



AI EXPLAINED

NON-TECHNICAL GUIDE FOR POLICYMAKERS

Branka Panic, Founder and CEO, AI for PEACE
Co-Founder, Center for Exponential Technologies

February 2020

CONTENT



AI BASICS

AI STATE OF THE ART

WHAT IS AI?

WHY NOW?

MACHINE LEARNING

DIFFERENT TYPES OF AI



AI APPLICATIONS

AI APPLICATIONS ACROSS INDUSTRIES

AI AND MEDICINE

AI AND AUTOMATIVE INDUSTRY

AI AND EDUCATION

AI APPLIED FOR GOOD OR BAD?



AI CHALLENGES

ETHICS

BIAS

SAFETY

PRIVACY



THE FUTURE OF AI

THE ROLE OF POLICYMAKERS

HOW TO NAVIGATE THROUGH THE GUIDE



THE GUIDE IS DIVIDED IN FIVE COLOR-CODED CHAPTERS, TO MAKE IT EASIER TO NAVIGATE.

PAY ATTENTION TO “ICONS”, SOME OF THEM CONTAIN LINKS TO MORE DETAILED INFORMATION OR A SPECIFIC RESOURCE ABOUT THE TOPIC.



DEEP DIVE

Click on magnifier to get a deeper insight into that chapter's topic.

AI BASICS

5 **AI STATE OF THE ART**

6 **WHAT IS AI?**

10 **WHY NOW?**

11 **DIFFERENT TYPES OF AI**

16 **MACHINE LEARNING**



”

**Artificial
Intelligence (AI) is
likely to be either
the best or worst
thing to happen to
humanity.**

Stephen Hawking

AI STATE OF THE ART TECHNOLOGY OF THE PRESENT OR FUTURE?

Today, everyone is talking about AI and how it is going to change the way we live our lives. But AI is not only about the future. **AI is already here, today, affecting people across the globe.**

AI is changing the life of individuals. We use AI in our everyday life: google maps, driving directions, spam email filters, song suggestions, friend suggestions, searches online, even Instagram filters, all this is powered by artificial intelligence. AI applications have begun to transform major industries, including healthcare, finance, manufacturing, retail, education, and entertainment.

Hence, AI is not far or foreign to our everyday lives as we sometimes think. And it has immense implications to many spheres of our lives shaped by various policies.

This guide is meant to help policymakers understand basics of AI and how it affects our society. It offers explanations and additional resources to help policymakers prepare for the current and future AI developments.

WHAT IS AI?



artificial intelligence noun

Definition of *artificial intelligence*

- 1 : a branch of computer science dealing with the simulation of intelligent behavior in computers
- 2 : the capability of a machine to imitate intelligent human behavior

There is no single agreed and accepted definition of artificial intelligence.

It is typically defined as the **ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving.**

AI is not a single technology; it is rather a collection of different technologies that can enable machine to act in a human-like levels of intelligence.

The current state of the art uses machine learning, deep learning, and reinforcement learning to identify patterns, produce insights, enhance knowledge-based work and automate routine tasks.

Examples of AI technologies that solve problems today are robotics and autonomous vehicles, computer vision, language and virtual agents.

AI research uses tools and insights from many fields, including computer science, psychology, philosophy, neuroscience, cognitive science, linguistics, operations research, economics, control theory, probability, optimization and logic.



Explore EAAA Digital Library to learn more about defining and explaining AI related terms.



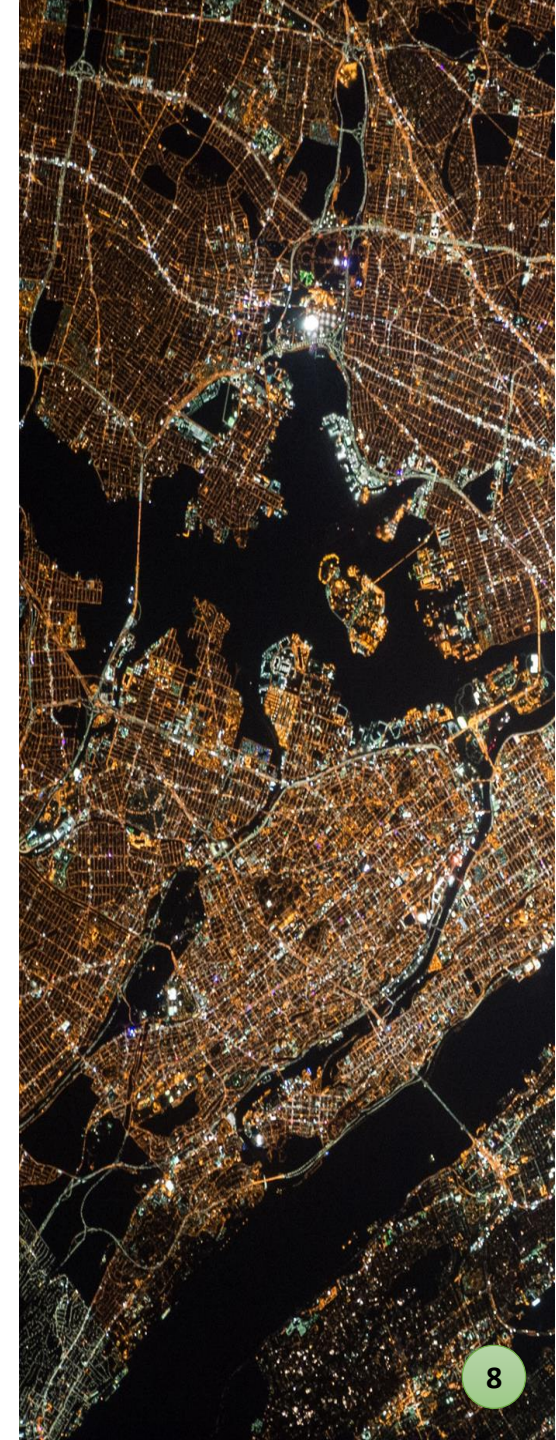
John McCarthy, who coined the term in 1956, defines it as "the science and engineering of making intelligent machines."



Oxford Dictionary defines it as the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.



Economist Intelligence Unit defines AI as a set of computer science techniques that enable systems to perform tasks normally requiring intelligence.



AI IS NOT A NEW CONCEPT

The term “artificial intelligence” was coined in 1956.

Turing proposed the famous Turing test in 1950 and the term “artificial intelligence” was first used at a workshop at Dartmouth College in 1956. Attendants were given millions of dollars to create a machine as intelligent as a human but couldn’t reach the goal. This was the first so called “AI winter”, of AI not being able to fulfill expectations, followed by other “winters” to come. Until now.

