How Can We Balance AI's Potential And Ethical Challenges?

23-04-2025 ~ While AI is transforming industries with powerful capabilities, challenges like data quality, bias, transparency, and privacy concerns must be addressed to ensure fairness and accuracy, especially in areas like fraud detection.

Artificial intelligence is transforming industries by automating processes, improving efficiency, and detecting patterns that humans might miss. However, as AI continues to evolve, so do the challenges associated with its implementation. Issues such as data quality, bias, transparency, and privacy concerns raise critical ethical questions. Ensuring that AI operates fairly and effectively requires continuous improvement and careful oversight, especially in sectors such as insurance, where accuracy and trust are crucial.

The primary issues of concern include:

Low data quality: The effectiveness of AI largely depends on the quality of the data it uses. If the data is inaccurate or incomplete, the AI's performance will suffer.

Data quality is crucial in artificial intelligence because it directly impacts AI models' performance, accuracy, and reliability. Poor data quality is the primary obstacle to deploying and executing artificial intelligence and machine learning projects and operations. "Garbage in, garbage out" (GIGO), a concept familiar to computer users for generations, is just as applicable to AI. If an AI model is of poor quality, inaccurate, or irrelevant, the system's output will also be of poor quality, inaccurate, or irrelevant.

Even the most sophisticated AI algorithms can produce flawed results, leading to poor performance and failure. A high-quality AI model should aim for accuracy, consistency (meaning that the data follow a standard format and structure to facilitate processing and analysis), completeness (to avoid missing essential patterns and correlations), timeliness, and relevance.

To ensure the AI model is efficient, developers need to collect relevant data, which depends on the choice of sources from which to draw the data. This challenge is compounded by the need to maintain quality and standards to eliminate duplicate or conflicting data. Then, the data must be labeled correctly, a process that can be time-consuming and prone to errors. At the same time, data must be stored to prevent unauthorized access and corruption. Data poisoning is another risk: it refers to a deliberate attack on AI systems, where attackers inject malicious or misleading data into the dataset, resulting in unreliable and even dangerous outputs.

Bias in AI Models

Sometimes, AI can be biased, meaning it might unfairly treat certain groups of people differently. For example, if an AI system is trained on biased data, it may make decisions that discriminate against specific individuals based on factors such as race, gender, or other characteristics.

There are two basic types of bias: explicit and implicit. An explicit bias refers to a conscious and intentional prejudice or belief about a specific group of people. An implicit bias operates unconsciously and can influence decisions without a person realizing it. Social conditioning, the media, and cultural exposure all contribute to these decisions.

Algorithmic bias can creep in because of programming errors, such as a developer unfairly weighting factors in algorithm decision-making based on their own conscious or unconscious biases. For example, indicators like income or vocabulary might be used by the algorithm to discriminate against people of a certain race or gender unintentionally. People can also process information and make judgments based on the data they initially selected (cognitive bias), favoring datasets based on Americans rather than a sampling of populations worldwide.

Bias in AI is not merely a technical issue but a societal challenge, as AI systems are increasingly integrated into decision-making processes in healthcare, hiring, law enforcement, the media, and other critical areas. Bias can occur in various stages of the AI pipeline, especially with data collection. Outputs may be biased if the data used to train an AI algorithm is not diverse or representative of the actual data. For instance, training that favors male and white applicants may result in biased AI hiring recommendations.

Labeling training data can also introduce bias since it can influence the interpretation given to the outputs. The model itself might be imbalanced or fail to consider diverse inputs, favoring majority views over those of minorities. To make AI more accurate and fairer, researchers need to retrain it regularly. Companies, especially insurers, must ensure that they use accurate, complete, and up-to-date data while also ensuring their models are fair to everyone.

Transparency

Transparency is a key issue, as it can be challenging to explain how AI makes its decisions. This lack of clarity can be a problem for both customers and regulators who want to understand how these systems work. Transparency in AI is essential because it provides a clear explanation for why AI's decisions and actions occur, allowing us to ensure they are fair and reliable.

Using AI in the workplace can help with the hiring process, but understanding how AI does so without bias can only be achieved if it is transparent. As AI becomes increasingly important in society, business, healthcare, the media, and culture, governments and regulators need to establish rules, standards, and laws that ensure transparency in the use of AI.

Transparency is closely related to Explainable AI (XAI), which allows outsiders to understand why AI is making its decisions. Such explainability builds customer trust. This is referred to as a glass box system, as opposed to a black box system, where the results or outputs from AI are transparent and the reasons for their decisions are known, sometimes even to the system's developer.

Errors in AI Predictions

AI can sometimes produce incorrect results, known as false positives or false negatives. This happens because the data used to train AI systems is often imperfect, leaving room for errors. It's human nature to overestimate a technology's short-term effect and underestimate its long-term effect. This tendency certainly applies to AI predictions. The question, of course, is how long the long run is.

The rise of generative AI confronts us with key questions about AI failure and how we make sense of it. As most experts (and many users) acknowledge, AI outputs, as astonishing and incredibly powerful as they can be, may also be fallible,

inaccurate, and, at times, completely nonsensical. A term has gained popularity in recognition of this fallibility—"AI hallucination."

