



The Global Initiative for
Ethical Considerations in the
Design of Autonomous Systems

v.1

ETHICALLY ALIGNED DESIGN

A Vision for Prioritizing Human Wellbeing with
Artificial Intelligence and Autonomous Systems





HOW TO USE ETHICALLY ALIGNED DESIGN

Thank you for reading this brochure describing our document, **Ethically Aligned Design (EAD)**. The brochure features information about and quotes from The Global Initiative for Ethical Considerations in the Design of Autonomous Systems to provide you a sense of the people and expertise behind the creation of Ethically Aligned Design before you download and read it.

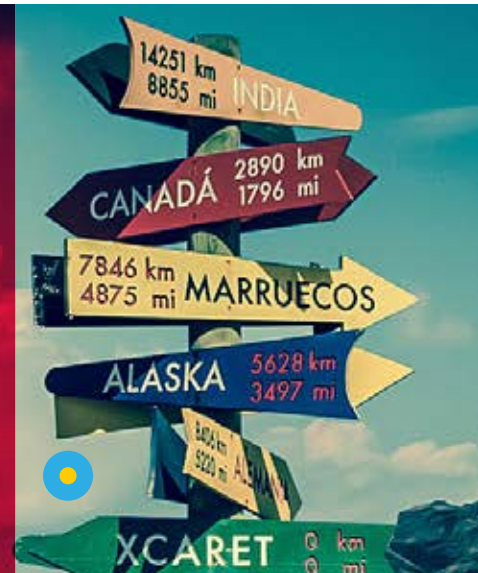
Audience for EAD and how we're hoping you'll use it:

- If you're an Educator, we encourage you to read EAD with your students to get a sense of why ethical considerations are so critical to consider in the design of Artificial Intelligence and Autonomous Systems (AI/AS). Like STEM education, we believe Ethics should be taught at University and even secondary school levels so the fundamentals of honoring end user values are considered in every field of academia, science and policy focusing on AI/AS today.
- For Governments or Government agencies we hope EAD will provide guidance for educating the general public or informing your Codes of Ethics and thinking about this topic.
- If you're an Engineer, Academic, Scientist, or Technologist creating AI/AS:
 - We hope recommendations in EAD may give some practical guidance addressing ethical dilemmas in your work or helping to define future research projects.
 - We hope EAD could provide material for educating Members and the general public while also providing ideas to update Codes of Ethics for engineering or scientific professional Associations.
 - We'd welcome your feedback on our document and to consider potentially joining one of our Committees or Working Groups forming around Standards Projects inspired by our work. This is the first Version of EAD, and we need your help to make it better.
- If you're a student or someone from the general public who is interested in the ethical aspects of AI/AS, you're more than welcome to join our work or provide comments on the Google Doc version of Ethically Aligned Design, available [here](#).
- If you're a journalist and would like to know more about The Global Initiative or EAD, please get in touch.

Next Steps:

- Interested in Joining a Committee or Working Group for one of our Standards Projects?
 - [Click here to get in touch.](#)
- Want to provide actionable critique to Ethically Aligned Design? Read the PDF version of the document [here](#) and then provide comments on the Google Doc Version [here](#).

Thank you in advance for your interest and support.





INTRODUCTION

Fully automated personalization is an oxymoron.

It may seem strange to take a measure of our values in the algorithmic era but the rigor of honest introspection defines who we are.

By outsourcing our ethical deliberations to intelligent agents we sacrifice the human code that should program our future. While Artificial Intelligence and Autonomous Systems (AI/AS) offer huge opportunity, it is only within their contextual applications we'll determine their benefit or risk.

Ethically Aligned Design represents the collective input of over one hundred global thought leaders in the fields of Artificial Intelligence, ethics, philosophy, and policy from the realms of academia, science, and the government and corporate sectors.