WHY NOW?

AI has been around for a long time, but only now we are seeing a surge in the technological advancement of AI. There are a couple of main reasons for this acceleration:

PROCESSING POWER

The processing power and storage ability of today’s machines is so much bigger than anything we saw before.

LOWERED COSTS

At the same time, the cost of cloud access, data storage, and processing has fallen, and the availability to store large inexpensive datasets has risen.

DATA AVAILABILITY

Large availability of “big data”, extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

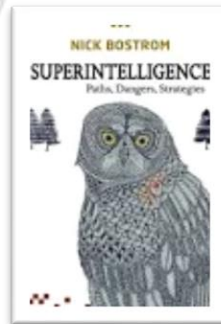
This immensely improved our capacity to build intelligent systems that can learn.

DIFFERENT TYPES OF AI

IMPORTANCE OF CLEAR UNDERSTANDING

Given the fears surrounding AI technologies, it is very important to understand the distinction between different levels of AI so that the potential threats and benefits can be rationally and realistically assessed.

AI has evolved significantly and today it can be used to fill the gaps in human abilities. But AI systems are not as nearly as good as humans at reasoning and thinking creatively and strategically.



ANI ARTIFICIAL NARROW INTELLIGENCE

Narrow AI or Weak AI, is focused on one single narrow task. It possesses a narrow-range of abilities. Narrow AI is something most of us interact with on a daily basis (Google Assistant, Google Translate, Siri, or Alexa).

AGI ARTIFICIAL GENERAL INTELLIGENCE

General AI or Strong AI is much closer to human intelligence. However, AGI is still an emerging field. The human brain is the model for creating General Intelligence, and we still lack comprehensive knowledge of its functionality. Therefore, experts don't agree how likely we will have AGI soon.

ASI ARTIFICIAL SUPER INTELLIGENCE

Artificial Super Intelligence is an AI more powerful and sophisticated than a human's intelligence. To reach this point and to be called an ASI, an AI will need to surpass humans at absolutely everything. The ASI type is achieved when AI is more capable than a human.



IBM Deep Blue is an example of ANI, developed by IBM in 1977. It is known as the first chess-playing computer to win both a chess game and a chess match against a world champion Gary Kasparov.

Fourteen years later IBM applied AI to the challenge of Jeopardy, the quiz show that was covering questions on just about anything. **IBM Watson** incorporated aspects of AI, machine learning, deep question answering and natural language processing to play and, ultimately defeated the game's greatest champions.

AGI IN 2040

Survey of Expert Opinion on the Future Progress in Artificial Intelligence, conducted in 2011, suggested that the median estimate was a 50% chance that artificial general intelligence (AGI) would be developed by 2040–2050.

MACHINE LEARNING

Machine learning is a branch of artificial intelligence, the core of AI, based on the idea that **systems can learn from data, identify patterns and make decisions with minimal human intervention.**

The process is powered by an **algorithm**, a function that is able to improve its performance over time by training itself using methods of data analysis and analytical modelling. Machine-learning algorithms use statistics to find patterns in big amounts of data. Data includes a lot of things - numbers, words, images, clicks. If it can be digitally stored, it can be fed into a machine-learning algorithm.

Machine learning is the process that powers many of the services we use today - **recommendation systems** like those on Netflix, YouTube, and Spotify; **search engines** like Google and Baidu; **social media feeds** like Facebook and Twitter; **voice assistants** like Siri and Alexa.

DEEP LEARNING

Deep learning is a subfield of machine learning, modeled after the human nervous system, known as artificial neural network. It uses a technique that gives machines an enhanced ability to find and amplify even the smallest patterns. This technique is called a **deep neural network** - deep because it has many layers of simple computational nodes that work together to work through data and deliver a final result in the form of the **prediction**.

ALGORITHM is a set of steps for a computer program to accomplish a task.



How can live videos be transferred across internet so quickly?



Audio and visual compression algorithm



How does google maps give us driving directions?



Route finding algorithm



How can NASA rearrange and schedule movement of solar panels on the International Space Station?



Optimization and scheduling algorithm



Learn more about machine learning at free Coursera course [AI for Everyone](#).

THE NUMBER OF ENTERPRISES IMPLEMENTING AI GREW

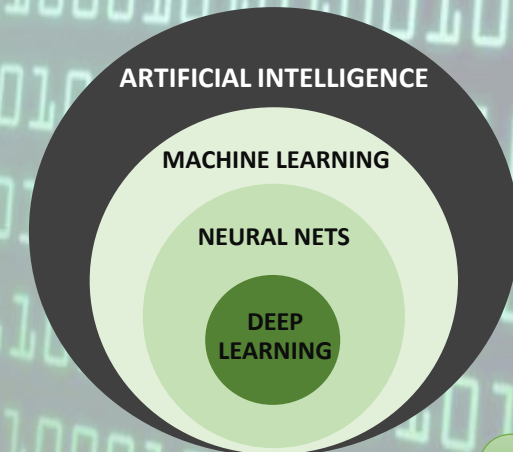
270%

IN THE PAST FOUR YEARS AND

tripled

IN THE PAST YEAR

Gartner, [2019 Survey](#)



HOW CAN A MACHINE LEARN?

TRAINING DATA

Machine learning and deep learning rely on training data to learn relationships, increase the model's efficiency, and improve its ability to achieve the desired output.

Training data refers to a data set that has been collected, prepared, and provided to the model for the purpose of teaching prior to active deployment. The quality, quantity, structure, and contents of training data are key determinants of how machine learning and deep learning models will function in a real environment.

There are different types of machine learning:

SUPERVISED

The data is labeled (organized and described) to tell the machine exactly what patterns it should look for. The algorithm then deduces the salient features that characterize each label and learns to recognize those features in new data.

UNSUPERVISED

The data has no labels. The machine just looks for whatever patterns it can find. Algorithm must structure data, discover patterns, classify inputs, learn functions, and produce outputs without external validation or support.

REINFORCED

A reinforcement algorithm learns by **trial and error** to achieve a clear objective. It tries out lots of different things and is rewarded or penalized depending on whether its behaviors help or hinder it from reaching its objective.

? "Dog or Muffin?"

How do we teach a computer the difference between visually almost identical objects in color and form? At first site it can be challenging to see the difference. Challenge was accepted by online community, and AI was put to work. Images were passed through image recognition APIs and AI was able to see the difference with high accuracy (with some occasions of confusing muffin with stuffed animal 😊).



“Deep Learning is a **superpower**. With it you can make a computer see, synthesize novel art, translate languages, render a medical diagnosis, or build pieces of a car that can drive itself. If that isn't a superpower, I don't know what is.”

