PRINCIPLES FOR
Trustworthy AI in Recruiting & Hiring

© The Center for Industry Self-Regulation, 2023. All Rights Reserved. BBB National Programs Charitable Foundation d/b/a Center for Industry Self-Regulation.
I. Goals for the Use of Artificial Intelligence (AI) in Recruiting and Hiring

Algorithmic systems using AI technologies, when responsibly designed, deployed, and monitored, can promote fairer outcomes and inclusivity by reducing the potential for bias that may exist in human-led decision-making during the recruiting and hiring process. The development and implementation of trustworthy AI therefore should focus on the following goals:

1) Ensuring systems are valid and reliable
2) Promoting fair outcomes, with harmful bias managed
3) Increasing inclusivity
4) Facilitating compliance, transparency, and accountability
5) Striving for systems that are safe, secure, resilient, explainable, interpretable, and privacy-enhanced

II. Purpose of the Principles

Employers must be committed to robust, achievable, and fair standards for the use of AI in recruitment and hiring. Practical baseline standards for trusted business practices are necessary to provide appropriate guidance for Employers and Vendors seeking to properly leverage AI technology in recruitment and hiring processes, provide protection for Applicants and employees, send marketplace signals to showcase best-in-class examples of responsible implementation and utilization, and demonstrate to regulators that new technologies are being developed and deployed responsibly.

Employers will evaluate the potential benefits and societal risks associated with these methods, as they may have the ability to identify conscious and unconscious human biases and barriers to inclusion. The ultimate objective is to promote positive changes in the hiring process. Important decisions are subject to human direction and review as necessary, the extent of which depends on the application and the impact it has on the decision-making process (e.g., final selection, filtering selection, testing, etc.).

III. Scope

These self-regulatory principles serve as a global baseline standard for the use of artificial intelligence applications by employers for the purpose of recruitment and hiring. The Principles are meant to inform best practices when advanced adaptive algorithms are used as part of the employment selection process. Excluded from scope are other Employer activities that may make use of AI applications, including Passive Hiring, as defined below, and non-Selection Activities. Also excluded from scope are those Selection Activities that make use of simple algorithms that do not meet the definition of AI Application below (e.g., functions in a spreadsheet). The Principles do not supersede or replace national or local laws, regulations, or best practices. Organizations (both Employers and Vendors) that commit to these principles agree to apply them as a consistent minimum set of practices across their global operations.

IV. Definitions

A. Applicant: Individual who has expressed and maintained interest in a specific position in the manner specified by the Employer’s application requirements.

B. Artificial Intelligence (AI) Application: an algorithmic system that includes machine learning or deep learning techniques—or future technologies with similar complexity—and is used as part of a Selection decision.

C. Artificial Intelligence (AI) Processing: the use of an AI Application.

---


2 The Principles are compatible with a holistic governance approach to AI risk management, such as that defined in the NIST AI RMF 1.0.
D. Employer: organization seeking to hire individuals.

E. Passive Hiring: the process of waiting for candidates to respond to job postings or career pages; the Employer does not affirmatively seek to identify relevant and qualified non-Applicants to fulfill a current or future staffing need.

F. Personal Data: information about a natural person or data linked to such information, where processed to uniquely identify a natural person.

G. Recruiting Activities (or “Sourcing”): the process of identifying, researching, and networking with potential job candidates with the intent to generate Applicants.

H. Selection Activities: the process of evaluating Applicants for existing staffing needs and selecting which Applicants move forward in the hiring process or, ultimately, are selected for hire. (E.g., reviewing submitted resumes or screening candidates through mechanisms such as interviews and/or assessments.)

I. Sensitive Personal Data: Personal Data that reveals information with enhanced legal protections (e.g., racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, or biometric data.).

J. Vendor: company that creates/markets AI Applications for end-user use by Employers.

V. Transparency

A. Objective

For Applicants: If AI Applications are being used, Applicants should be able to make an informed choice about whether to proceed at the appropriate stage in the recruiting or hiring process.

For Employers: Employers using AI should partner with Vendors to understand how the AI Application operates and to ensure the AI Application is designed and maintained consistent with the other self-regulatory principles.

B. Employers: General Notice if the Use of AI Processing

1. Timing of the notice. Employers should provide Applicants with an opt-out notice that AI Processing will be applied to personal data collected about them. Such notice may be provided to the Applicant at the time of application, or in any case prior to the Employer’s utilizing AI Processing to Select/Hire the Applicant.

2. Content of the notice. The initial notice of AI Processing should be written in clear, plain language and should link to the Employer’s privacy policy or other policy describing the AI Processing. Following is a model template that can be tailored by the Employer to fit specific use scenarios:

   To help us select the most qualified candidates from the many applications we receive, while reducing the potential risk of human bias, we use AI processing to evaluate your application. The AI processing used meets or exceeds the standards for fairness and inclusivity established in the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES. We may use information you submit to consider you for other positions as well. For more information, see [company privacy

---

3 The opt out described in this sample notice is meant to be separate and distinct from any ADA accommodation obligations. See Section VI(B).