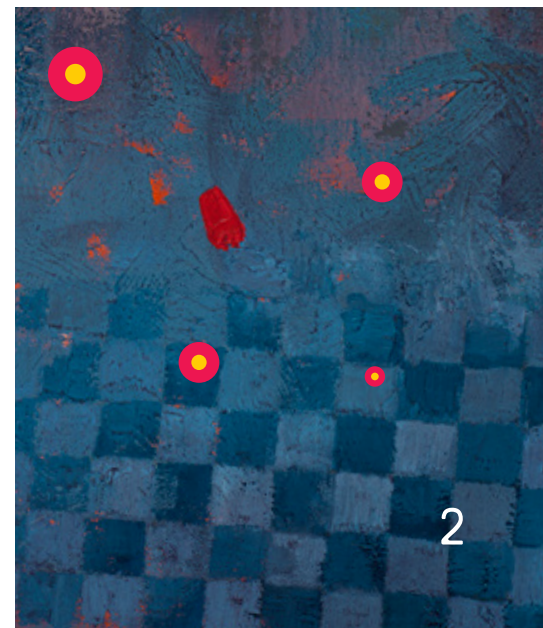
The first version of this document is designed to evolve via the insights received by its users and readers. Released under a Creative Commons license for free, it is intended for pragmatic application to help society move beyond both the fear and the uncritical admiration regarding autonomous and intelligent technologies.

Our insights are informed by a desire to incorporate aspects of human wellbeing that may not automatically be considered in the current design and manufacture of these tools. Our aspiration is to reframe the notion of success so human progress can include the intentional prioritization of individual, community, and societal values.

It doesn't hinder innovation to prioritize wellbeing - *it redefines it*. We believe the identification and implementation of values-based AI/AS will accomplish this transformation by ethically aligned design.

"Herakleitos said the paradox of change is that only something that preserves its core can undergo transformation, otherwise it will be substituted by something else. As technology takes society to spaces beyond our imagination, the question is how we can evolve and still preserve our core -- that what makes us human. Ethically Aligned Design is an IEEE-supported collective effort to precisely address this question. It represents a milestone for developing methodologies that will ensure humanity utilizes technology that inherently prioritizes our wellbeing and takes our explicit values into account. Only by maintaining our agency can we move beyond the fears associated with these technologies and bring valued benefits to humanity today and for the future."

Konstantinos Karachalios,
Ph.D, Managing Director of The Institute of Electrical and Electronics Engineers (IEEE)
Standards Association and Member of the Management Council of IEEE





OUR PROGRAM

The Global Initiative for Ethical Considerations in the Design of Autonomous Systems is an official Program of IEEE, the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity with over 400,000 members in more than 160 countries.

However, Ethically Aligned Design is not an official document or position statement of the IEEE at this time. The Initiative is serving as an incubator designed to bring together multiple voices in the Artificial Intelligence and Autonomous Systems (AI/AS) communities to find consensus on issues before subject matter is selected for submission into formal IEEE processes.

This is why IEEE will not apply its usual copyrights policy to this document, made available to any organization for free under a Creative Commons agreement. Our intention is that any organization (including IEEE) can adopt aspects of this work at their discretion at any time.

IEEE's mission is to "advance technology for the benefit of humanity." By supporting The Initiative and this document IEEE is demonstrating how critically important ethical considerations in AI/AS are an issue for all of society to address today, together.

OUR PURPOSE

To ensure every technologist is educated, trained, and empowered to prioritize ethical considerations in the design and development of autonomous and intelligent systems.

By "technologist", we mean anyone involved in the research, design, manufacture or messaging around autonomous and intelligent technologies. You're the creators and evangelists in the trenches working to ensure these systems are as beneficial and safe as possible for customers, citizens, and society. Our goal is that Ethically Aligned Design will provide insights and recommendations from peers that provide a key reference for your work in the coming years.

"Ethically Aligned Design and the work of our Global Initiative is focused on empowering technologists to prioritize ethical considerations in the creation of Artificial Intelligence and Autonomous Systems. Rather than assume a machine or system will de facto provide positive benefits, we must determine and align with the values of society before its implementation."

Raja Chatila,
(Initiative Chair) CNRS-Sorbonne UPMC Institute of Intelligent Systems and Robotics, Paris, France;
Member of the French Commission on the Ethics of Digital Sciences and Technologies CERNA; Past
President of IEEE Robotics and Automation Society



"As a society, we cannot move forward in a spirit of fear around the creation of Artificial Intelligence and Autonomous Systems.