Andrew Ng, Founder of deeplearning.ai



AI APPLICATIONS

20

AI APPLICATIONS ACROSS INDUSTRIES

21

AI AND MEDICINE

23

AI AND AUTOMATIVE INDUSTRY

25

AI AND EDUCATION

27

AI APPLIED FOR GOOD OR BAD?



”

AI is not just a game-changer; it will disrupt the entire playing field.

Neil Jacobstein
Singularity University Artificial
Intelligence and Robotics Chair

AI APPLICATIONS ACROSS INDUSTRIES

APPLIED FOR GOOD OR BAD?

Adoption of AI boomed across a wide range of areas in the previous couple of years. The technology behind AI is being deployed in making music and writing books, in medicine, pharmaceuticals and healthcare, in warfare and humanitarian action, in making decisions about our credit and insurance. In short, it's making decisions that affect our lives whether we like it or not.

AI has been used to develop and advance numerous fields and industries, including finance, healthcare, education, transportation, and more.

An algorithm can't choose where, when, or how it will be used, or if it will be used for good or bad intentions. This decision is in our hands as well as the challenge of ensuring ethical use of AI.

As AI develops and its applications grow, **there is a great opportunity, but also a great responsibility to make sure it contributes to the public good** and benefit to all, with fairness, reliability, security, and where appropriate transparency and privacy ensured.

AI AND MEDICINE

AI IMPROVES ACCURACY OF BREAST CANCER DETECTION

AI was applied as a model that read and interpreted the findings of digital breast tomosynthesis images, three-dimensional mammography that takes multiple pictures of the breast to detect possible cancers.

Results showed that the deep learning tool was able to improve the accuracy of detection and cut reading times in half. Sensitivity, a measure of the test's ability to correctly identify those with the disease, increased from 77 percent to 85 percent when AI was employed.

The time required by radiologists to read and interpret the findings was cut from over a minute to 30 seconds with the assistance of the AI. Such a decline in reading time means radiologists will be able to interpret more images in the same amount of time and diagnosis more individuals efficiently.

»» improved efficiency and accuracy

ALONG WITH BREAST CANCER. RESEARCHERS HAVE BEEN LOOKING AT HOW AI CAN IMPROVE THE EFFICACY AND EFFICIENCY OF CARE FOR LUNG, BRAIN AND PROSTATE CANCER, IN ORDER TO MEET DIAGNOSIS DEMANDS.

*Cancer is the second-leading cause of death globally (one in six deaths is due to cancer). Breast cancer in the most common cancer in the world, along with lung cancer, with **2.09 million cases**.*



World Health Organization, 2018

 Google AI Blog

Even Google is looking at how AI can be used to diagnose cancer. They trained an algorithm to detect metastasized tumors, with a

99%

SUCCESS RATE

For deep dive in AI applications in medicine, healthcare and pharmaceuticals click on magnifier.



AI AND AUTOMOTIVE INDUSTRY

THE MENACE AND THE PROMISE OF AUTONOMOUS DRIVING

AI is used to enable the cars to navigate through the traffic and handle complex situations. AI software is combined with other IoT sensors, such as cameras, to ensure proper and safe driving. Autonomous vehicles are equipped with sensors, cameras, and radars harvesting data, trying to perceive the surrounding, lanes, trees, traffic lights.

There are different levels of driving automation: automation for driver assistance, partially, highly, fully automated driving and completely automated cars. While automation for driver assistance and partially automated cars are in commercial use, the remaining types are still under test conditions.

The impact of AI is not limited to autonomous cars. Car manufacturers and factories are seeking ways to utilize AI to enhance production, which brings fears about loss of jobs due to automation of work.

INTERNET OF THINGS (IoT) describes the network of physical objects - "things" that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet. These devices range from ordinary household objects to sophisticated industrial tools.

NEARLY
1.25
million
PEOPLE DIE IN
ROAD CRASHES
EACH YEAR
ON AVERAGE

3,287

DEATHS
PER DAY

Association for Safe
International Road Travel



AUTOMOTIVE
MOTOR MARKET

6.5%

GROWTH RATE
2019-2023

Global Automotive Motor
Market 2019 Industry Report



In this era of digitalization and smart technologies, the use of autonomous vehicles as a mode of transport is not a question of 'if' but 'when'.

Professor Chen Tsuhan
National University of Singapore



**"BUILDING WORLD'S
MOST EXPERIENCED
DRIVER"**

Right now, a minivan with no one behind the steering wheel is driving through a suburb of Phoenix, Arizona. Waymo, the autonomous vehicle division of Alphabet, Google's parent company is first to put fully self-driving cars on US roads without a safety driver.



Read about challenges of regulating self-driving cars.

AI AND EDUCATION

NEXT GENERATION PERSONALIZED LEARNING

Educational technology is the process of integrating technology into education in a positive manner that promotes a more diverse learning environment. As AI becomes more prominent in this age of big data, it is more adopted in education field as well.

Existing school systems are designed under the assumption that all students progress at the same pace. AI has the potential to create highly differentiated, self-paced classrooms, not replacing, but helping teachers to support students by sharing the workload and recommending the best course of action. AI applications can personalize the learning experience by suggesting individual learning objectives, selecting instructional approaches and displaying exercises that are based on the interests and skill level of every student.

There are five core areas where AI technology is being developed and used for education: **hardware** – smart devices, robotics and **software** powered by AI; **vision and emotion recognition** to detect confusion and engagement and for virtual learning; **text-to-voice interfaces** are used in literacy and language training; **natural language processing** is used in assessment, feedback, and plagiarism detection; **algorithms** are being applied to create personalized, adaptive learning paths for students.



2019 Artificial Intelligence & Global Education Report

BIAS AND PRIVACY CONCERNS

Organizations must now consider what type of data is being collected, how this information is being used and what controls are in place to safeguard student privacy.

The Future of AI in Schools

PERSONALIZING PERFORMANCE

AI can collect and analyze data, helping educators create personalized learning plans for each student.

BREAKING BIAS

AI can help leverage solutions capable of grading exams and evaluating papers using established rubrics and benchmarks to both automate completion and eliminate bias.

AGGREGATING ASSISTANCE

AI-driven assistants can release time and help teachers dedicate more times to connecting with students.



Check out how K-12 schools have adopted AI in 2019.



IN 2018 \$1 BILLION WORLDWIDE WAS SPENT ON AI IN EDUCATION, AND THAT NUMBER IS EXPECTED TO GROW TO OVER A \$6 BILLION BY 2025.

AI APPLIED FOR GOOD OR BAD?

