IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Wipro

Ethics and bias are starting to have an impact on organizations that are deploying AI-enabled applications and processes. The need to provide ethical and transparent AI algorithms and implementation practices will become a major factor for organizations in the years ahead.

The Impact of Ethics and Bias on Artificial Intelligence

February 2019

Written by: David Schubmehl, Research Director, Cognitive/AI Systems, and Adelaide O'Brien, Research Director, IDC Government Insights

Introduction

Artificial intelligence (AI) promises to change the way we run our businesses, empowering us to make better decisions more quickly. If you look at the potential for the problems it can solve in humankind in the world, it's very large. We already see that, with the ability to help fight human trafficking, the ability to reunite missing kids with their parents, and for education services and security services, there is a huge amount of good happening in the world right now based on using these types of machine learning services. This is a future of business that gets lost in discussions of robots replacing humans. The true promise of artificial intelligence is to improve how we as humans run our businesses today and to allow us to be more productive in our work than we otherwise would be on our own. That's an exciting future.

AT A GLANCE

WHAT'S IMPORTANT

Recent advances in AI have made it smarter and faster, and yet AI can provide answers and recommendations that seem biased. The rise and promise of AI brings with it the need to enable environments that foster societal and organizational trust.

KEY TAKEAWAYS

Organizations should strive for their algorithms and AI models to be transparent, secure, and consistent in behavior. Explainability of AI is a key attribute.

However, the rise of AI brings with it questions of who is responsible for making sure these powerful technologies are used for good and not evil. How do we enable environments that foster societal trust and build upon a common vision of human values? With any technology, there is the potential for some to use it irresponsibly or unethically. In one of the worst scenarios, malevolent users manipulated Microsoft's AI chatbot Tay into tweeting racial slurs and genocidal comments. Microsoft quickly solved the problem and has learned from that experience. How do we increase "human intelligence" in a world where artificial intelligence is burgeoning? Recent advances in AI have made it smarter, faster, and more human-like. But with all the data, opinions, and interactions from a global community, AI can inherit our flaws. The CEO of Amazon Web Services, Andy Jassy, believes that, to combat this, it's important to set the right types of standards so that people use the technology responsibly.

Definitions

For those new to AI, the terminology used can be the first challenge to overcome. What is the difference between AI, cognitive computing, and machine learning? Or are these concepts interchangeable?

In fact, they are related in a nuanced manner and support one another when used in concert. The discussion warrants more space than available here, but to aid in the discussion when we use these terms, we mean the following:

- AI is the study and research of providing software and hardware that attempts to emulate a human being.
- Cognitive computing is computing focused on reasoning and understanding that is inspired by human cognition. It is a subset of AI.
- Machine learning is the process of creating a statistical model from various types of data that perform various functions without having to be programmed by a human. Machine learning models are "trained" by various types of data (often, lots of data).
- General-purpose cognitive/AI software platforms are used to build intelligent applications that provide predictions, answers, or recommendations and are a platform for the development of cognitive applications. These applications automatically learn, adapt, and improve over time using information access processes combined with deep/machine learning.
- Conversational AI software platforms are a subset of cognitive/AI platforms that are specialized for the development of intelligent digital assistants and conversational chatbots. Conversational AI platforms use content analytics, information discovery, and other technologies to communicate with human beings.
- Natural language processing (NLP) is the ability to extract people, places, and things (also known as entities) as well as actions and relationships (also known as intents) from sentences and passages of unstructured text.
- Natural language generation (NLG) is the ability to construct textual/conversational narratives from structured or semi-structured data.

Key Trends

The demand for artificial intelligence (AI) software platforms that can provide advice, recommendations, and predictions will continue to be strong. IDC estimates that, by 2022, over US$9.5 billion will be spent on AI software platforms worldwide. Organizations are deploying AI-enabled applications and services, and bias can derail AI development and can cause potentially significant compliance and regulatory issues for organizations.