The Center for Industry Self-Regulation (CISR) is a 501(c)3 non-profit organization (FEIN: 84-5021924) and is a DBA of BBB National Programs Charitable Foundation. 1676 International Drive, Suite 550 | McLean, VA 22102.
C. Employers: Detailed Notice of the Use of AI Processing

In its privacy policy, a dedicated notice, or communication from HR, the Employer should specify, in clear, plain language:

1. Whether its AI Processing meets the Principles for Trustworthy AI in Recruiting and Hiring and the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES; or similar protocols as warranted.

2. The Personal Data elements, tasks, and inferences subject to AI Processing with sufficient detail to enable the Applicant to determine whether an accommodation is necessary and consistent with current legal accommodation requirements, and/or providing an opt-out option.

3. The sources of data elements subject to AI Processing, if not collected from the Applicant.

4. In accordance with disability regulations, the option for applicants to request an accommodation to proceed (separate from the general opt-out option).

5. Any choices the Employer provides to limit or review AI Processing, such as human review of outputs.

6. The functions of the AI Processing, such as the stages of the selection process at which it may be applied (e.g., resume review, assessment stage of the hiring process, etc.) the modalities through which it operates (text, images, audio), and the dimensions being assessed.

7. The nature of any specific automated decisions made through AI Processing (e.g., behavioral or skill assessments, video interviewing).

8. Whether the purpose of the AI Processing is limited to recruitment and selection.

9. Whether data will be used for future hiring decisions on other job openings.

D. Employers: Periodic Review of AI-Enhanced Outputs

1. Employers commit to appropriate examination of outputs where AI-enhanced methods are used for Selection Activities, including AI Processing to promote fairer and more inclusive outcomes and to reduce the potential for biased results, consistent with the U.S. Uniform Guidelines on Employee Selection Procedures (41 CFR 60-3) or other applicable legal requirements. The examination will be appropriately tailored to the Employer’s individual circumstances and specific uses of AI Applications.

2. When making use of AI Applications developed by Vendors, Employers agree to appropriately vet Vendors using either:
   a. Employer’s own industry-standard internal validation processes or
   b. A requirement for all Vendors to demonstrate:
      (1) a third-party-certified process validation or

---

4 If the Employer wishes to provide an alternative selection process for those who opt-out from AI processing, they may do so consistent with these Principles, but it is not required. At a minimum, Employers commit to provide the choice to opt-out from consideration for those who prefer their personal data not be subject to AI processing.

5 In conducting a review, employers are encouraged to utilize a third party, or, if using an internal process, implement safeguards to ensure independence of the review.
(2) an internal validation by the Vendor designed to demonstrate basic fairness and to facilitate employer compliance with the U.S. Uniform Guidelines on Employee Selection Procedures (41 CFR 60-3) or other applicable legal requirements.

E. Vendors: Transparency about Commitment to Standards

In a dedicated notice to Employer or Applicant (as appropriate) the Vendor will specify:

1. Whether its AI Processing meets the Principles for Trustworthy AI in Recruiting and Hiring and the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES;

2. A list of the steps it takes to assure the inputs and functions of AI Processing have been evaluated to ensure a level of data set bias that is minimal and job related;

3. A list of the techniques it deploys to promote fair, achievable, and bias-neutral outputs; and

4. The specific purpose(s) for which the AI Application is designed to be used by Employers. Vendors are not expected to validate an Employer’s use of the AI Application, unless contractually obligated to do so.

VI. Fairness, Non-Discrimination, Technical Robustness, and Safety

Throughout the AI lifecycle (pre-design, design, development, and deployment) AI Processing should be evaluated with reference to three components: inputs, functions, and outputs. Optimally, organizations should rely on the scientific method and a holistic approach to testing, evaluation, validation, and verification (TEVV).

A. Inputs of AI Processing should be evaluated with the following general goals as applicable to the technology used:

1. Rely on scientifically rigorous approaches to AI Application development/techniques.

2. Ensure model data set integrity to reduce the impact of systemic biases and increase probability of non-discriminatory results.

3. To avoid the potential for discrimination by proxy, recognize that socially derived constructs (e.g., education, geography, group memberships, or resume gaps) may disadvantage groups.

4. Treat cultural differences fairly (e.g., eye contact, visual cueing, and accents).

5. Reduce computational and statistical biases, including selection bias (over- or under-representing populations) and reporting bias (under-reporting all available information).

6. Anticipate accommodations needed for various physical or cognitive disabilities and ensure proper mechanisms are in place to allow for such accommodations and that such accommodations do not disadvantage the Applicant.

B. Functions of AI Processing should be evaluated periodically for:


7 For an explanation and comparison of systemic bias, computational and statistical bias, and human-cognitive bias, and the role of mitigation efforts throughout the AI lifecycle, see id. at 18.