AI FOR GOOD builds on previous ideas of using latest technological advancement to solve societal challenges, namely the Tech for Social Good idea of using digital technologies to tackle some of the world's greatest challenges.



Microsoft started “**AI for Good**” program, providing technology, resources, and expertise to empower those working to solve humanitarian issues and create a more sustainable and accessible world.



Google initiated “**Google AI**” AI for social good program, working to use AI to address societal challenges through research, engineering, and initiatives to build the AI ecosystem.



Partnership on AI brings together researchers, academics, businesses and policy makers to invest more attention and effort on harnessing AI to contribute to solutions for some of humanity's most challenging problems, including health and wellbeing, transportation, education, and the sciences.



AI for Good
Global Summit

An **ITU** experience

AI for Good is a United Nations platform, centered around annual Global Summits, that fosters the dialogue on the beneficial use of AI through the development concrete projects.

The AI for Good series connects AI innovators with those seeking solutions to the world's greatest challenges to identify practical applications of AI that can accelerate progress towards the UN Sustainable Development Goals (SDG's).



Read more about AI applications for Social Good in 2018 Discussion Paper.

Rather than asking what sort of impact AI will have on humanity, we should start by asking what sort of impact we *want* it to have. The people working on AI, not AI itself, are ultimately responsible for how much good or harm will be done.

MALICIOUS USE OF AI

How can we forecast, prevent, and mitigate the harmful effects of malicious use of AI?

This is the question posed in 2018 in a report written by 26 authors from 14 institutions, from academia, civil society, and industry. The authors, representatives from the Future of Humanity Institute, the Center for the Study of Existential Risk, OpenAI, and the Center for a New American Security, argue that AI is not only changing the nature and scope of existing threats, but also expanding the range of threats we will face. The purpose of the report is to survey the landscape of security threats from intentionally malicious uses of AI.

The authors suggest there are three primary security domains: **digital security**, which largely concerns cyberattacks; **physical security**, which refers to carrying out attacks with drones and other physical systems; and **political security**, which includes examples such as surveillance, persuasion via targeted propaganda, and deception via manipulated videos.

Read the entire report here



AI CHALLENGES

31

AI CHALLENGES AND CONCERNS

33

AI AND ETHICS

35

AI AND BIAS

37

AI AND SAFETY

39

AI AND PRIVACY



”

Like all technologies before it, artificial intelligence will reflect the values of its creators. So inclusivity matters - from who designs it to who sits on the company boards and which ethical perspectives are included.

Kate Crawford

Research Professor at NYU
Co-founder of AI NOW

AI CHALLENGES AND CONCERNS

WHAT SHOULD **PRIORITIES** OF POLICYMAKERS BE?

As we showed, AI already brings benefits to economic, financial and social development, medical research and better healthcare, it protects security and human rights, and helps solving environmental problems. It has a huge potential to help solving world's hardest problems and improve countless lives.

However, **with successes come new concerns**. In order to responsibly realize its full potential, the challenges associated with it have to be addressed. [The Future of Life Institute](#) is offering an extensive list of 14 topics of particular concern for the safe and beneficial development of AI, in the near and far future.

Some of these concerns include the economic impacts, labor shifts and inequality, accountability, transparency and explainability, surveillance, privacy and civil liberties, fairness, ethics and human rights, AI safety, security and cybersecurity, political manipulation and computational propaganda, autonomous weapons.

We are singling out four general concerns: ethics, bias, safety, and privacy, believing that they are a crosscutting challenge for many AI industry applications. With brief overview we are offering some examples, proposals and links for further readings.

Addressing these challenges should be a priority for policymakers trying to use AI for good while at the same time mitigating potential threats.

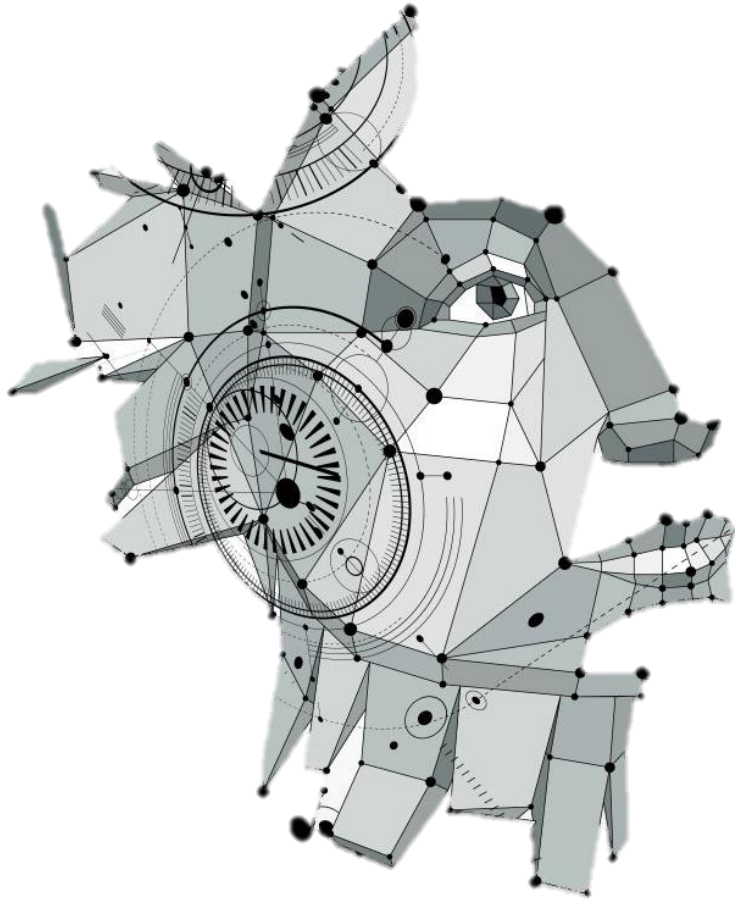
AI CHALLENGES

AI AND ETHICS

AI ethics is a broader field, including more specific topics of discrimination, fairness, bias, and human rights. Today computer sciences and AI lack gender, racial and other forms of diversity, which can lead to false assumptions, skewed design, and discriminatory unfair products.

Algorithms can reproduce and magnify social biases and discrimination from using training data that mirror existing bias in society or that have a skewed representation. If programmers are isolated from the problems they are trying to solve, chances of unintended negative consequences are increasing. Ethical AI has to begin with diverse voices and experiences included at the very beginning of the design process.

The ethics of AI ultimately lies at the intersection of ethical quality of its predictions and their outcomes and the ethical quality of the impact it has on people and societies.