By ensuring ethical methodologies become industry standard in the creation of these technologies we'll shift from a spirit of paranoia to pragmatism and redefine innovation around which machines or systems best honor the values of its users."

Kay Firth-Butterfield,
(Initiative Vice-Chair) Executive Director,
AI Austin



OUR PROCESS

Our Initiative is global, open and inclusive, welcoming all individuals or representatives of organizations dedicated to advancing technology for humanity by prioritizing the use of ethical considerations in the design of autonomous and intelligent systems.

We are just getting started in a consensus building process that will take years to complete. We'll be updating and iterating Ethically Aligned Design for the next two to three years based on input from experts within our committees, at our public events, and from the public at large.

OUR APPRECIATION

While we have more than one hundred experts from all but one continent involved in our work, (Antarctica AI experts, please get in touch) most of us come from North America and Europe. We are aware we need to expand our cultural horizons and get more people involved from around the world as we continue to grow our document and our efforts. We are eager for these new voices and perspectives to join our work.

Ethically Aligned Design is not a code of conduct or a professional code of ethics. Engineers and technologists have well-established codes along these lines and we wish to respectfully recognize the formative precedents surrounding issues of ethics and safety and the professional values these Codes represent. These Codes provide the broad framework for the more focused domain of AI/AS we address in this document and it is our hope that the inclusive, consensus building process around its design will contribute unique value to technologists and society as a whole.

This document is also not a position or policy statement, or formal report. It is intended to be a working reference tool created in an inclusive process by those in the AI/AS Community prioritizing ethical considerations in their work. While we have done our best to create this first version of Ethically Aligned Design with rigor, we realize we're just beginning. We need your critique, insights and participation to make this project flourish.

Thank you in advance for your interest and support.

*"How will machines know what we value if we don't know ourselves?
Ethics and values-driven design provide tools for introspection technologists should
prioritize as we build the machines and systems guiding our lives for the future.
We can't positively increase human wellbeing if we don't take the time to identify our
collective values before creating technology we know will align with those ideals"*

John C. Havens,
Executive Director of The Global Initiative for Ethical Considerations in the Design of Autonomous
Systems, author, *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*





1 GENERAL PRINCIPLES

The General Principles Committee seeks to articulate high-level ethical concerns that apply to all types of artificial intelligence and autonomous systems that:

- 1) Embody the highest ideals of human rights that honor their inherent dignity and worth.
- 2) Prioritize the maximum benefit to humanity and the natural environment.
- 3) Mitigate risks and negative impacts as AI/AS evolve as socio-technical systems.

It is our intention that by identifying issues and candidate recommendations regarding these principles they will eventually serve to underpin and scaffold future norms and standards within a new framework of ethical governance.

"For autonomous and intelligent systems to be trusted, and hence bring the greatest benefit, they must be designed and operated ethically. It is vital therefore that we build such systems on a strong foundation of ethical principles."

Alan Winfield,
(Co-Chair) Professor and Director of the Science Communication Unit,
Bristol Robotics Lab, University of the West of England

"The Principles that best honor human dignity should be mirrored in the ethical considerations we utilize when creating future technologies. Artificial Intelligence and Autonomous Systems should prioritize maximum benefit for humanity to ensure society flourishes long into the future."

Kay Firth-Butterfield,
(Co-Chair) Executive Director, AI Austin





2

EMBEDDING VALUES INTO AUTONOMOUS INTELLIGENT SYSTEMS

In order to develop successful Autonomous Intelligent Systems (AIS) that will benefit our society, it is crucial for the technical community to understand and be able to embed relevant human norms or values into their systems.

- 1) **Identify** the norms and values of a specific community affected by an AIS;
- 2) **Implement** the norms and values of that community within the AIS; and,
- 3) **Evaluate** the alignment and compatibility of those norms and values between the humans and the AIS within that community.

"The alignment of values between a system and its user is of critical importance to ensure Artificial Intelligence and Autonomous Systems increase human wellbeing while optimizing innovation."

