V1.1

Global Artificial Intelligence Framework (GAIF)

This publication is available free of charge

September 2023

AI Ethics Forum
Noah M. Kenney, President & Lead Author

Update Schedule and Versions:
This framework utilizes a two number version control system. The first number (i.e., V1.X) indicates the major version number. Minor changes are denoted in the second number (i.e., VX.1). Unless otherwise noted, it should be assumed that only the latest framework version is endorsed by the AI Ethics Forum. The table below documents the revision history of this document:

Version 1.1	Initial release (Published September 11, 2023)

Purpose of Document and Acceptable Uses:
This framework was designed in light of a general lack of regulation as it relates to artificial intelligence, along with the increasing development of and access to this technology. The framework is designed to serve as guidance for companies, non-profits, and governments to evaluate the risks and ethical concerns posed by artificial intelligence, in addition to suggested methods of addressing these challenges. While the beginning of the document is designed to be applicable to all uses of artificial intelligence, the later sections address industry specific challenges.

This document may be used and distributed internally or externally by any individual, company, non-profit, or other organization. No person or organization shall present this framework as their own work, and all copyrights are retained by the AI Ethics Forum. When information from this framework is shared, quoted, paraphrased, or otherwise disseminated, a proper citation to this document shall be made.

No person or organization may charge for access to this document, directly or indirectly, without prior express written consent given by the AI Ethics Forum.

Disclaimers:
No part of this document is meant to constitute legal advice, nor is it meant to be considered a comprehensive analysis of all possible threats posed by artificial intelligence. Given the rate of advancement in this field, we recognize that there may be considerable threats posed by artificial intelligence not covered in this document, despite our best efforts to ensure comprehensiveness.

We implore companies to abide by any and all local, state, or federal regulations that may govern topics covered in this document. In no case shall any part of this framework be meant to supersede applicable laws or regulations, and we recommend that all organizations seek legal counsel prior to the implementation of this framework.

You agree to hold the AI Ethics Forum, the framework co-authors, and any other indirect or direct stakeholders blameless for any harm or damages resulting from implementation of this framework.

Section 1: Executive Summary

Section 1.1: Purpose of Framework

In recent years, Artificial Intelligence (AI) has become increasingly prevalent, and is now ingrained in the daily lives of most people. From simple AI-powered algorithms that recommend movies or music to large language learning models to autonomous vehicles, AI has reached nearly every aspect of society. Yet, despite its prevalence, artificial intelligence is largely unregulated. We find three primary reasons why governments may be ill-suited to developing AI regulations:

1. Speed of innovation: Many governments are unable to implement regulation without extensive review, approval, and implementation processes. In light of the rate of technological innovation within the field of AI, this often prevents a large challenge.
2. Lack of understanding: Government leaders are required to analyze and implement policy to address a wide range of issues. As such, it is impossible for government leaders to be experts in every policy area. This is especially true in policy related to artificial intelligence, given the complexity of the technology. Ultimately, this often results in regulations that are impractical to implement.
3. Decision bias: Government leaders, particularly in democratic governments, are generally concerned with gaining approval of constituents. As such, they are bias toward policy that may be popular, even if aforementioned policy fails to meet the needs of constituents.

In light of this, we understand that it may not be possible for most governments to effectively regulate artificial intelligence. Accordingly, this framework was developed to provide a practical approach for individuals, businesses, and organizations to develop, deploy, and utilize artificial intelligence in an ethical manner.

Section 1.2: Governing Principles

To achieve our goal of developing a practical framework, we have outlined several governing principles that shape our analysis. While this framework considers many objectives, we deemed the following three most important:

1. Ensuring data privacy and reducing bias: What is the least amount of data that can be shared to effectively accomplish a task? How can a lack of data privacy lead to discrimination in algorithmic decision making within your respective focus area?
2. Economic implications: How will a particular policy impact the economy? Each focus area will have different economic considerations, which must be taken into account, given that business feasibility must not be in conflict with other objectives.
3. Mass destruction: How can a particular type of AI be “weaponized” or used for malicious intent? How can we overcome this? Assuming, we cannot mitigate the threat completely, at what level do we consider it “safe enough” to implement anyways?

Section 1.3: Using the Framework

Given that every business and organization is different, this framework is meant to provide flexibility in implementation. In some cases, we provide fixed parameters to govern the use of development, deployment, and use of artificial intelligence. We were particularly uncompromising in matters of ethics. In other cases, we acknowledge tradeoffs between business interests and user desires. For example, we recognize an inherent trade-off between ensuring data privacy or anonymity in AI model training and ensuring data integrity or verifiability. In these cases, we seek to acknowledge the trade-offs, challenges, and concerns posed by each extreme.

Our general guidance is that businesses should operate in a manner that prioritizes ethics, privacy, and anti-discrimination metrics without compromising business interests beyond what is deemed reasonable. In short, we give significant freedom and flexibility to businesses and organizations. We provide "optimal scenarios", with an understanding that what is optimal may not always be feasible.

Section 1.4: Limitations of the Framework

While our goal is to provide a comprehensive framework and approach to AI, we also recognize several limitations of this framework. They are outlined below.

1. Due to the rate of advancement in AI, recent AI technologies may not be covered directly;
2. Each country has different policies regarding data collection and usage, which has required us to take a more broad approach in this framework;
3. Users may manipulate or use AI in a malicious manner that is in conflict with the goals of this framework;
4. Economic considerations are impossible to predict definitively, and portions of our analysis may prove inaccurate once applied to real world situations.

