VERMONT

State of Vermont

Artificial Intelligence Code of Ethics

1 Introduction

1.1 Authority, Precedence

The Agency of Digital Services is directed to create a Code of Ethics pertaining to the use of Artificial Intelligence within the State of Vermont in 3 V.S.A. § 5012(b)(1) and (c)(3). If laws or regulations require more stringent requirements than stated in this code, agencies and departments may develop policies with the consent of the Agency of Digital Services, provided nothing in the requirements and standards in such policies is construed as diminishing or weakening those in this code.

1.2 Definition

For the purposes of this code, "Artificial Intelligence Systems", "Systems employing Artificial Intelligence", "Systems", and "AI" are used interchangeably to refer to the systems defined in 3 V.S.A. § 5011. For clarity, any system as defined in that section, whether it is implemented as a learning system or as a pre-trained model, is an Artificial Intelligence System for the purposes of this code.

This code differentiates among systems that make decisions directly impacting the provision of services or the freedoms of Vermonters ("direct impact" systems) and systems whose decisions improve the efficiency of State processes ("indirect impact" systems).

For the purposes of this code, the "State of Vermont" refers to the executive branch of Vermont's State Government.

1.3 Purpose

Artificial intelligence is a rapidly expanding field of research and new applications of AI are constantly being developed. The State of Vermont seeks to enable the use of AI systems within State Government to improve quality of services the State delivers to Vermonters and to enable the State of Vermont to continue to be a leader in the thoughtful application of AI technologies in government.

This code is intended to establish minimum requirements for AI systems in use by the State of Vermont. While both direct impact and indirect impact systems are prone to similar kinds of challenges, direct impact systems deserve an increased level of scrutiny prior to and after implementation commensurate with their potential to impact Vermonters.

1.4 Scope

This code establishes requirements and standards for the Executive Branch of the State of Vermont.

2 Use of Artificial Intelligence

Artificial Intelligence shall be used by the State of Vermont to further the goals set out in Vermont's Constitution: for the common benefit, protection, and security of the people.

In general, Artificial Intelligence Systems should be used in a human-centered way that recognizes the dignity and value of all persons and their contributions to society. AI Systems must always be used in ways that comply with State and Federal Laws and maintain the rights of persons as enshrined in those laws.

The following broad use cases are provided as a general framework for where AI Systems should or should not be applied. These are intended as general principles, not to explicitly allow or disallow any particular AI system.

2.1 Appropriate Uses of Artificial Intelligence

Within the State of Vermont, AI should be used in ways that:
- Maintain the freedoms and liberties of Vermonters
- Are Transparent and Maintain trust in Vermont's institutions
- Improve the delivery of services to Vermonters
- Make work more creative and meaningful for Vermonters and their Public Servants
- Improve efficiency in government

2.2 Inappropriate Uses of Artificial Intelligence

Within the State of Vermont, AI should not be used in ways that:
- Limit the rights of Vermonters or monitor the free exercise of those rights
- Create new capabilities for government
- Make decisions without human intervention that could have a significant impact on individuals or groups
- Create confusion about the capabilities of the systems being used or the roles of the people involved in using them

Recognizing that AI Systems operate in the context of human systems, processes, and environments, it is also inappropriate to use AI Systems in ways that make overly optimistic assumptions about their capabilities, such as that they are free of bias or less fallible than humans.

2.3 How AI Systems should be Created and Implemented

AI systems should be implemented in ways that recognize they are part of human systems, and that while in use within that system they may be subject to conditions different from the intended operating environment. Additionally, these human systems may change over time.

Every phase of the creation and implementation of systems employing artificial intelligence should seek to ensure AI systems behave ethically and avoid the risk of unexpected outcomes. For direct impact systems it is important that there is continual monitoring to ensure the outcomes of the system are as expected.

2.4 Conflicts of Interest

If an artificial intelligence system is provided by an entity which has a reasonable expectation of gaining a material benefit from the State's use of the system – whether for that entity, their partners, or 3rd parties – such benefits will be described in procurement documentation and any

contract. This includes financial and non-financial benefits such as access to data about use, users, or Vermonters, recommendations that may increase remuneration to the entity, or any other material benefit.

The State will consider the impact of these actual or potential conflicts of interest during the procurement process and will mitigate the impact of such potential or actual conflicts of interest as far as reasonably possible.

3 Characteristics of AI Systems

The sections below apply to all AI systems except where explicitly stated.

3.1 Bias and Representativeness

Systems employing artificial intelligence should be tested and continually monitored for disproportionate impacts to any group or individual based on group membership.

Bias can be introduced at many layers of an AI System, including in the data selected for training, the algorithm itself, and the way users interact with the system.

Bias relating to group membership, often known as cultural bias, is a crucial type of bias to monitor and test for. It usually stems from the dataset itself. There are many other kinds of bias that can show up in a data set, including but not limited to historical, temporal, and aggregation bias.

AI systems should be trained and tested using data that is representative of the types of cases it will encounter in production. For direct impact systems, this means ensuring training materials and testing include cases representing cohorts and personas of the populations served. For indirect impact systems, representativeness means including samples of different cases that may occur in production, including edge cases.

The remainder of this subsection are requirements for Direct Impact Systems and strongly preferred for any system employing AI.