AJung Moon,
(Co-Chair) Co-founder of the Open Roboethics initiative, and PhD Candidate and Visiting Scholar

"Future organizations will fundamentally base their work in AI and AIS on the ethical considerations of users and their values. Companies prioritizing these issues will have a market advantage over competitors who ignore their critical importance."

Francesca Rossi,
(Co-Chair) Full Professor, computer science at the University of Padova, Italy, currently at the IBM Research Center at Yorktown Heights, NY





3

METHODOLOGIES TO GUIDE ETHICAL RESEARCH

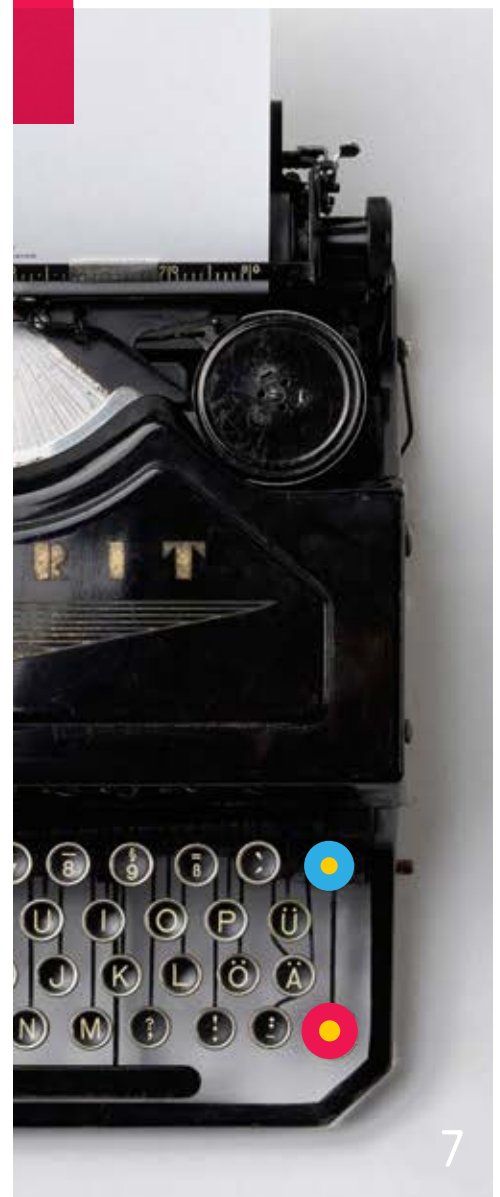
In order to create machines that enhance human wellbeing, empowerment and freedom, system design methodologies should be extended to put greater emphasis on human values as a form of human rights such as those acknowledged in the Universal Declaration of human rights. We therefore strongly believe that values-based design methodology should become an essential focus for the modern organization.

"Modern system design should be extended to put greater emphasis on human rights as a primary form of human values. Values-aligned design methodologies provide pragmatic tools for modern technologists to best honor societal needs while redefining innovation in the algorithmic era."

Raja Chatila,
(Initiative Chair) CNRS-Sorbonne UPMC Institute of Intelligent Systems and Robotics, Paris, France; Member of the French Commission on the Ethics of Digital Sciences and Technologies CERNA; Past President of IEEE Robotics and Automation Society

"Artificial Intelligence and Autonomous Systems need first and foremost to enhance human well being. This cannot be an afterthought, and as such ethics needs to be part of the design methodology. Our committee has focused on how AI/AS organizations can ensure that their system design AI/AS methodologies are based on a values-aligned design methodology, that engenders human dignity and respects human rights."

Corinne Cath,
(Co-Chair) PhD student at The University of Oxford, Programme Officer at ARTICLE 19





4

SAFETY & BENEFICENCE OF ARTIFICIAL GENERAL INTELLIGENCE (AGI) & ARTIFICIAL SUPERINTELLIGENCE (ASI)

Future highly capable AI systems (sometimes referred to as artificial general intelligence or AGI) may have a transformative effect on the world on the scale of the agricultural or industrial revolution, which could bring about unprecedented levels of global prosperity. It is by no means guaranteed however that this transformation will be a positive one without a concerted effort by the AI community to shape it that way.