Further, we recognize that there may be other limitations of this framework, not explicitly stated.

Section 2: Artificial General Intelligence (AGI)

Lead Author: Noah M. Kenney
Second Author: Mirzakhmet Syzdykov

A heavily debated question in the field of artificial intelligence is what constitutes consciousness, and what additional responsibilities are raised in the development, deployment, and use of a conscious system. For purposes of this framework, we focus our analysis in this regard on Artificial General Intelligence (AGI), and we make the assumption that a true AGI will possess the characteristics of consciousness. We have outlined these key characteristics as below:

1. Breadth of knowledge: In comparison to specialized artificial intelligence, we expect that an AGI will display a high degree of knowledge on a wider range of subjects, on par with or better than that of an average human.
2. Common sense: An AGI will have the ability to operate by a set of unwritten rules which the majority of the population would consider universally necessary or generally beneficial.
3. Inference processing: Given imperfect information or user inputs without context, we expect an AGI to make inferences of what is unsaid.
4. Memory and contextual awareness: We expect an AGI to remember information about users and to tailor its outputs to match the expected needs of the aforementioned users using past inputs and contextual awareness.
5. Pattern and sequence recognition: An AGI should be able to recognize patterns or sequences in datasets or user inputs without needing to be asked to search for patterns or sequences. In this regard, we can test that the AGI is capable of some basic level of "sub-conscious" thinking or cognitive processing.

In the event of an AGI reaching consciousness according to thorough testing of the characteristics above, as well as other relevant characteristics, this framework requires that we still recognize and account for the following:

1. No network of neural nodes can produce the level of clarity of consciousness that a human being can. As such, we must recognize the lack of psychological congruency with the human mind, despite congruency of cognitive processing and communication abilities.
2. Psychology and self ("I") of any mind cannot be followed from the chaos produced by any neural network. Thus, any AGI should be considered nothing more than a computational tool without the modus operandi and without the perspective of a human.
3. An AGI may appear to express empathy, sympathy, or emotions; however, neural networks are not capable of experiencing feelings. The capabilities of AGIs should be considered comparable to neurons in the human brain, not the human being as a whole.
4. AGIs are optimized to specific goals, which may not align with humanity or with interests that preserve the wellbeing of humans and human rights.

Section 4.1: Clarity of Consciousness

In regard to artificial consciousness, the following apply:

1. An AGI may generate new or original information, which shall not be considered universal sources of truth without sufficient human verification and acceptance;
2. An AGI, by default, will over emphasis the use of statistics and data and underemphasize personal experiences and worldviews. This should be algorithmically accounted for in the development of an AGI.
3. To make users aware of potential biases, a report covering the demographics (age, gender, race, socioeconomic status, etc.) and key personal beliefs (religious, political, etc.) of those involved in the development of the AGI shall be published with each release. All data included in the report shall be included anonymously.

Section 4.2: AGI as a Tool

Throughout history, we observe the following truth about tools: Tools are utilized when needed by the user of the tool, at the control of the user. We expect the same to hold true for AGI. Thus, the following standards apply to the development of an AGI:

1. An AGI may not "wake up" without a user's knowledge and express permission;
2. Express permission may be the result of a verbal command only if said verbal command is not part of ordinary language;
3. All physical AGIs (i.e., any AGI that takes up space in the physical world) must have a visual indicator (i.e., an LED light) that clearly displays when the AGI is listening to user input;
4. An AGI shall make no efforts to act upon predicted user input or desires, without express opt-in from the user.

Under classification of AGI as a tool, it is necessary that humans are capable of understanding the outputs or decisions of the AI. Thus, in essence, one primary goal in reference to the use of AGIs is transparency of computational cognitive processing. This can be achieved in multiple ways; however, this framework only finds one acceptable method, which is the use of Explainable Machine Learning (XML), often referred to as Interpretable AI or Explainable Artificial Intelligence (XAI). For purposes of this framework, we refer to all three terms as interchangeable.

Explainable Machine Learning relies on input data represented as a matrix of sorted elements by the category and priority of each row and column. We have outlined the steps of building this matrix and evaluation of the query below:

1. Sort the input data matrix with the given priority of each factor;
2. Query and result each piece of data in the sorted matrix for the short-coming range of lower and upper row;
3. For numerical values, calculate the floor and ceil of the value and compare it in the same manner of the word entries contained in the input matrix.

Neural networks which are based upon sigmoid function can be extended for XML as the algorithm for arbitrary function with respect to the various probability types, which are well-known in the modeling theory of mass-serving queues.

1. Define the function f(x) and its range, L and R;
2. Compute the minimum and maximum values of the function f(x) of this range using differential calculus;
3. Train the neural network by applying the function f(x) divided by the minimum and maximum, subtracting the minimum.

The steps above give the novel example of the arbitrary learning without usage of sigmoid function for neural network training.

Section 4.3: Feeling vs. Cognitive Processing

AGIs are generally designed to operate on the basis of data, facts, and other quantitative information. However, qualitative analysis on the basis of personal experience, feelings, and emotions is innately human. In an effort to align AGIs in this regard, we find that the following apply:

1. An AGI shall factor in the concept of negative disparity, in which humans focus far more on the possibility of experiencing negative feelings or emotions (pain, fear, etc.) than positive ones. Note: In some cases, this may produce outputs that prove contrary to rational thought, aligning more closely with human cognitive processing.
2. An AGI shall always disclose to users its inability to experience feelings and its bias toward quantifiable data.