80%

OF AI PROFESSORS
ARE MEN



3.3%

OF GOOGLE'S
FULLTIME WORKERS
ARE BLACK AND

26%

SHARE OF WOMEN
IN COMPUTING



5.8%

LATINX



FATE AI

BEST PRACTICES AND STANDARDS IN
FAIR, ACCOUNTABLE,
TRANSPARENT, AND ETHICAL AI.

PRINCIPLED ARTIFICIAL INTELLIGENCE

Mapping Consensus in Ethical and
Rights-based Approaches to Principles
for AI



Berkman Klein Center for
Internet and Society

There are five biggest sectors
developing on the
intersection of ethics and AI:

HEALTH

from potential misdiagnoses by flawed
algorithms to gene editing

AUTONOMOUS DRIVING

making life-and-death decisions for
human passengers

INSURANCE

algorithms can make discriminatory decisions
against minorities

JOURNALISM

from "fake news" to "deepfake" videos

MILITARY

using AI tech for deadly drone strikes

Many individuals and organizations are
working on developing ethical
principles and guidelines for the
creation and deployment of AI
technologies. **These ethical standards
can have measurable effect only if they
come with enforceable accountability.**

EXAMPLES OF AI ETHICS PRINCIPLES



IEEE Ethically Aligned Design



ITI AI Policy Principles



DeepMind Ethics and
Society Principles

“NEW JIM CODE”

“I present the concept of the “New Jim Code” to explore a range of discriminatory designs that encode inequity: by explicitly amplifying racial hierarchies, by ignoring but thereby replicating social divisions, or by aiming to fix racial bias but ultimately doing quite the opposite.”

Ruha Benjamin, Princeton University, Author of *“Race After Technology”*

AI CHALLENGES

AI AND BIAS

AI can help reduce bias, but it can also reproduce it and scale it. AI systems ‘learn’ based on the data they are given, and data reflects social, historical and political conditions in which it was created, including implicit racial, gender, or ideological biases.

As a result, algorithmic bias can impact individuals’ access to resources and services, the level of surveillance they experience, their treatment by police and government, access to education, insurance or loan. In this way, if not implemented properly, AI can introduce or reinforce disparities in societies, leading to inaccurate and unfair outcomes to diverse populations

Human judgment is still needed to ensure AI supported decision making is fair.

PREDICTIVE POLICING

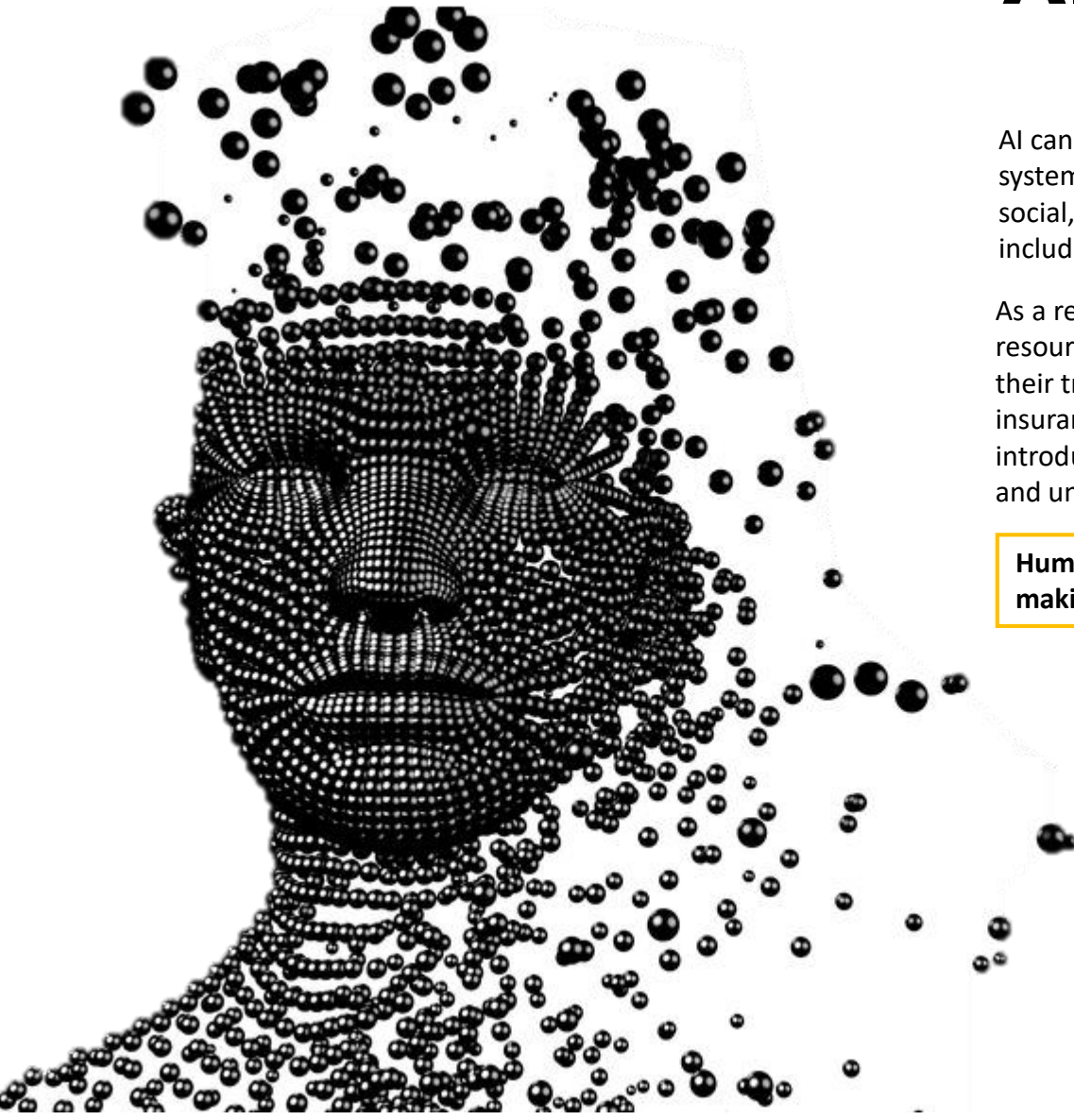
Around the United States AI algorithms are deciding on whether to detain or release defendants, providing recommendations on prison sentences. A ProPublica Report showed evidence that the AI system was biased against black defendants as it constantly recommended longer sentences in comparison to white counterparts for the same crime.