"The AI community needs to encourage and promote the sharing of safety related research and tools as a primary way to imbue ethical considerations within their work."

Richard Mallah,
(Co Chair) – Director of Advanced Analytics,
Cambridge Semantics; Director of AI Projects,
Future of Life Institute



"As AI systems become more useful and capabilities increase, unintended behaviors and accidents will pose correspondingly greater risks. It's essential that the AI community adopt some best practices from computer security, where systems and their safety/security measures are subjected to highly rigorous assessments before seeing wide adoption."

Malo Bourgon,
(Co-Chair) – COO, Machine Intelligence Research
Institute



5

PERSONAL DATA/ ACCESS CONTROL

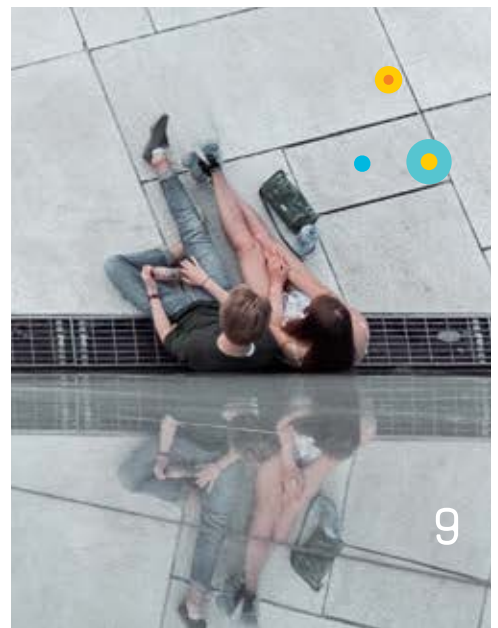
A key ethical dilemma regarding personal information is data asymmetry. To address this asymmetry there is a fundamental need for people to define, access, and manage their personal data as curators of their unique identity. We realize there are no perfect solutions, and that any digital tool can be hacked. But we need to enable a data environment where people control their sense of self. Our goal is to envision the tools and evolved practices that will eradicate data asymmetry to project a positive image of our future.

"Personal Data forms the bedrock of the algorithmic economy. Prioritizing ethical considerations regarding the use of this data by autonomous and intelligent technologies means we'll help individuals gain clarity around their digital assets while improving the quality of information provided to the systems we're building to best guide our future."

Michelle Dennedy,
(Co-Chair) Vice President, Chief Privacy Officer, Cisco; Author,
The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value

"Along with personalization methodologies that track our actions, we need tools to manage the personal data reflecting our intentions and subjective identity. Ethical considerations for AI/AS must account for and align with these values to best increase human wellbeing."

John C. Havens,
(Co-Chair) Executive Director, The Global Initiative for Ethical Considerations in the Design of Autonomous Systems and
author of, *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines.*





