

ALGORITHMIC DETERMINISM

Addressing the Impact of Artificial Intelligence on Society

VALUES & PRINCIPLES

ASCERTAINING CORE VALUES

At Accenture, our core values and principles with regard to AI are:

- **Fairness** (identifying how bias can creep in and mitigating)
- **Accountability** (governance, systems of redress/address)
- **Transparency** (interpretable and understandable)
- **Trustworthiness** (safe, reliable, secure)
- **Agency** (human-centric, sustainable, beneficial)

While we embody these values in our internal and external AI, we recognize that our clients and partners may have a different set of core values and principles. As part of our Responsible AI journey with clients, we assist with an examination of corporate core values and mission statement, working with key stakeholders in the organization to identify a list of priorities unique to that organization. By doing this, we are able to link the use of AI to a corporation's already-identified desired impact on society.

ETHICAL IMPLICATIONS OF AI

One of the most pressing aspects of digital cooperation is the explosion of activity around Artificial Intelligence (AI). Computers are increasingly mimicking traditionally-human behaviors, such as the ability to sense the world around them, construct internal models to understand the world, act with individual discretion, and learn from mistakes.

Unfortunately, AI systems only know a limited amount about the world. They can only understand the data used to train them. When that input data is skewed, non-representative, or biased, the system inevitably acts in the way it learned: it replicates that bias and hinders individual freedom. As a global IT Advisory firm, Accenture has noticed these issues in the public and private sectors, across industries and countries.

In the [public sector](#), issues have ranged from the life-altering (The Flint Water Crisis, military technology) to the mundane (school bus routing in Boston). In the [technology industry](#), no company is free from criticism, be it for racial bias in search results (Google), enabling discrimination in housing (Facebook), or building gender-biased hiring tools (Amazon). In [Financial Services](#), credit models have long drawn criticism for potential bias – and are again today. And in [Health Care](#), some insurance companies condition coverage on AI supervision.

ALGORITHMIC DETERMINISM

The U.N. should recognize these issues and work to address one critical aspect: Algorithmic Determinism. Algorithmic Determinism describes how AI can magnify biases in the world. An initial measurement bias combines with an AI feedback loop to cause behavior that it purports to measure, influencing people impacted by an AI system in unforeseen ways. This is particularly insidious when it is influenced by immutable

characteristics (such as age): systems can learn to penalize individuals based on characteristics that cannot be changed.

Some places are taking this challenge head on. For example, California recently banned car insurance companies from using [gender in determining insurance rates](#). But more must be done. Algorithmic Determinism addresses the foundations of AI's impact on society. It has the potential to hinder the ability of individuals to interact and cooperate with each other, especially in situations where that ability may be initially limited by socioeconomic circumstance.

THE IMPORTANCE OF DIGITAL COOPERATION

Fairness emerges as a key value in AI. Because this technology is used in ever increasing facets of life, from retail such as ordering fast food at [KFC where facial recognition AI is used to make menu suggestions](#), to criminal justice for determining a [defendant's risk of recidivism and prison sentence](#), to [financial services](#) and even in [schools](#), even minor fairness issues have the potential to have profound and lasting impacts. Algorithmic Determinism is a prime example. Two key values emerge: Fairness and Agency.

Defining fairness is the first challenge in addressing the problem of Algorithmic Determinism. No one definition of fairness exists (nor can it). Currently, developers creating AI decision-making programs and the businesses, governments, and other organizations using them do not have any multilateral standards or requirements around fairness. Questions of agency and justice must be considered when crafting a definition of fairness for both end users and other stakeholders. The U.N. needs to guide these organizations – particularly governments – to define agreed-upon standards for what constitutes fair and unfair applications or impacts of AI.

Agency is the other principal value in AI: as artificial intelligence tools increasingly make recommendations on a wide range of topics, they have slipped into a decision-making rather than advisory role. On a small scale [filter bubbles](#) on social media may simply influence and individual and lead to [polarization](#), whereas on a large scale AI tools have been used to determine [food stamp retail authorizations](#) affecting entire communities. As this technology becomes more complex and widespread, its impact on the [agency of individuals](#) will only increase, and it is necessary to put safeguards in place governing its use.

These two values represent two of the core challenges of digital cooperation. For effective collaboration to occur, access needs to be distributed equitably – a condition that necessitates fairness. For that collaboration to be fruitful, individuals need real (not just perceived) agency to act individually, free from coercion. Fairness and agency are the preconditions for effective cooperation.

CASE STUDY: SOCIAL CREDIT SYSTEM - CHINA

The [social credit program](#) that is currently being piloted in China is a perfect example of the risks to fairness and agency in the broad use of AI. China claims the system is meant to increase societal cohesion: by creating transparency into individual

behavior, the score can make [individuals feel safer](#) in trusting their compatriots. However, the system has faced criticism outlining dangers that are likely to undercut that purpose, hindering fairness and individual agency throughout the country. In effect, it is a large-scale program predicated on Algorithmic Determinism.