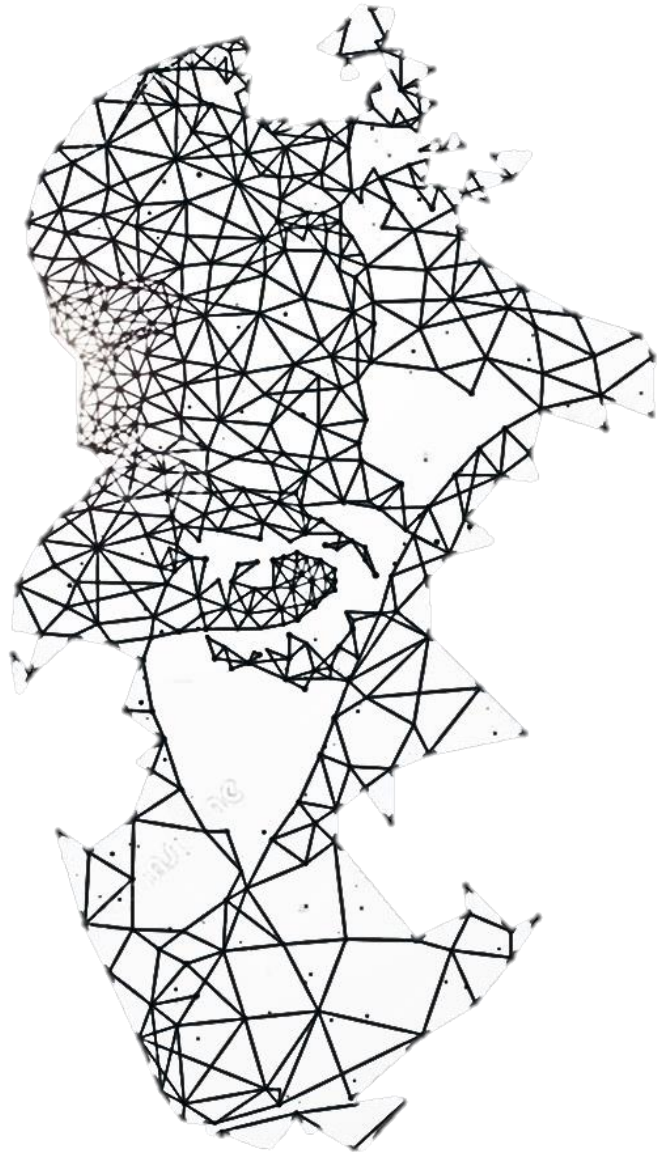


[Read full ProPublica Report](#)

A 2018 study conducted at UC Berkeley found that consumer lending technology discriminates against minority applicants.

Consumer-Lending Discrimination in the FinTech Era





AI CHALLENGES

AI AND SAFETY

AI safety refers to the technical design of AI systems that aims to avoid potentially negative impacts, accidents, unintended and harmful consequences of poor design.

Researchers at Open AI define three areas of AI safety: **SPECIFICATION**, which ensures that an AI system's behavior aligns with operators' true intentions; **ROBUSTNESS**, which ensures that an AI system continues to operate within safe limits upon perturbations; **ASSURANCE**, which ensures that we can understand and control AI systems during operations.

Open AI defines a goal of long-term AI safety as ensuring that advanced AI systems are aligned with human values, that they reliably do things that people want them to do.

i MICROSOFT CHATBOT

Microsoft's chatbot, Tay, is a recent example of how an AI can learn negative behavior from its environment, producing results quite the opposite from what its creators had in mind.

i TESLA ACCIDENT

In 2016, the Tesla car on autopilot had an accident, in which the vehicle mistook a white truck for a clear sky. The accident took a driver's life. This is an example of an AI misunderstanding its surrounding and taking deadly action as a result.

"We would want the solution to the safety problem before somebody figures out the solution to the AI problem."

Nick Bostrom, philosopher, author of "Superintelligence"

VALUE ALIGNMENT PROBLEM

Any consideration of AI safety must also include value alignment: **how can we design artificial intelligence that can align with the global diversity of human values?**

GLOBAL EXISTENTIAL RISKS

Existential risk from artificial general intelligence (AGI) is the hypothesis that substantial progress AGI could someday result in human extinction or some other unrecoverable global catastrophe.

A 2008 SURVEY BY THE FUTURE OF HUMANITY INSTITUTE ESTIMATED A

5%

PROBABILITY OF EXTINCTION BY SUPERINTELLIGENCE BY 2100



Future of Humanity Institute

AI CHALLENGES

AI AND PRIVACY

In a world of smart and connected devices in our homes, workplaces and public spaces, there is vast amount of data those devices, networks, and platforms generate, process, or share. Facing the fact that this data can be exploited becomes increasingly pressing. The possibility of misusing data for harmful profiling, tracking or identifying people becomes an increased concern for policymakers, and has led to greater efforts in protecting data privacy.

Concerns over consumer and citizen privacy have peaked in recent years, in line with, among other things, the rise of advanced technologies like AI. AI relies on large amounts of data to make accurate inference, predictions and decisions, which brings additional challenge for societies, industries and governments in **balancing the need for more data with the protection of consumers and citizens.**

TECHNICAL SOLUTIONS TO PRIVACY QUESTION



DIFERENTIAL PRIVACY

Differential privacy, which randomly perturbs data, so that any individual's record is obscured, while amalgamated features are mostly left intact.



HOMOMORPHIC ENCRYPTION

Allows computations to be performed on fully encrypted data, so that the raw data is never seen by the machine learning system.

EXAMPLES OF POLICY RESPONCES



Europe's General Data Protection Regulation is the privacy legislation that went into effect in May 2018.



The California Consumer Privacy Act (CCPA), enacted in 2018, creates new consumer rights relating to the access to, deletion of, and sharing of personal information that is collected by businesses.



Similar laws are being pursued in a handful of other states.

SURVEILLANCE ISSUE

Growing number of states are deploying advanced AI tools to monitor, track, and surveil citizens. Sensor networks, social media tracking, facial recognition, affect recognition are some of the ways AI has been used in surveillance in US, China and many other countries.

FACIAL RECOGNITION

Facial recognition potentially presents risks to human rights and civil liberties in many countries throughout the world. It can reinforce skewed and potentially discriminatory practices from criminal justice to education.

AFFECT RECOGNITION

It's a technology that attempts to read inner emotions by a close analysis of the face and is connected to making claims about people's mood, mental health, level of engagement, and guilt or innocence. It is already being used for discriminatory and unethical purposes, often without people's knowledge.