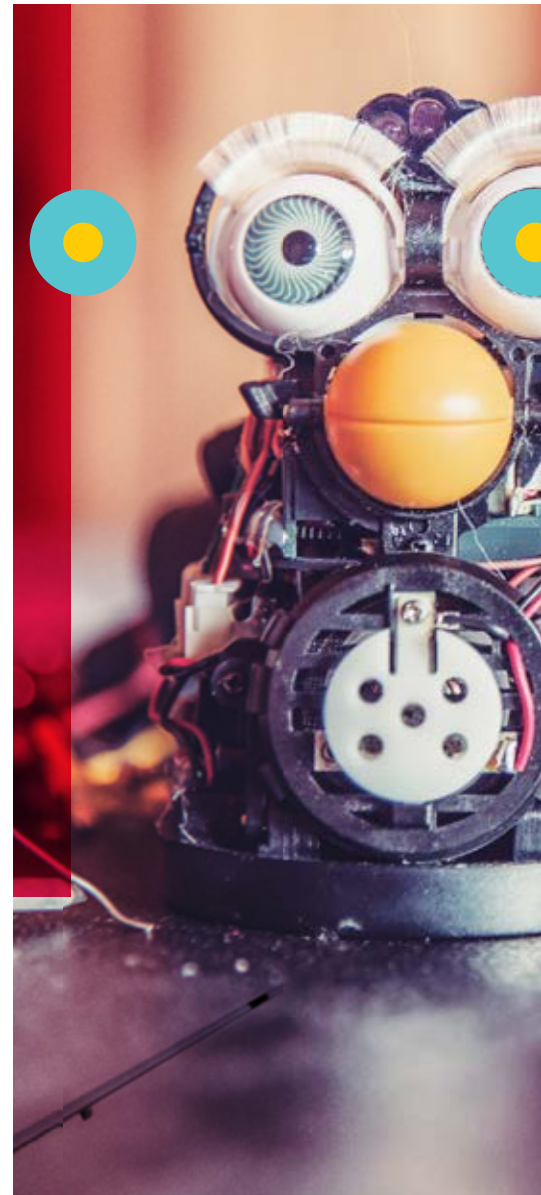
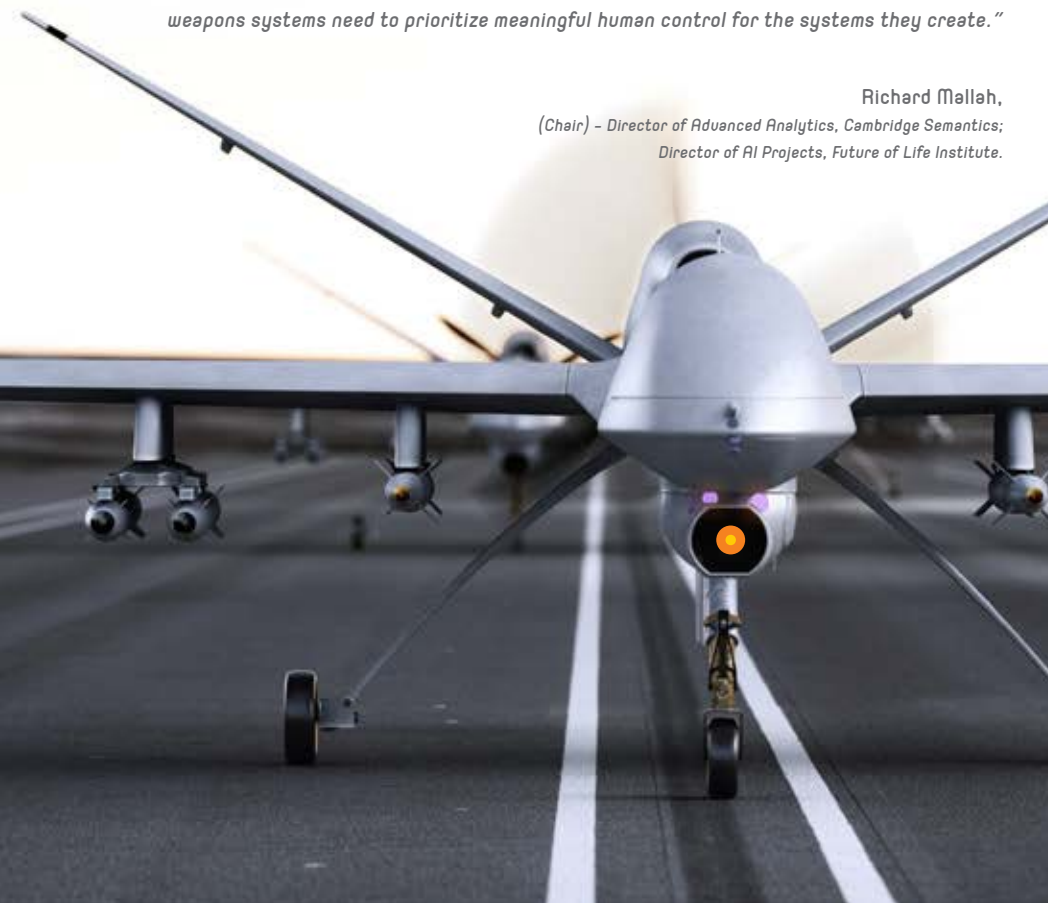
6