Some overall trends and predictions that IDC is seeing include the following:

- By 2021, algorithm opacity, decision bias, malicious use of AI, and data regulations will result in the doubling of spending on relevant governance and compliance staff and explainability teams.
    - IT will be expected to proactively monitor and maintain a roster of deployed software utilizing AI algorithms and work with specialist governance and compliance staff.
    - IT will be expected to support governance and compliance efforts, plus handle cybersecurity.

» By 2020, 35 U.S. states and 5 non-European countries will have passed GDPR-like laws, making privacy a global requirement and driving growth in outsourced privacy risk and third-party data services.

- Companies need to be able to demonstrate consumer consent to access and use data, whether done directly or through a third party.
- Customer data stewardship and consent management will become increasingly important preferences for consumers, especially for users of online services.

» By 2024, 50% of structured repeatable tasks will be automated and 20% of workers in knowledge-intensive tasks will have AI-infused software or other digitally connected technology as a "coworker."

- IT support expands beyond technology acquisition, deployment, configuration, and support and must consider security, privacy, and compliance implications.
- Data quality, data governance, and data utilization is even more important, as the ability of AI-enabled automation software to deliver quality outcomes is predicated on quality data inputs and historical data sets.

The reason that ethics is so important is that now we have machine intelligence that sits between us and the organizations that we are dealing with. AI algorithms aren't neutral. They are built by humans, and it leaves them exposed to bias as they are programmed or used. Instances of bias are found in image searches, hiring software, financial searches, and so forth.

Over the past six years, the New York City police department has compiled a massive database containing the names and personal details of at least 17,500 individuals it believes to be involved in criminal gangs. The effort has already been criticized by civil rights activists who say it is inaccurate and racially discriminatory. "Now imagine marrying facial recognition technology to the development of a database that theoretically presumes you're in a gang," Sherrilyn Ifill, president and director-counsel of the NAACP Legal Defense Fund, said at the AI Now Symposium in New York in October 2018.

Not only is facial recognition imperfect, studies have shown that the leading software is less accurate for dark-skinned individuals and women. By Ifill's estimation, the police database is 95–99% African American, Latino, and Asian American. "We are talking about creating a class of people who are branded with a kind of criminal tag," Ifill said.

Meanwhile, police departments across the United States, the United Kingdom, and China have begun adopting facial recognition as a tool for finding known criminals. In June, the South Wales police released a statement justifying their use of the technology because of the "public benefit" that it provides. Indeed, technology often highlights peoples' differing ethical standards — whether it is censoring hate speech or using risk assessment tools to improve public safety.

Another example is of Amazon, where its machine learning specialists uncovered a big problem with its AI recruiting tool, realizing that its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way. The company had to scrap the tool.

Lawyers, activists, and researchers emphasize the need for ethics and accountability in the design and implementation of AI systems. But this often ignores a couple of tricky questions: Who gets to define those ethics, and who should enforce them? The Data & Society Research Institute published a proposal for using international human rights to govern AI. The

report includes recommendations for tech companies to engage with civil rights groups and researchers and to conduct human rights impact assessments on the life cycles of their AI systems.

Algorithms produced by different companies must be constantly benchmarked and refined so that they are as accurate as possible. There should be clarity on how they are recommended to them using those services. For instance, if facial recognition is used for matching celebrity photos, then it may be acceptable to have a confidence level or threshold that is around 80%. But if facial recognition is used for law enforcement or something that can impact people's civil liberties, then the threshold target should be 99%, and even then, it shouldn't be the sole determinant in making a decision. There should be humans involved, and there should be multiple inputs.

Considering Wipro HOLMES™ and the ETHICA Program

Wipro is a leading global information technology, consulting, and business process services company. It harnesses the power of cognitive computing, hyper-automation, robotics, cloud, analytics, and emerging technologies to help its clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability, and good corporate citizenship, Wipro has over 160,000 dedicated employees serving clients across six continents.

Wipro HOLMES™, Wipro's Artificial Intelligence platform, helps enterprises automate processes, redefine operations, and reimagine their customer journeys. Through algorithmic intelligence and cognitive computing capabilities, Wipro HOLMES™ accelerates the digital journey of enterprises and enhances operational efficiency, effectiveness, and user experience across applications, infrastructure management, and key business processes. Wipro HOLMES™ has been successfully deployed in data and information-driven verticals, including banking and financial services institutions, retail, manufacturing, and telecommunications.