AT LEAST 75 OUT OF 176 COUNTRIES GLOBALLY ARE ACTIVELY USING AI TECHNOLOGIES FOR SURVEILLANCE PURPOSES



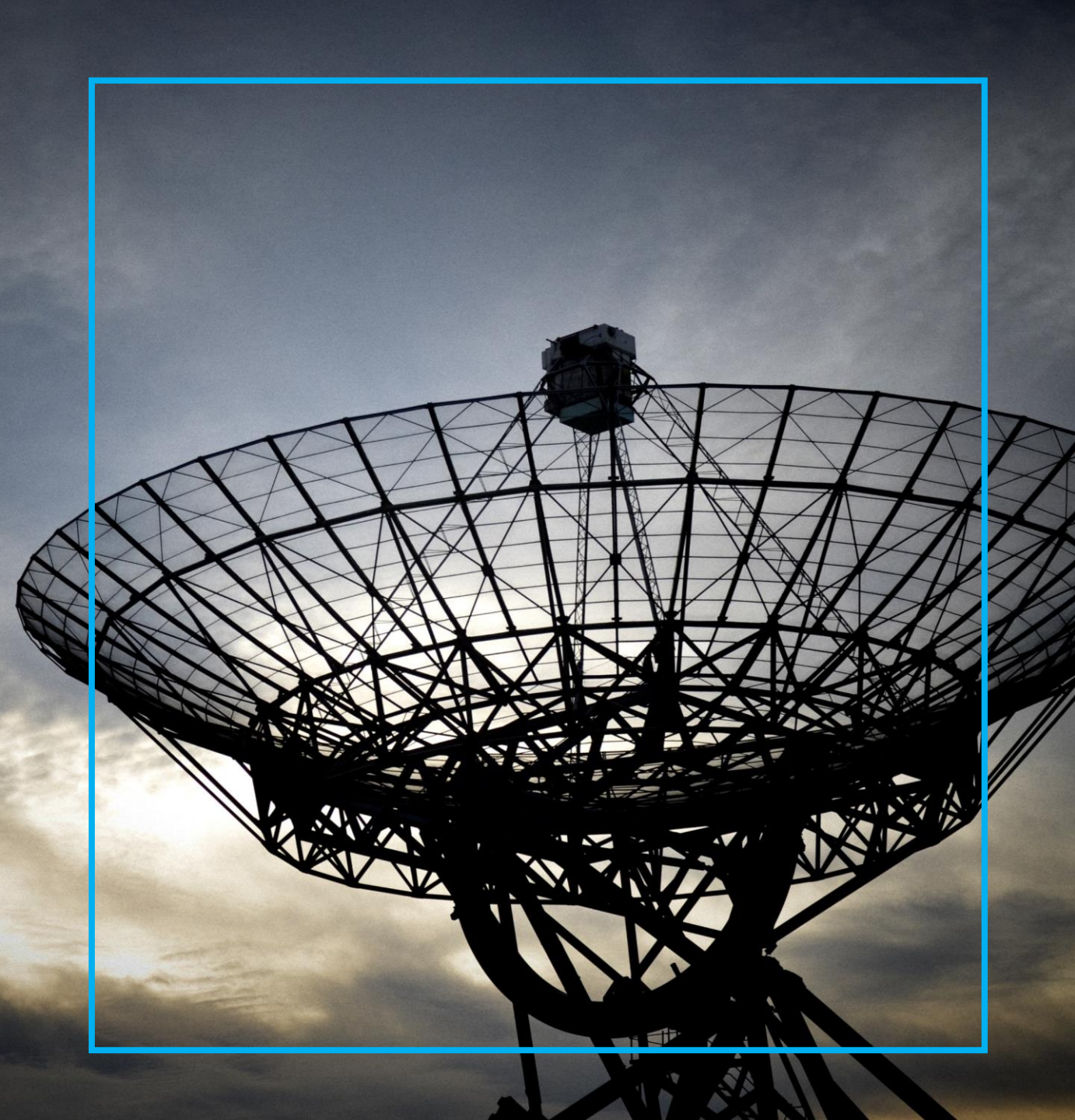
51% OF ADVANCED DEMOCRACIES AND

37% CLOSED AUTOCRATIC STATES



The above data is from AIGS - AI Global Surveillance Index

The index does not distinguish between AI surveillance used for legitimate purposes and unlawful digital surveillance.



THE **FUTURE** OF AI

43

THE ROLE OF POLICYMAKERS



”

By the time we get to the 2040s, we'll be able to multiply human intelligence a billionfold. That will be a profound change that's singular in nature. Computers are going to keep getting smaller and smaller. Ultimately, they will go inside our bodies and brains and make us healthier, make us smarter.

Ray Kurzweil

Director of Engineering at Google,
Co-founder of Singularity University

THE FUTURE OF AI

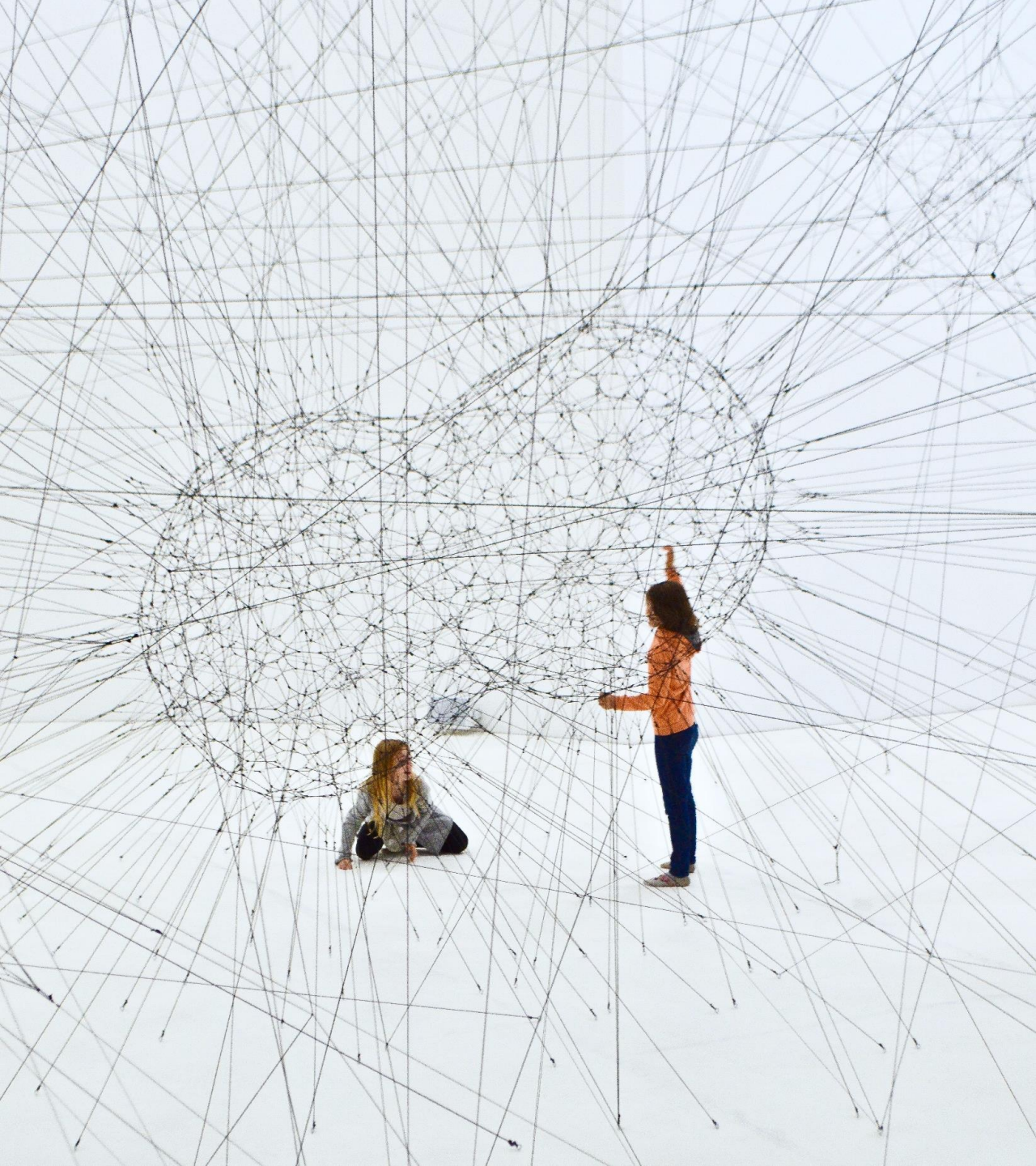
WHAT IS THE ROLE OF POLICYMAKERS?

