



THE ALGORITHMIC JUSTICE LEAGUE'S 101 OVERVIEW

OVERVIEW

Through a combination of art, research, policy guidance and media advocacy the Algorithmic Justice League is leading a cultural movement towards equitable and accountable AI. This requires both looking at how AI systems are developed and preventing harmful use. Our goal is to empower communities and galvanize decision makers to take action that mitigates AI bias and harms.

EQUITABLE AI

Equitable AI is focused on empowering people.

Agency and Control

Equitable AI requires that people have agency and control over how they interact with an AI system. Agency requires awareness of how these systems are used – for example, at airports, stadiums, schools, hospitals and in hiring and housing – the entities involved in creating and deploying the system – from business, government and academia – and the risks and potential harms.

Affirmative Consent

Equitable AI requires securing **affirmative consent** from people on how or whether they interact with an AI system. To provide affirmative consent, people must understand exactly how their data will be used along with a guarantee that it will only be used for that express purpose. The defaults for affirmative consent are “opt-in.” If people elect not to opt-in, affirmative consent requires that they will not be denied access to a platform, service or opportunity as a result. Unlike the terms of service that tech companies require people to click through to use their platforms, affirmative consent for AI cannot be offered on a take it or leave it basis.

Centering Justice by Focusing on Impermissible Use

In addition to providing agency, equitable AI respects human life, dignity and rights.

Considering these values requires prohibiting certain government and commercial uses of AI. This includes AI that would identify targets where the use of lethal force is an option (for example, when being stopped by the police) and AI used in government and citizen-led surveillance. Justice requires that we prohibit AI from being used by those with power to increase their default level of control, particularly where it would automate long-standing patterns of injustice such as racial profiling in

law enforcement, gender bias in hiring and overpolicing of immigrant communities. Justice demands protecting those that would be targeted by these systems from their misuse.

ACCOUNTABLE AI

Accountable AI is focused on holding decision makers responsible.

Meaningful Transparency

For an AI system to demonstrate **meaningful transparency** it must provide an explanation of how the system works, how it was designed and its purpose. Critically, meaningful transparency allows people to clearly understand the intended capabilities and known limitations of the AI.

To demonstrate this standard, companies and governments must share information about how AI is being used in their own decision-making and sold to others. The goal is for people to understand the societal risks every time they encounter an AI and how their data is being used to make decisions that affect them. Sharing this information is supported by reporting requirements that are mandated through law or agreed to through codes of conduct like the SAFE Face Pledge.

Continuous Oversight

AI systems are constantly evolving. As a result, we must implement methods of **continuous oversight** to hold companies and government agencies accountable for harms. Continuous oversight is enabled by law and policies that support independent review and mandate changes if harmful or discriminatory impacts are revealed. Requirements might include: maintaining on-going documentation, submitting to audit requirements and creating channels of access for researchers and civil society organizations to engage without friction.

Redress Harms

Accountable AI provides people who have been harmed with access to remedy, meaning that there is a working pathway for people to **contest and correct** a harmful decision made by an AI.

For example, if an AI incorrectly denied a welfare benefits check, remedy would entail an easy way for the recipient to call attention to this error and receive payment plus interest for time lost. If an AI was suspected of disqualifying a job applicant based on gender or race, remedy would allow the applicant to discover how the decision was made and provide a basis for challenging the decision in court.

WHY “EQUITABLE” AND “ACCOUNTABLE”?

There are many different terms that have been used to describe a policy approach to AI. We want to be clear about what we mean by equitable and accountable as separate from these other approaches.

Limitations of Ethical AI

The notion of “Ethical AI” has been leveraged by big tech companies – investors and executives strategically aligned with academia and government – to push for voluntary principles over government regulation. The idea of using ethics is not problematic in itself, but has led to a proliferation of “AI Principles” with limited means for translating these principles into practice. A system of AI ethics allows companies to keep along as they see fit, accountable only to rules that they have set. The ball is in their court from beginning to end. Appeals to ethical AI can also be leveraged by the government to justify questionable policies that have not been settled in law. This is a limited approach from our perspective because it does not create any mandatory requirements or ban certain uses of AI. Our focus is instead on creating action that bridges the gap from principles to practice.

Limitations of Inclusive AI

While calls for Inclusive AI may be well-intended, inclusion alone does not support progress. At times, respecting life, dignity and rights may require that a system gather more data with affirmative consent, for example to support accuracy in precision medicine across diverse groups. At other times, including more data can mean improving upon a system that unfairly subjects vulnerable populations to additional targeted scrutiny. In service of “inclusion,” data may also be collected in violation of privacy and without consent.

While we do have an interest in creating robust AI systems where appropriate, how the data is collected and the purpose for which it is being used is critical to evaluating inclusion. On the one hand, improvements to AI may limit discriminatory outcomes, – for example, correcting for a self-driving car that fails to detect darker pedestrian faces. At the same time, **we must** always call into question whether the AI use is supported by our values and should be permitted at all. If we focus only on making improvements to datasets and computational processes, we risk creating systems that are technically more accurate but also more capable of being used for mass surveillance and to enhance discriminatory policing practices that allow for the use of lethal force.

FOCUSED ON HARM

Technical standards are not enough to ensure that AI systems will not be deployed to threaten civil liberties and amplify existing patterns of discrimination. The use of facial recognition in surveillance threatens the civil liberties of all citizens, including freedom of expression, freedom of association and due process.

In addition, because these systems are much less accurate for vulnerable populations, those groups – including immigrants and communities of color – face an even greater risk to their rights when confronted with an invasive AI system like facial recognition, which has a greater likelihood of targeting and misidentifying these individuals, leading to a loss of physical liberty, jobs and other serious consequences. These groups already face societal discrimination, for example through racial profiling and over-policing. Introducing AI risks making these problems and the conditions for inequality in our society much worse.

Process over Product

All of our recommendations supporting equitable and accountable AI are intended to focus on **the process** embedded at every level of operations **over any individual product**. An effective process can provide a reliable standard for outside evaluation and be applied no matter what the specific technology in question. It also allows for the voice and choice for the people who will be impacted to be incorporated throughout the lifecycle and ultimate decisions regarding the use of the AI. Companies and governments must evaluate their practices with respect to these priorities.