Wipro has launched a framework called ETHICA, which stands for Explainability, Transparency, Human-first, Interpretability, Common sense, and Auditability. This framework and program are all about how organizations can ensure ethical and unbiased AI solutions. Technology algorithms are really not biased, but when training and data is introduced, this is where potential biases can come in. However, taking certain steps can help to eliminate the potential bias upfront. Some approaches are:

- **Masking some types of data can help to eliminate potential bias.** For example, if a consumer is applying for a bank loan, name, credit score, social security number, gender, and other attributes are critical to identifying them as a person and making sure that the right person is applying. But once that authentication is done, you may not need all these parameters for the actual loan processing itself. So, masking or eliminating that type of data in the learning model could help to potentially alleviate downstream bias.
- **Deploying ethics transparency and explainability as part of the development process.** An example of this is Wipro's Know Your Customer (KYC), which is being run for banks for example. How do you actually go ahead and onboard a customer without looking at the parameters that were discussed previously such as background, gender, and race. Instead, the algorithms use other factors that are easily explainable, such as purchase and payment history, instead of the factors that can lead to potential bias.
- **Using proper anomaly detection.** Anomalies are based on patterns, where developers look at not just a rule-based engine but any anomaly that could come up in terms of duplication or fraud, irrespective of the background and

irrespective of the type of activity that took place. For example, anomaly detection has been used in travel expenses, payment fraud, and insurance fraud. These anomalies are based on historical data without biases, rather than the detection being biased based on who committed the fraud or what actually caused the anomaly from a biased perspective.

- **Unbiased revenue forecasting.** This is something Wipro is focusing on, where it can predict the revenue of a company and look at multiple parameters without biases. For example, looking at a company without considering origin or ownership structure, using such data as credit history and social media profiles that discuss the company in an unbiased manner.
- **Human-based auditing.** This is where an organization wants to make sure that, every time a critical action is taken, there is a human in the loop. There should always be a human monitor to make sure that, should any bias originate, that monitor can detect and correct the action and then feed that information back into the learning models.

Wipro ETHICA is based on the belief that humans will always remain responsible, and the organization is a key partner in that responsibility. This is all about Wipro HOLMES™ embodying Wipro's core values where customers are considered first, trust must be inherently built into the applications, and the overall organization needs to engrain the values of integrity, explainability, and anti-bias into all the AI-infused solutions it builds. It also includes the controls and compliance capabilities pre- and post-deployment to ensure that no nasty surprises await Wipro's customers or their customers.

Challenges

Some of the most important challenges for organizations and their partners like Wipro revolve around two key factors: technology and people. Within the areas of technology, the focus on models using deep learning lack algorithmic transparency and make it challenging for developers to identify exactly why a particular decision or recommendation was reached. Organization and industry partners need to adopt techniques and algorithms that foster transparency and explainability.

Another key technical issue revolves around data. The use of data that is not broadly based or is indicative of various types of bias can wreak havoc on the creation and use of AI models. The use of small homogeneous data sets can be especially problematic. Organizations and industry partners need to find and use the broadest and most varied data sets possible to eliminate the opportunity for data bias to creep in.

In terms of people, some of the key challenges revolve around the lack of staff experienced in the data science used to create AI models and the lack of education in AI models for auditing staff. Data scientists are a challenge to hire today, and the lack of these data scientists will extend and prolong development times as in-house staff come up to speed with data science. In the same way, internal auditing teams are typically not experienced in auditing algorithms or AI models. This is also an area where time will be spent educating and training staff who have been predominantly involved with auditing financial statements or business processes, not AI models.

Conclusion

Keeping pace with the speed of technological innovation in AI is challenging for many organizations, as industry leaders are constantly developing new tools and techniques. It's important to stay abreast of relevant innovations and technological trends, and organizations should leverage investments made by industry leaders in this field. Organizations should also expect their industry partners to have deep government, analytics, and technology expertise. IDC recommends seeking partners that have responsible and ethical policies governing the use of AI and have developed tools and techniques that test for and detect unintended consequences such as gender, racial, and ethnic bias in AI software. Industry partners can assist organizations in preparing data so it's ready to be used for AI. They can also train employees to be ready for and work together with AI and help organizations build data labs for continuous analytics.