The social credit program collects vast sets of data on all the country's citizens, across multiple facets of their lives. The gathered data is divided into four distinct segments: a government trust system, a commercial credit system, a social trust system and a judicial trust system. Based on their score in the aggregate system, an individual – and even local governments – can be barred from certain activities. Individuals can be barred from critical infrastructure and activities, ranging from [transportation](#) to [universities](#).

As this system proliferates, it guides individuals to act in the same ways that they have in the past. Without continued access to the same resources as those with higher scores, people with lower scores lack effective recourse to improve. This absence only leads to a lower score. It separates people into "types" that form the basis for future societal interaction, threatening fairness, and nudges them into repeating past behavior, threatening individual agency. Insofar as this system is meant to form a basis for social cohesion and trust, it embeds Algorithmic Determinism into the fabric of society.

METHODS & MECHANISMS

THE NEEDS

Effective guidance and low-level control are key to the benefit of AI systems. Existing mechanisms to achieve this control – both low-level technical mechanisms, as well as high-level governance mechanisms – are insufficient. Not only are they limited within their respective sectors, they are woefully inadequate to address interaction effects.

To foster collaboration and norms that can provide the basis for this guidance, it is necessary to address the problem at two levels. First, technical solutions drive low-level transparency and facilitate the identification and remediation of immediate issues. Second, governance provides the high-level tracks through which larger concerns, among them Algorithmic Determinism, can be addressed.

TECHNICAL SOLUTIONS

Technical solutions are the first way firms can address Algorithmic Determinism. Generally, technical solutions focus on *post hoc* evaluation. Accenture launched the Fairness Tool in June 2018, providing an enterprise solution for addressing issues of algorithmic bias. Soon after, both IBM and Google released tools to identify issues with algorithmic fairness, using a variety of methodologies developed by academics.

While helpful, these tools are often unable to address complex fairness issues. For example, the impossibility theorem states that of the three primary ways of defining fairness (demographic parity, equalized odds, and predictive rate parity), any algorithmic fix can solve for two but not all three. Also, these metrics address group fairness, [but not individual fairness](#).

Technical solutions are helpful in identifying wrongs and providing solutions. However, as concepts like ‘fairness’ have no singular definition, additional frameworks are necessary to ensure agreed-upon and standardized methods of addressing issues of bias. These higher-level questions need to be addressed by humans, not code.

THE ROLE OF GOVERNANCE

Multi-lateral frameworks facilitate governance. Current frameworks are often good at governing the ethical creation and use AI programs. But they are insufficient to address second-order and collaborative challenges of AI, such as Algorithmic Determinism. Furthermore, frameworks are developed in disparate sectors and are not standardized, such that it is not effective to apply them across a broad range of programs and sectors. Stakeholders from all areas need to be involved.

Several frameworks have been proposed to address practical ethical concerns with Artificial Intelligence, of which the following are use widely: the Future for Privacy Forum’s [Four Types of Harm](#), and the [Three Questions](#) proposed by The Authority of “Fair” in Machine Learning. These frameworks represent two distinct approaches to ethics: harm mitigation and contextual fairness. As they come from two different communities – advocacy and academia – they also characterize two different starting

points: possible harms and the development process. However, neither framework sufficiently captures a full range of harms, including the harms that result from Algorithmic Determinism.

To remedy the insufficiencies of the above framework, we propose two new questions to ask. Like the above frameworks, they are not sufficient in and of themselves to mitigate all ethical concerns. However, they supplement the above frameworks to address Algorithmic Determinism. The questions are:

- To what degree does the system encode or predict false assumptions of personal mutability?
- To what degree can the system prompt or cause behavior which it purports to independently measure?

These questions aim to identify the circumstances in which Algorithmic Determinism tends to arise. Questioning personal mutability allows designers to understand what types of behaviors they are incentivizing or requiring, and the impacts thereof. Addressing the impact of systems on measurement can help identify feedback loops, and the measurement biases that feed them.

FILLING THE GAP

Collaboration between some of the stakeholders in tech and academia and government already exists, but it is typically focused on the development of specific solutions. Google, Amazon, and Microsoft are all working on AI technologies for the government. These relationships are geared towards the development of very specific technologies, and hence do not offer models for ongoing, broad collaboration.

Technology advisory firms, trade associations, and technology product companies are each in a position to address this issue as leaders in technology. As businesses are a significant conduit through which AI impacts the lives of individuals, information-sharing and the development of comprehensive best practices can support integrated, rather than fragmented, problem-solving. Moreover, major companies' presence in several global markets means that these companies can set and enforce norms internationally.