Machines are predicted to be better than humans at translating languages by 2024, writing high school essays by 2026, driving a truck by 2027, working in retail by 2031, writing a bestselling book by 2049 and surgery by 2053. In fact there is a 50 per cent chance that machines will outperform humans in all tasks within 45 years, say respondents to the survey conducted by the University of Oxford and Yale University in 2015. [The Survey](#) was distributed to 352 AI experts, who shared their views on the future of AI.

Not all experts agree when and if the AI superintelligence will be developed and how it will coexist with humans. The researchers put only a 5 percent chance on computers bringing about outcomes near the level of human extinction. A quarter of the eminent AI researchers [surveyed by Etzioni in 2016](#) said they thought superintelligence would never materialize at all.

[An Oxford study](#) claimed that 47% of jobs were at risk of becoming automated in the next 20 years, which increased concern for job losses. Still, other experts claim that AI and automation will primarily impact specific industries, while potentially creating more jobs in the process.

All these topics and others mentioned in this guide need to be on the agenda of policymakers.



THE ROLE OF POLICYMAKERS

Policymakers have a responsibility to engage in AI development, planning, and discussions. AI is deeply pervasive and influences many spheres of our lives, impacting security, democracy, freedom, human rights.

Stopping AI is not an option, and **we need to be quicker in comprehending this development and impact, as well as shape it in ethical, safe and responsible ways through effective policymaking.** The future of AI is not the one we dream of or the one we fear of, but the one we plan for and the one we carefully design for. AI can help create a better world, if we build it right.

Only with thoughtful and responsible policymaking we will be able to successfully incorporate AI into our society.

ADDITIONAL AI POLICY RESOURCES



There needs to be more cross-talk between industry, civil society, and the academic organizations working to advance these technologies and the government institutions that are going to be representing them.

Terah Lyons, Founding Executive Director, Partnership on AI



INCLUSIVE AI RESOURCES

This list is compiled by the CITRIS Policy Lab at CITRIS and the Banatao Institute, UC Berkeley.



TOWARD AI SECURITY: GLOBAL ASPIRATIONS FOR A MORE RESILIENT FUTURE

Jessica Cussins Newman; The Center for Long-Term Cybersecurity; February 2019



UNITED NATIONS: AI REPOSITORY

AI related projects, research initiatives, think-tanks and organizations that can accelerate progress towards the SDGs.



AI NOW 2019 REPORT

Kate Crawford et al. AI NOW Institute; New York; December 2019



AI GOVERNANCE: A RESEARCH AGENDA

Allan Dafoe; Governance of AI Program; Future of Humanity Institute; University of Oxford; August 2018



THE AI SHIFTS: IMPLICATIONS FOR POLICYMAKERS

Sarah Villeneuve and Nisa Malli; The Brookfield Institute for Innovation + Entrepreneurship; Ryerson University; May 2018



THE EUROPEAN ARTIFICIAL INTELLIGENCE LANDSCAPE

Charlotte Stix; European Commission; April 2018



ARTIFICIAL INTELLIGENCE: A POLICY ORIENTED INTRODUCTION

Anne Bowser, Michael Sloan, Pietro Michelucci, Eleonore Pauwels; The Wilson Center; November 2017



MAKING THE AI REVOLUTION WORK FOR EVERYONE

Nicolas Mialhe, Cyrus Hodess; The Future Society at Harvard Kennedy School; 2017

ABOUT THE AUTHOR

BRANKA PANIC



AI

Peace

Branka Panic is the Founder and Executive Director of AI for Peace, a nonprofit ensuring artificial intelligence (AI) benefits peace, security and sustainable development and where diverse voices influence creation of AI and related technologies.

Branka is a passionate advocate for positive peace, with 10 years of experience in policymaking, working with governments and think-tanks across the globe. She is a co-founder and Board Member of the Center for Exponential Technologies, connecting policy and the tech world.

(E)



She holds a Master's degree in International Development Policy, from Duke University, Sanford School of Public Policy, and a Master's degree in International Security from the University of Belgrade. She studied International Peace and Conflict Resolution at the University of North Carolina.

She is a global citizen who found her temporary habitat in the Bay Area, CA.

ACKNOWLEDGMENTS

The author is grateful for the contributions of many colleagues from the policy and AI worlds. This guide would not have been possible without them sharing their views and experiences about the implementation of AI and intersections with policymaking.

WE WELCOME QUESTIONS AND COMMENTS ABOUT THE GUIDE

YOU CAN FIND US AT:

 www.aiforpeace.org



@AI4Peace
#AI4Policymakers
#AIGuide

YOU CAN CONTACT THE
AUTHOR DIRECTLY AT



Medium



PUBLISHED BY



SUPPORTED BY



IN COOPERATION WITH

