand.
4. Human Compatible: AI and the Problem of Control, Russell, S.J; 2020; Penguin Books, Limited. See also: <https://www.penguin.co.uk/books/307/307948/human-compatible/9780141987507.html>
5. Artificial Intelligence in Healthcare, Nature Biomedical Engineering, Vol 2, Kun-Hsing Yu, Andrew L. Beam and Isaac S. Kohane; 2018. See also: <https://www.nature.com/articles/s41551-018-0305-z>
6. Artificial Intelligence for clinical trial design, Trends in pharmacological sciences, Stefan Harrer, Pratik Shah, Bhavna Antony, Jianying Hu; 2019. See also: <https://www.thetalkingmachines.com/sites/default/files/2019-07/piis0165614719301300.pdf>
7. <https://www.melloddy.eu/>
8. To conduct our business in a manner that respects the rights and dignity of all people. We will strive to prevent, mitigate and remedy adverse human rights impacts throughout our workplace, business operations and in the communities in which we work. We want to protect people from abuse by those who are more powerful.
9. To deploy Artificial Intelligence (AI) systems in a transparent and responsible way. We will ensure that the use of AI systems has a clear purpose that is respectful of human rights, and is accurate, truthful, not misleading, and appropriate for their intended context
10. <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>
11. We will educate our people on inclusivity and provide all associates with equal opportunities to contribute to our company and advance their careers. We will listen to different communities with a learning mindset, to do what we can to contribute to building a world that is safer and more inclusive
12. To conduct our business in a manner that respects the rights and dignity of all people. We will strive to prevent, mitigate and remedy adverse human rights impacts throughout our workplace, business operations and in the communities in which we work. We want to protect people from abuse by those who are more powerful.
13. We will respect the rights, safety and dignity of individuals and communities, protect scientific integrity and strive to advance the practice of medicine. We will make sure that any data or information that we create or are responsible for, is true, accurate and fair. We do not make false or misleading statements.
14. To use personal information that we are entrusted with in a responsible way. We will adhere to our Data Privacy principles and ensure our external service providers also commit to these principles.
15. <https://www.novartis.com/sites/www.novartis.com/files/novartis-data-privacy-principles.pdf>
16. <https://www.novartis.com/sites/www.novartis.com/files/novartis-data-privacy-principles.pdf>
17. To deploy Artificial Intelligence (AI) systems in a transparent and responsible way. We will ensure that the use of AI systems has a clear purpose that is respectful of human rights, and is accurate, truthful, not misleading, and appropriate for their intended context
18. Human Compatible: AI and the Problem of Control, Russell, S.J; 2020; Penguin Books, Limited. See also: <https://www.penguin.co.uk/books/307/307948/human-compatible/9780141987507.html>
19. To protect our data and technology and ensure that information is kept safe from theft, loss, misuse or disclosure. We will take accountability for the information and technology we handle.
20. <https://www.nature.com/articles/s41591-020-1034-x>
21. <https://www.nature.com/articles/s41591-020-1037-7>
22. Leading a Digital Transformation in the Pharmaceutical Industry: Reimagining the Way We Work in Global Drug Development, Clinical Pharmacology & Therapeutics, Luca A. Finelli, Vas Narasimhan; 2020. See also: <https://ascpt.onlinelibrary.wiley.com/doi/full/10.1002/cpt.1850>
23. To minimize the environmental impact of our activities and products over their lifecycle. We will strive for a positive effect on climate, by reducing our carbon footprint, waste and water usage and making efficient use of natural resources.
24. To maintain high standards of ethical business conduct. We are committed to the same high standards of ethical business conduct wherever we do business.
25. Human Compatible: AI and the Problem of Control, Russell, S.J; 2020; Penguin Books, Limited. See also: <https://www.penguin.co.uk/books/307/307948/human-compatible/9780141987507.html>
26. Artificial Intelligence for clinical trial design, Trends in pharmacological sciences, Stefan Harrer, Pratik Shah, Bhavna Antony, Jianying Hu; 2019. See also: <https://www.thetalkingmachines.com/sites/default/files/2019-07/piis0165614719301300.pdf>
27. Artificial Intelligence for clinical trial design, Trends in pharmacological sciences, Stefan Harrer, Pratik Shah, Bhavna Antony, Jianying Hu; 2019. See also: <https://www.thetalkingmachines.com/sites/default/files/2019-07/piis0165614719301300.pdf>
28. <https://www.census.gov/about/policies/quality/standards/glossary.html>
29. [https://www.ll.mit.edu/sites/default/files/page/doc/2018-05/22\\_1\\_6\\_Gadepally.pdf](https://www.ll.mit.edu/sites/default/files/page/doc/2018-05/22_1_6_Gadepally.pdf)
30. Principled Artificial Intelligence, Berkman Klein Center. See also <https://cyber.harvard.edu/publication/2020/principled-ai>