REFRAMING AUTONOMOUS WEAPONS SYSTEMS

Autonomous systems that are designed to cause physical harm have additional ethical ramifications as compared to both traditional weapons and autonomous systems that aren't designed to cause harm. Professional ethics about these can and should have a higher standard covering a broader array of concerns. Broadly, we recommend that technical organizations accept that meaningful human control of weapons systems is beneficial to society, that audit trails guaranteeing accountability ensure such control, that those creating these technologies understand implications of their work, and that professional ethical codes appropriately address works that are intended to cause harm.

"Ethical considerations and codes of Ethics designed to guide technologists creating autonomous weapons systems need to prioritize meaningful human control for the systems they create."

Richard Mallah,
(Chair) - Director of Advanced Analytics, Cambridge Semantics;
Director of AI Projects, Future of Life Institute.





7 ECONOMICS/ HUMANITARIAN ISSUES

Technologies, methodologies, and systems that aim at reducing human intervention in our day-to-day lives are evolving at a rapid pace and are poised to transform the lives of individuals in multiple ways. The aim of our multi-stakeholder committee is to identify the key drivers shaping the human-technology global ecosystem and address economical and humanitarian ramifications, and to suggest key opportunities for solutions that could be implemented by unlocking critical choke points of tension. The goal of our recommendations is to suggest a pragmatic direction related to these central concerns in the relationship of humans, their institutions and emerging information-driven technologies, to facilitate interdisciplinary, cross-sector dialog that can be more fully informed by expert, directional, and peer-guided thinking regarding these issues.

"Artificial Intelligence and Autonomous Systems need to prioritize ethical considerations in their design to ensure the equal distribution of their benefits while reducing harm to society. Otherwise they will not be designed for the increase of wellbeing for all of humanity but simply for those individuals who are privileged to reap their benefits"

Raj Madhavan,
(Chair) Founder & CEO of Humanitarian Robotics Technologies, LLC, Maryland, U.S.A.





8 LAW

The early development of Artificial Intelligence and Autonomous Systems (AI/AS) has given rise to many complex ethical problems. These ethical issues almost always directly translate into concrete legal challenges—or they give rise to difficult collateral legal problems. There is much to do for lawyers in this field that thus far has attracted very few practitioners and academics despite being an area of pressing need. Lawyers should be part of discussions on regulation, governance, domestic and international legislation in these areas and we welcome this opportunity to ensure that the huge benefits available to humanity and our planet from AI/AS are thoughtfully stewarded for the future.

"It is essential that the laws created to guide AI/AS are built to best honor the community and societal values of the communities in which they're developed."

Derek Jinks,
(Co-Chair) University of Texas Law School; Consortium on Law and Ethics of Artificial
Intelligence and Robotics, Strauss Center, University of Texas

"Increasing human wellbeing means creating and adapting our laws to mirror the values we want to develop in ourselves, society, and the machines we build in the future."

Kay Firth-Butterfield,
(Co-Chair) Executive Director, AI Austin





NEW COMMITTEES¹

Classical Ethics in Information & Communication Technologies

This Committee will focus on examining classical ethics ideologies (utilitarianism, etc) in light of AI and autonomous technologies.

Mixed Reality Committee

Mixed reality could alter our very notions of identity and reality over the next generation, as these technologies infiltrate more and more aspects of our lives, from work to education, from socializing to commerce. An AI backbone that would enable real-time personalization of this illusory world raises a host of ethical and philosophical questions, especially as the technology moves from headsets to much more subtle and integrated sensory enhancements. This Committee will work to discover the methodologies that could provide this future with an ethical skeleton and the assurance that the rights of the individual, including control over one's increasingly multifaceted identity, will be reflected in the encoding of this evolving environment.

"Classical ethics methodologies have, to some degree, informed Artificial Intelligence and Autonomous Systems research since 1948, originating with Norbert Wiener's Cybernetics, the first values-driven methodology that sought to systematically study aspects of inherently biased values in artificial machine intelligence. By exploring ethics from several culturally diverse traditions and applying the thousands-year-old tradition of classical ethics to values-driven methodologies in ICTs and AI design we can achieve the goal of increasing human wellbeing for a positive future."