The scholar and bestselling author Naomi Klein argued in an article for the Guardian in May 2023 that the term "hallucination" only anthropomorphized a technical problem and that, "by appropriating a word commonly used in psychology, psychedelics and various forms of mysticism, the tech-industry is feeding the myth that by building these large language models, we are in the process of birthing an animate intelligence."

Nonetheless, all major AI developers, including Google, Microsoft, and OpenAI, have publicly addressed this issue, whether it is called a hallucination or not. For instance, an internal Microsoft document stated that "these systems are built to be persuasive, not truthful," allowing that "outputs can look very realistic but include statements that aren't true." Alphabet, the parent company of Google, has admitted that it's a problem "no one in the field seems to have solved." That means AI outputs cannot be entirely relied upon for their predictions and need to be verified by reliable sources.

Privacy Concerns

In many cases, real data cannot be used to train AI due to privacy issues. Instead, fake data is created based on real data, which can lead to inaccuracies and lower performance in the AI system.

AI privacy refers to the protection of personal or sensitive information that is collected, used, shared, or stored by AI systems. One reason AI poses a greater data privacy risk than other digital technology is the sheer volume of information AI needs to be trained on: terabytes or petabytes of text, images or video which often includes sensitive data such as healthcare information, personal data from social media sites, personal finance data, and biometric data used for facial recognition.

As more sensitive data are being collected, stored, and transmitted, the risks of exposure from AI models rise. "This [data] ends up with a big bullseye that somebody's going to try to hit," Jeff Crume, an IBM Distinguished Engineer, explained in an IBM Technology video.

Data leakage from an AI model can occur through the *accidental exposure* of

sensitive data, such as a technical security vulnerability or procedural security error. Data exfiltration, on the other hand, is the theft of data. An attacker, hacker, cybercriminal, foreign adversary, or other malicious actor can choose to encrypt the data as part of a ransomware attack or use it to hijack corporate executives' email accounts.

It's not data exfiltration until the data are copied or moved to some other storage device under the attacker's control. Sometimes, the attack may come from an insider threat—an employee, business partner, or other authorized user who intentionally or unintentionally exposes data due to human error, poor judgment, ignorance of security controls, or out of disgruntlement or greed.

The Future of AI in Fraud Detection

As AI technology improves, it is expected to become more effective at detecting and preventing complex fraud. For example, let's consider phone insurance fraud. Phone insurance fraud, also known as device insurance fraud (because it can refer to fraud involving laptops and tablets as well as smartphones), occurs when someone intentionally makes a false claim on their device's insurance company, falsely asserting that their device was lost, stolen, or damaged, or exaggerating the extent of the damage.

One survey showed that 40 percent of all insurance claims are fraudulent. For companies, fraud can result in significant losses and increase the cost of premiums for consumers. Rates of fraud incidents have increased significantly. A survey by Javelin Strategy Research found that fraudulent claims on mobile phones increased by 63 percent between 2018 and 2019.

Phone theft has also become more sophisticated and organized, leading to phishing attacks and social engineering used to access stolen devices and perpetrate fraud and false claims. In some instances, phone owners will buy multiple policies on the same phone and then claim theft, loss, or damage to obtain money for the phone from numerous insurance providers.

There's another type of fraud to be aware of. According to Jonathan Nelson, director of product management for Hiya, an insurance and finance provider, insurers need to be mindful of how their customers are being misled or unwittingly targeted. "The most common thing that you'll experience when you're becoming a victim of… an automobile, insurance, or warranty scam, is what we

call illegal lead generation. Effectively, the goal is to manipulate the recipient into signing up with a different third-party insurance company [that] may or may not be aware of the fact that their new customers are coming through this illegal sort of scam-like channel."

AI to the Rescue

One promising development is the use of deep learning models, which can quickly compare new insurance claims against millions of past claims. These models look for unusual patterns that might suggest fraud, such as strange damage descriptions, multiple claims from the same person, or inconsistencies in location data.

These advanced models don't just follow fixed rules; they learn and improve with every new piece of data they analyze. For example, they can examine pictures of damaged phones, compare them with large databases, spot signs of image manipulation, and assess whether fraud may be involved.

The Internet of Things will enhance these fraud prevention efforts by connecting data from various devices, such as smartphones and wearables. This will allow insurers to gather real-time information about how devices are used, where they are located, and any unusual activity. Additionally, new platforms are being developed to help insurance companies share anonymous data on fraud, making it easier to identify repeat offenders and stay ahead of evolving fraud tactics.

As AI continues to develop, striking a balance between innovation and ethical considerations will be key. While AI has the potential to revolutionize fraud detection and many other industries, it is essential to address biases, improve data quality, and ensure transparency. With proper oversight and responsible implementation, AI can be a powerful tool that benefits both businesses and consumers.

By Gaurav Mittal

Author Bio: Gaurav Mittal is a cybersecurity, data science, and IT expert with two decades of experience leading high-performing teams in cloud computing, machine learning, and data security. A thought leader in AI and automation, Mittal has published articles on optimizing ML deployment, securing email communications, and automating workflows. Mittal is an AWS-certified cloud

practitioner and a Lean Six Sigma-certified professional.

Credit Line: This article was produced by the Independent Media Institute.