AI & Society

THE JOINT COMPLEXITY OF AI SYSTEMS AND THE SOCIETIES INTO WHICH THEY ARE INTEGRATED DEMANDS A MULTIDISCIPLINARY APPROACH.

The JHU BDP Cluster on AI & Society will include scholars from across the social sciences, engineering and technology, to enable us to address key challenges from multiple perspectives, to help ensure the just and equitable integration of AI technology in society.

In order to achieve this vision, the cluster leads identified eight research focus areas of interest to guide the recruitment process. These research focus areas include: AI Ethics & Human Values, AI Policy & Governance, Participatory Engagement, Impact of AI on Information & Democracy, Cultural Competence, Impact of AI on Health, AI-Human Interaction, and Social Robotics.

AI Ethics & Human Values

AI has quickly become pervasive within and across many sectors of society. As is often the case with emerging technologies, broad use has preceded the conceptual and empirical research and norm development necessary for the integration of this new technology into our lives. Unlike many traditional biomedical technologies, the technologists developing AI have little or no exposure to the ethics and norms related to human subject research and the social scientists and humanists involved in this space. This gap or disconnect has implications at all stages of conceptualization, research, development, and deployment of AI-based technologies.

Without building ethical principles and assurance into human-AI interaction, AI-enabled systems are likely to fail to respond to broadly-held human values and may perpetuate and exacerbate social inequities. It is also imperative to understand the conditions under which trust and distrust in human-AI interaction arise in applications (and the institutions that deploy them) and how to embed safety, security, fairness, reliability, transparency, ethics, and justice into human-AI interaction.

Close collaboration between the technology and social science communities is critical. This will ensure not only that ethical and societal considerations are taken into account at all stages of conceptualization, research, development, and deployment, but also facilitate the development of ethics-driven AI applications and technically-informed conceptual work that will be critical to a human understanding of the meaning of AI in itself and as part of our lives.

AI Policy & Governance

Policymakers have always struggled to keep up with advances in technology and the often-unanticipated consequences for individuals and society, and AI technology is no exception. There is a dire need for trusted entities to inform and counsel governmental leaders regarding the benefits, limitations, and pitfalls of AI, and even the ways in which AI can be misused by adversaries. It is also important to think broadly about all the ways in which AI can be governed,

including not only through 'hard' governance, like laws and regulations, but also through 'soft' governance approaches, such as guidelines, professional standards, and market forces.

In addition to the myriad AI application areas we see today, there are also looming questions around emerging uses of AI-enabled autonomous systems and the governance of AI that verges on having consciousness and other characteristics that have historically been the sole domain of humans.

To address these issues and to constructively move the governance conversations forward, we need faculty who are accomplished in both technology and in technology governance, both to conduct research at this crucial intersection and to act as translators and bridges between the technology and policy communities.

Participatory Engagement

As AI-enabled technologies have rapidly become pervasive across society, their impact on individuals and communities has been profound. To ultimately achieve beneficial outcomes from AI-enabled technologies, it is important to understand what people and communities consider as safe, secure, reliable, ethical, trustworthy and beneficial. Participatory engagement is an important and evolving field that seeks to develop methodologies both to elicit understanding of people's perceptions – in this case, of AI-enabled technologies – as well as provide models for direct stakeholder input into the technology development process. The field of participatory engagement includes research on, among others, the advancement of human-centered design methodologies, participatory technology and algorithmic development processes, and civic tech engagement models.

Impact of AI on Information and Democracy

Historically, democracies have depended on the dissemination of both facts and opinions through word of mouth, newspapers, and more recently, electronic broadcast media (radio and television). The advent of the internet has changed the structure of the spread of information to be less dominated by institutions such as mass media, turning instead to special-interest social media platforms and news sources with increasing levels of anonymization, targeting, and opacity. AI-based methods are being developed that exacerbate the concerns that exist about the veracity and fairness of the dissemination of information in society, and what the impact of AI technology will be on the functioning of democracy.

Cultural Competence

As AI technologies are increasingly designed to interact with humans in natural and intuitive ways, it is important to improve the cultural competence of AI. Culturally competent AI will be able to effectively respond to and interact with people from different backgrounds in a culturally appropriate manner. An underlying component of cultural competence is cultural intelligence which is the ability to interpret a person's behavior or communication in a way similar to that person's community. Advancements in natural language processing (NLP) have increased the linguistic competence of algorithms across a broad array of languages, particularly in the deployment of voice assistants and translation technologies. Moving beyond linguistic competence to achieve diverse cultural competence of AI-enabled systems will require new research focus areas at the intersections of social science and technology development.

Impact of AI on Health

The healthcare system relies on technology more and more, from managing electronic patient records, to interpreting patient medical imaging, to performing medical procedures. The increased use of AI to make decisions about insurance coverage, care plans, to perform procedures, or to advise healthcare professionals on diagnoses and treatments introduces myriad opportunities and challenges. In addition to new questions about the appropriate role of AI, especially the degree of autonomy, in these kinds of healthcare decisions, the promise of AI also introduces questions regarding research, development, and deployment, informed consent (e.g., of patients, providers), and the potential of risks such as perpetuating or exacerbating bias and inequity that must be addressed.

AI-Human Interaction

How will AI impact humans? How can we design interactive intelligent systems that are usable and beneficial to humans while respecting human values? Under a framework that integrates computer science and social science, interactive technologies powered by AI can be used to collect data and identify the places in the data pipeline where human actions and interactions with AI and other humans may drive machine learning. This new approach will revise the existing, biased algorithms and think both optimistically and critically of what AI systems can do and how they can and should be integrated into society.

We need to examine the interactive nature and process of human learning (e.g., from teacher-pupil, doctor-patient) for a target task (e.g., tutoring mathematics, MRI imaging for disease diagnosis). What are the overarching (prescriptive) social norms governing those roles? What are the emerging norms and their enforcement and sanction mechanisms governing the ongoing interactions between two or more individuals who are taking the roles in diverse communities of different social and demographic makeups?

What we learn from human-human interaction norms and their enforcement and sanction in human learning can be used to inform machine learning algorithms. This can begin an iterative, dynamic process of introducing human learning to AI machine learning, and AI learning to human learning, to adapt and advance both the human side and the AI side. Such mutual learning exemplifies what we conceive as AI-human interaction.

Social Robots

Social robotics is an example of use-inspired AI, particularly manifesting the human-AI interaction foundational concept. Taking "aging in place" and community elder care as an example, AI systems can be built to increase opportunities of social interactions for largely isolated aging individuals living at home. Serving as a knowledgeable "companion", this type of social robot unifies the human learning model and the machine learning model for the specific task of providing cognitively stimulating conversations, storytelling, consultations, psychological counseling, and game playing, etc. These social robots may enhance the mental health of the aging population, but also raise questions about society's responsibility to our elders, the replacement of human caring and touch with objects that may fulfill some of the roles of human

companions but cannot 'care' or be in a true relationship with their human users, and older individuals' understanding of the limits of those objects.