Care should be taken that selected inputs do not bias the output in favor of or against any group, unless such biases are intentional and clearly documented.

Systems with known biases should be reviewed and have a remediation plan put in place to ensure adequate safeguards are in place to prevent inappropriate outcomes. The remediation plan may include decommissioning or replacement of systems where adequate safeguards are impractical.

3.2 Transparency and Explainability

It is crucial that AI Systems used by the State of Vermont build public trust in Government's processes and oversight. A core challenge of many AI systems is their "black box" nature – they are trusted because of their strong capabilities and performance in testing scenarios, but in government use cases black box AI poses a risk both of negative outcomes and to public trust. AI systems must be implemented and used in a way that instills public trust.

The remainder of this subsection are requirements for Direct Impact Systems and strongly preferred for any AI system.

Transparency refers to having full, clear documentation of how the AI was developed, any known issues in the training datasets employed, the training process, what testing was performed and its findings, how it is implemented, and the impact of its decisions. The documentation should clearly define the ways in which the AI system is intended to be used, and any known weaknesses. The lineage of and any assumptions about data provided as inputs to the system should be clearly documented. Any data and decisions created by the AI system must have their limitations and scope of use documented. Additionally, the data elements described in 3 V.S.A. § 3305 b must be provided.

Explainability refers to the ability of the AI system to explain how a decision made or result produced to a layperson. AI Systems should be designed in a way that a decision can be clearly explained and justified. Additionally, unless business requirements dictate otherwise, decisions provided to Vermonters should have a point of contact for review or appeal.

Any modifications to AI systems must be made with the consent of the State of Vermont and a changelog with any known changes to behavior noted, especially as it new or known weaknesses in the system.

3.3 Autonomy

In any cases where an AI system is capable of making decisions autonomously, those decisions must comply with Vermont and Federal Law, regulations, policy, and this code.

In general, the State of Vermont seeks to encourage the use of AI systems that support and enhance human-led processes. Autonomous systems are only acceptable in the case where an incorrect decision would have a small and transient impact. Examples of low-risk direct decision systems include use cases like chatbots, recommendation engines, and expediting of "happy path" applications or forms.

Autonomous capabilities need to be clearly documented and implemented in narrow contexts. Direct Decision systems that make decisions autonomously must include a point of contact for support, review, or appeal.

In all cases, autonomous decision systems must include more regular and comprehensive outcome monitoring (see next section).

4 Outcomes

Artificial Intelligence is not capable of reducing poor outcomes merely because it is added to a process. However, AI systems are expected to maintain or improve outcomes when used thoughtfully.

AI systems must be considered in the context of the human systems they exist within. Sometimes AI systems may operate correctly, but unexpectedly exacerbate existing issues due to the context

they operate within or changing conditions in their environment. The outputs of the AI system and the outcomes of the full process must be continually monitored to ensure the desired goals are being met and no adverse impacts are identified.

4.1 Accountability and Trustworthiness

Within the State of Vermont, every decision made by a System using Artificial Intelligence must have a human accountable for it. This will usually be the division or department head responsible for the process the AI system supports. The organization within the State of Vermont using the AI system is responsible for testing and monitoring the system to ensure consistently good outcomes.

Developers of Systems using Artificial Intelligence are responsible to ensure they have met the requirements of this code, applicable laws, regulations, and other policies; and relevant standards and best practices in their fields.

The Agency of Digital Services is responsible to ensure that Systems using Artificial Intelligence developed and procured for use within the State certify compliance with the provisions of this code and applicable laws and regulations, and will support organizations within the State of Vermont to test and implement AI Systems.

Any roles and responsibilities specific to the system especially as it relates to Artificial Intelligence components must be clearly defined, documented, and enforced.

AI systems must be used in ways that justify the trust Vermonters place in their government. Systems must take a human-centered approach to decision making. Systems must operate in ways that are clearly just and protect the rights of Vermonters. See "Transparency and Explainability" above.

4.2 Impact, Fairness, and Discrimination

The impacts of decisions made my AI systems must be clearly documented and monitored to ensure they are proportionate and that the rights, liberties, and interests of Vermonters and the State of Vermont are protected. See "Autonomy" above for further considerations on AI system impacts.

The design of AI systems, standards for implementation, and procedures for use must protect the rights of persons and groups affected by their decisions.

Processes using direct impact systems must have components that continually monitor both the outputs and outcomes of AI decisions to ensure fair and just outcomes are achieved, and that decisions are not having disproportionate impacts on any group.

AI systems should be designed to be resistant to manipulation through the alteration of inputs that should not have bearing on the decision being made.

4.3 Confidentiality and Privacy

The design of AI systems, standards for implementation, and procedures for use must not adversely affect the privacy rights of Vermonters or the privacy of the State of Vermont. Data and decisions must be stored securely and in compliance with applicable laws, regulations, and policies. People, roles, and groups with access to data or decisions must be clearly documented.

In general, AI systems should avoid operating on confidential data whenever possible.

4.4 Safety and Security

The design of AI systems, standards for implementation, and procedures for use must not decrease the overall safety and security of Vermonters or the State of Vermont. AI systems must prevent the system from being altered and the data in the system from being accessed or modified by unauthorized persons.