However, a technology driven solution from the tech or business sector alone will be insufficient to address the dynamic ethical concerns that emerge as AI continues to evolve. The UN must play a crucial role in this space by leading the discussion and setting governance norms around this subject. As an impartial party, the UN can bring together stakeholders from tech, business, government, etc. on equal footing for discussions around norms, frameworks, and solutions.

Although some collaboration already exists, The U.N. needs to bring it to the fore. It is especially important to include input and representation from marginalized groups, to ensure that new frameworks are inclusive and fair for all users.

ILLUSTRATIVE ACTION AREAS

TRADITIONAL ACTION

Exploring the issue of digital cooperation through the lens of Algorithmic Determinism shows that there is potential for improvement in two action areas traditionally addressed by the UN: inclusive participation in the digital economy and the protection of human rights. AI has the potential to threaten the equal treatment of groups that are already vulnerable; the U.N. can support mitigation of that challenge.

The U.N. has long engaged with emerging technologies through the work of the Economic and Social Council (ECOSOC), the General Assembly, and their [subsidiary bodies](#). The Commission on Science and Technology for Development provides the General Assembly and ECOSOC with high-level advice on relevant science and technology issues, and the UN Technology Innovation Labs foster partnerships and provide a global perspective on technology, connecting local innovation ecosystems across the world.

Both organizations engage a range of stakeholders to further business and policy. Most recently, the Commission has worked on public policy issues pertaining to the internet – fairness in AI and Algorithmic Determinism is the next challenge it should address. The longer this issue remains unaddressed, the more extensive and grievous the harm caused by Algorithmic Determinism will become.

CASE STUDY: PREDPOL

As artificial intelligence programs are increasingly used across a range of real-life applications, the impacts of Algorithmic Determinism have the potential to spill over into the non-digital world and have nefarious consequences for already marginalized individuals and communities. A common example of this problem is found when examining the use of predictive policing AIs such as PredPol.

PredPol is a software algorithm used by police forces across the United States and Europe to dispatch police resources. However, there are two serious problems with the use of this program. First, misalignment between stakeholders has led to the tool being used in a manner inconsistent with its design – the nuance between predicting where crime is recorded vs. where it occurs is crucial in this [case](#). Second, the algorithm itself is unduly deterministic and considers biased factors.

PredPol works as an observational program – it ‘observes’ the rates and locations at which crimes are recorded (whether reported or witnessed by police) and then predicts where future crime will be recorded. Those predictions are used to dispatch police. In practice, because the system implicitly assumes all relevant crimes are recorded, it tends to send more police to places where more crimes are recorded (not necessarily where more crimes occur). Because more police are sent, they find more crimes. This leads the system to believe it should send *even more police* who, predictably, find more crime. This occurs independent of how much crime there actually is in a neighborhood.

The program and its implementation compound biases and lead to the disproportionate policing and incarceration of already marginalized communities – in violation of their basic rights to freedom and safety. The data used for PredPol is gathered from several decades of [biased police practices](#) and will therefore maintain the status quo at best. In Newark, NJ where Black people comprised 54% of the population, but 85% of pedestrian stops and 79% of arrests by police and were 2.5 times more likely to be stopped by the police than their white counterparts, the application of PredPol will not shift to more fair outcomes. The situation is similar across many American cities.

NEW OPPORTUNITIES FOR SOLUTIONS

The Commission on Science and Technology for Development is in the position of setting the agenda and determining key issues. By making AI and Algorithmic Determinism a focus it will signal this issue as a priority for member states.

The UN Technology Innovation Labs are then well-positioned to address the more specific issues around AI in their host country and geography. The development of better educational resources and understanding of the capabilities, risks, and limitations – such as Algorithmic Determinism – of artificial intelligence and machine learning is necessary and is a crucial first step towards ensuring informed and fair use across stakeholders.

Academic and non-governmental organizations are well-positioned to collaborate and work on this issue, as they do not face many of the constraints and interests that businesses and governments in this space typically have. In addition to developing better educational resources around the technologies and issues, collaboration between governments and human rights organizations is necessary to develop norms and standards for decision-making AI programs, which can then serve as the basis for regulations. This is important not only for the protection of normal consumers and their access and participation in the digital economy, it is also crucial for sensitive applications such as law enforcement and financial services that have the potential to severely restrict human rights and equality.

As a whole, the UN is well positioned to address issues around Algorithmic Determinism internationally. By increasing the focus on better education around these issues and shared learning and knowledge, these UN organizations can effectively reduce the current impact of these issues and foster the prosperity that effective AI can bring. Addressing Algorithmic Determinism is a first step in ensuring AI benefits people, businesses, governments and the global community.