

BUILDING CONFIDENCE & TRUST IN ARTIFICIAL INTELLIGENCE SYSTEMS

Advances in software-enabled data analytics and artificial intelligence (AI) are fueling job and economic growth in the United States and around the world, producing societal gains, and transforming how businesses operate, how governments provide important services, and how consumers perform daily tasks. As we seek to maximize these economic and societal benefits, we also recognize the importance of raising awareness about how AI systems operate. This issue snapshot provides information about both the benefits of AI and how companies can enhance understanding of their AI systems.

Benefits of AI

At their core, AI systems augment human intelligence. Specifically, BSA members' AI systems primarily assist in the analysis of data to find connections that can improve the quality and accuracy of human decision-making. There are myriad examples across a wide swath of industry sectors that illustrate the benefits of AI technologies.

- » **Accessibility.** AI systems, powered by data analytics, are at the heart of new devices and applications that can improve the lives of people with disabilities. For instance, AI is helping people with vision-related impairments interpret and understand visual content, such as photos and their physical surroundings. This technology opens new possibilities for people with vision impairments to navigate the world, giving them increased independence and greater ability to engage with their communities.
- » **Healthcare.** AI systems are helping hospitals provide more affordable healthcare while achieving better patient outcomes. Physicians can now harness AI to analyze large volumes of data, including patient information and medical test results, to identify the best possible treatment options for patients.

Whether it is improving health outcomes, detecting financial fraud, or enhancing educational programs, AI systems have the potential to provide immense economic and societal benefits.

- » **Financial services.** AI is improving fraud detection by recognizing suspicious behavior and providing companies with real-time information that helps them identify and investigate different types of fraud, reducing the losses attributed to fraudsters by billions of dollars.
- » **E-Commerce.** AI is enabling consumers to transact business online quicker and more securely by making it easier to authenticate identity and reduce the friction needed to establish trust in digital transactions of all kinds.
- » **Government services.** AI is helping governments improve constituent services in ways that save time, money, and lives. For example, cities are optimizing medical emergency response processes using AI-based systems, enabling them to more strategically position personnel and reduce both response times and the overall number of emergency trips.
- » **Education.** AI is helping educators quickly identify students that need particular attention and develop personalized lesson plans that tailor instruction, content, pace, and testing to individual students' strengths and interests.

Whether it is improving health outcomes, detecting financial fraud, or enhancing educational programs, AI systems have the potential to provide immense economic and societal benefits.

[more >>](#)

Principles to Improve Understanding of AI Systems

AI systems, like software, often use sophisticated algorithms. An algorithm is a systemized method or set of instructions that collects inputs and provides outputs according to a well-defined rule. The algorithms used in AI are often particularly well-suited to analyzing massive volumes of data from many different sources, and accounting for an enormous number of variables that may interact in complex and unexpected ways. Through this analysis, AI solutions can enhance perception, learning, reasoning, and decision-making aimed at improving the ability of people to solve complex and challenging problems, as humans alone can neither comprehend the same volume of data nor detect the same complex, unexpected interactions.

In some instances, the complexity of these technologies, which are designed to identify patterns that we could not otherwise extrapolate, makes it challenging to explain how certain results were achieved. As these technologies are increasingly deployed, BSA companies recognize the importance of increasing awareness of AI systems and providing meaningful information to enhance consumer understanding of these systems. At the same time, research has shown that disclosing the algorithms, source code, or associated data sets is ineffective in helping to provide explanations, in part because they cannot be meaningfully understood in isolation.

BSA therefore supports industry efforts to provide users of AI systems with the information necessary to instill confidence that such systems are operating as intended. Facilitating increased understanding and promoting trust in the use of AI technologies is an important priority. As companies explore how to enhance understanding of AI systems, BSA highlights below principles that could aid in this effort:

- » **Fairness.** Consider measures to evaluate AI systems to help recognize improper or unconscious bias.
- » **Accuracy.** Acknowledge the importance of data quality and, where feasible, identify sources of error in data inputs and system outputs.
- » **Data provenance.** Consider measures that could facilitate evaluation and documentation of data used to train AI systems, how data is collected, and how data is used over time within AI systems, consistent with any other data retention obligations.
- » **Explainability.** Explore how to provide reasonable explanations of how AI systems operate.
- » **Responsibility.** Consider whether processes are available to address unexpected issues that may arise after AI products and services are deployed.

Importantly, there are many forms of AI, which vary in complexity, and are used to enable different services. As a result, there is no one-size-fits-all approach. Instead, considering the variation in technologies and applications of AI, and ongoing research efforts, there are many possible ways to enhance understanding and build trust in AI systems, and they may overlap. For example, such practices could include:

- » Conducting in-house testing and evaluation of AI systems to ensure they meet their specified goals;
- » Developing guidelines and providing necessary resources to developers to help evaluate fairness and guard against improper bias;
- » Identifying persons with relevant expertise who are responsible for addressing significant problems identified with operating AI systems;
- » Ensuring subject matter experts, especially those with knowledge of the policy landscape in which the AI system will be deployed, are available to assist computational scientists in the design and implementation phases;
- » Providing descriptions of procedures used to assess the quality of data inputs and address errors identified in outputs;
- » Providing general descriptions, where appropriate, of training datasets that AI systems use to learn;
- » Developing mechanisms for consumers to request information, obtain guidance, and address potential concerns;
- » Continuing to monitor systems after product release to detect and address unintended outcomes;
- » Providing visual aids and/or plain language explanations that communicate important facts about AI systems and their operation; and/or
- » Supporting continued research and analysis of transparent modeling.

The development of AI technologies is growing and evolving rapidly, and relevant technical tools and processes that could aid in accomplishing some of these tasks are still being developed. As a result, as companies seek to address these important issues, it is vital that they maintain sufficiently flexible approaches to adapt to the changing landscape as these emerging technologies, and potential solutions to new challenges, continue to evolve.