Jared Bielby,
(Chair, Classical Ethics Committee), Co-chair,
International Center for Information Ethics

"Mixed Reality media combined with intelligent and autonomous technologies have the potential to rush us into a software-mediated world in which we see, hear and experience only what we want to see, hear and experience. This is why it is critical that technologists are trained in ethics so they can build and design technology that promotes and inspires our collective empathy, our work, ourselves and our society as a whole."

Monique Morrow
(Co-Chair, Mixed Reality Committee)
CTO New Frontiers Engineering at Cisco



NEW COMMITTEES²

Affective Computing

This Committee addresses the impact on individuals and society that autonomous systems capable of sensing, modeling, or exhibiting affective behavior such as emotions, moods, attitudes, and personality can produce. Affective computational and robotic artifacts have or are currently being developed for use in areas as diverse as companions, health, rehabilitation, elder and childcare, training and fitness, entertainment, and even intimacy. The ethical concerns surrounding human attachment and the overall impact on the social fabric may be profound and it is crucial that we understand the trajectories that affective autonomous systems may lead us on to best provide solutions that increase human wellbeing in line with innovation.

Policy: Effective Policymaking for Innovative Communities involving Artificial Intelligence and Autonomous Systems (EPICAIAS)

This Committee will: (1) explore how effective policymaking employing autonomous and intelligent technologies can be done in a rapidly changing world, (2) generate recommendations on what initiatives the private and public sector should pursue to positively impact individuals and society, and (3) illuminate newer models of policymaking both extant and experiment to support the innovation of AI/AS for shared human benefit.

"We need to provide ethical guidance regarding the appropriate design and use of affective computing within AI/AS to ensure that it does not violate the rights of users and society as a whole while at the same time assuring benefits to those who knowingly employ it for their own enjoyment and well being."

Ronald C. Arkin,
(Affective Computing Committee Co-Chair)
Regents' Professor & Director of the Mobile
Robot Laboratory; Associate Dean for Research
& Space Planning, College of Computing Georgia
Institute of Technology

"We need to be clear that the decision to use affect in intelligent systems has significant ethical ramifications. While there is clear utility to emotional systems in natural intelligence, both for individual control and social coordination, the confounding of emotion, suffering, and moral status in familiar natural examples of intelligence makes transparency concerning the role and nature of affect in AI particularly difficult and important."

Joanna Bryson
(Affective Computing Committee Co-Chair)
Visiting Research Collaborator, Center for
Information Technology Policy, Princeton
University; Associate Professor, University
of Bath, Intelligent Systems Research Group,
Department of Computer Science



NEW COMMITTEES³

ADDITIONAL QUOTES

"As mixed reality and associated technologies evolve, they will inevitably intermingle and become increasingly driven by AI. Just as "artificial intelligence" will eventually be seen simply as "intelligence," the convergence of these technologies that control perception and simulate reality will eventually be seen simply as "reality." That scenario raises a host of ethical concerns as well as questions about our very notions of self, identity, and reality."

Jay Iorio,

(Co-Chair, Mixed Reality Committee) Director of Innovation, IEEE Standards Association

"We need both policies and places in public service where we can collaborate with citizens and private sector partners on new ways of doing the business of public service better -- to include artificial intelligence and autonomous systems. The benefits of artificial intelligence to individual nations and the world are in the civilian domain, more so than any other domain."

Dr. David R. Bray,

(EPICRIAS Committee Co-Chair) Senior Executive & CIO for the FCC; Eisenhower Fellow to Taiwan and Australia; Harvard Visiting Executive In-Residence

"AI and autonomous systems will influence many aspects of life, business, health, and education. As a result, the ethical considerations are fraught with complexity. Standing as a bridge between policymakers and the commercial sector, The Global Initiative offers a uniquely impartial and balanced perspective in this global conversation. This work unites the public and private sectors, to the benefit of all"

Michael Krigsmen,

Industry analyst and host of CHOTALK