8 Id.
2. How the AI Application will be deployed and used to determine if reasonable accommodation for a disability under the Americans with Disabilities Act (ADA) or other applicable law is appropriate on a case-by-case basis, as needed.

3. Human-cognitive biases\(^9\) that can emerge while calibrating and interpreting models, such as confirmation bias (the tendency to search for, interpret, favor, and recall information in a way that confirms one’s pre-existing beliefs or hypotheses).

4. Providing sufficient safeguards to ensure data integrity and preventing accidental disclosure of Personal Data.

C. Outputs of AI Processing should be evaluated for, and adjustments made, if necessary, for:
   1. Discriminatory impact.
   2. Actual uses of the system conforming to expectations.
   3. Processes for monitoring, evaluating, and providing feedback including by engaging in ongoing monitoring, testing, and adjustment of AI Applications and processes, as described in Section VI above.
   4. Human decision-making controls at relevant stages.
   5. Safeguards in place to protect from disclosure consistent with jurisdictional privacy standards.

VII. Governance and Accountability

A. In General

Organizations that make use of AI Processing should implement robust mechanisms to manage risk\(^10\) and ensure data set, algorithm, and output integrity as well as rates of bias that meet or exceed baseline requirements.

Who is responsible?

1. Vendor: responsible for baseline reliable AI Application designed and tested to meet standards with clear disclosure of purposes for which the AI Application was designed.

2. Employer: responsible for proper AI Processing, execution, and validation of results, with the frequency of testing dependent on volume and other situational factors.

B. Vendor Responsibilities

1. Rigorous and industry-standard pre-testing and validation of the AI Application.

2. Appropriate confidence that AI Application meets or exceeds industry standards, including these Principles, based on, as appropriate and agreed by Employer and Vendor, internal testing, independent certification to the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES and/or third-party audit.

3. Reasonable basis for achievable outputs, limited to the described purpose for which the AI Application designed.

4. Scheduled third-party review (essential for open-loop learning models; recommended for closed-loop models).

\(^9\) Id.

\(^10\) For an explanation of risk in relation to AI systems, see NIST AI RMF 1.0 at 38 (Appendix B: “How AI Risks Differ from Traditional Software Risks”).
5. Educate Employers about results of any prior validations or third-party reviews to ensure that the Employer is informed of best practices to operate the AI tool effectively and in accordance with these self-regulation principles.

C. Employer Responsibilities

1. Confirm that Vendor meets or exceeds industry standards, including these Principles, through contractual controls (specific conditions detailed prior to engagement as necessary for the type of AI Application being used).

2. Ensure that AI Applications are appropriately certified to the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES, meet equivalent criteria, or other appropriate protocols under relevant law.

3. Employers are responsible for biased outcomes, but they can mitigate potential damages through a robust vetting and use of AI-based systems, including by engaging in ongoing monitoring and testing of AI Applications and organizational processes, as described in Section VI above.

D. Independent Certification

To ensure consistency and foster trust in the workforce, participating Employers may require Vendors to demonstrate independent certification of AI Applications against the parameters specified in the INDEPENDENT CERTIFICATION PROTOCOLS FOR AI-ENABLED HIRING AND RECRUITING TECHNOLOGIES.

The accompanying protocols specify the criteria for certification and the requirements for qualifying as a third-party certifying entity.

The independent certification process is designed to achieve the following goals:

1. AI Applications are only deployed after industry standard processes for testing and validation have been completed.

2. Uniformity in disclosures and communications between Vendors and Employers.

3. Employers that are properly informed about the purposes for which AI Applications have been developed and tested.

4. Enhanced levels of testing and accountability when appropriate (including when Sensitive Personal Data is processed in an AI Application).

5. Indicators to assist applicants in verifying whether Employers abide by industry standard practices for the use of AI Applications.
About the Center for Industry Self-Regulation

The Center for Industry Self-Regulation (CISR), BBB National Programs’ 501(c)(3) nonprofit foundation, was created to harness the power of independent, industry self-regulation to address the marketplace trust challenges businesses face today. CISR is dedicated to education and research that supports responsible business leaders developing fair, future-proof best practices, and to the education of the general public on the conditions necessary for industry self-regulation.

About BBB National Programs

BBB National Programs, a non-profit organization, is the home of U.S. independent industry self-regulation, currently operating more than a dozen globally recognized programs that have been helping enhance consumer trust in business for more than 50 years. These programs provide third-party accountability and dispute resolution services that address existing and emerging industry issues, create a fairer playing field for businesses, and a better experience for consumers. BBB National Programs continues to evolve its work and grow its impact by providing business guidance and fostering best practices in arenas such as advertising, child-and-teen-directed marketing, data privacy, dispute resolution, automobile warranty, technology, and emerging areas. To learn more, visit bbbprograms.org.