Organizations should seek industry partners who understand the ethical, legal, and social implications of AI, as well as those developing methods for designing AI systems that are responsible and ethical.

IDC also recommends that organizations seek industry experts that have deep expertise in design thinking, as AI algorithms compute but don't "think" and are only useful when designed properly by humans to fit the mission purpose. For example, an algorithm designed for the U.S. Post Office to recognize handwritten zip codes is useless in providing facial recognition of bad actors for U.S. Customs and Border Protection. Industry partners can assist in meshing the needs across agency functions to articulate requirements of the AI system and ensure analytic outputs are tailored to the unique needs of each agency.

Many companies serving the government and private organizations can assist with avoiding potential pitfalls and providing best practices for deployment in related regulated industries such as financial services and healthcare where protection of PII and responsible and ethical AI are top priorities. Organizations should seek industry partners that understand the ethical, legal, cultural, and social implications of AI, as well as those developing methods for designing AI systems that are responsible and ethical.

To establish a better set of behaviors, here are some principles that could be adhered to by technology providers like Wipro:

- **Utility:** Ensure that your algorithms are clear, useful, and satisfying (delightful) for the user.
- **Empathy and respect:** Validate that your algorithms understand and respect people's explicit and implicit needs.
- **Trust:** Strive for your algorithms to be transparent, secure, and consistent in behavior.
- **Fairness and safety:** Ensure that the algorithms are free of bias that could cause harm — in the digital or physical world, or both — to people and/or the organization.
- **Accountability:** Establish clear escalation and governance processes, and offer recourse if customers are unsatisfied.

IDC advises technology providers to strive for strong positive social outcomes and not unintended negative outcomes. Here's how they could help the data teams incorporate data ethically:

- **Establish holistic metrics.** Don't just goal yourself on revenue but think about the social outcomes. Try hard to measure the leading indicators of social outcomes.
- **Have diversity in your data teams.** Work to have a representation of the population where you would be deploying the algorithms. Avoid the blinders of the homogenous teams. Diverse teams will work to have diverse training data, more thoughtful feature sets, and less bias in the data.
- **Have centralized data teams to avoid line-of-business bias.** For example, data teams reporting to sales will lean toward the bias of the sales objectives.
- **Remain dedicated to refining design practices.** Create AI that is human focused and audited for biases.
- **Data teams should be chartered to be the "conscience" officers.**

Ethics must be included as a key component of AI application and services development. AI software platforms like Wipro HOLMES™ should include capabilities for verifying trust and compliance as elemental parts of the AI application development process. Tools to detect bias and to provide explainability are required, so that organizations can comfortably develop and deploy AI applications and services with a known level of risk for their employees, customers, and the public at large. ETHICA is a focused effort by Wipro to imbue Wipro HOLMES™ with the capabilities and tools it needs to deploy verifiable, trusted, and explainable AI-infused solutions.

About these analysts:

David Schubmehl, Research Director, Cognitive/AI Systems

Dave covers information access and artificial intelligence technologies including content analytics, search systems, unstructured information representation, cognitive computing, deep learning, machine learning, unified access to structured and unstructured information, Big Data, visualization, and rich media search in SaaS, cloud, and installed software environments.

Adelaide O'Brien, Research Director, Government Digital Transformation Strategies

Adelaide O'Brien is research director for IDC Government Insights responsible for Government Digital Transformation Strategies. Ms. O'Brien assists clients in understanding the full scope of efforts needed for digital transformation, and focuses on technology innovations such as Big Data, AI, cognitive, and cloud in the context of government use cases such as customer experience, data-driven benefits and services, and public health protection.

IDC Custom Solutions

IDC Corporate USA

5 Speen Street
Framingham, MA 01701, USA

T 508.872.8200

F 508.935.4015

Twitter @IDC

idc-insights-community.com

www.idc.com