UC BERKELEY STRATEGIC PLANNING

SIGNATURE INITIATIVES WORKING GROUP

SPRING 2018 REPORT
FINAL VERSION - DECEMBER 2018

APPENDIX A1: INCLUSIVE INTELLIGENCE

BRIEF SUMMARY:

INCLUSIVE INTELLIGENCE: A new approach to artificial intelligence and Data Science that is: 1) inclusive of individuals from all backgrounds to benefit the greater good; 2) inclusive of a broad community of scholars from engineering, business, the arts, humanities, psychology, neuroscience, sociology, politics, philosophy, history, and other disciplines; 3) inclusive of developing a broad array of related approaches and technologies such as data science, artificial intelligence, robotics, sensing, machine learning, etc.; and 4) inclusive of both human and artificial intelligence and the way they interact, complement, and enhance each other. This initiative will promote both continued technological innovation and a broad investigation of the societal and ethical implications of artificial intelligence, robotics, and data sciences.

EXECUTIVE SUMMARY:

"Inclusive Intelligence" is a proposed approach to the technologies of Data Science, Artificial Intelligence, Robotics, and Sensing (DAIRS). These innovations are made possible by the vast increases in computing power, networking capabilities, and data collection developed in the last fifty years. The approach is: 1) inclusive of individuals from all backgrounds to benefit the greater good; 2) inclusive of a broad community of scholars from engineering, business, the arts, humanities, psychology, neuroscience, sociology, politics, philosophy, history, and other disciplines; 3) inclusive of developing a broad array of related approaches and technologies such as data science, artificial intelligence, robotics, sensing, machine learning, etc., and 4) inclusive of both human and artificial intelligence and the way they interact, complement, and enhance each other. The approach recognizes that DAIRS technologies will have a transformative impact on how we work, on how we learn, on how we understand the world, and on how we live our daily lives, and it recognizes that these technologies also have the potential to dramatically increase inequality of income and wealth by concentrating power among a small number of individuals, companies, and governments. Ensuring that the benefits are shared inclusively and democratically requires attention to human and societal choices. The goal of "Inclusive Intelligence" is to advance research into a digital transformation in economic, political, and social conditions that will yield benefits for all. **How can humans and machines collaborate to advance our collective and inclusive intelligence?** Berkeley has a rich history of basic and applied research in these technologies and is uniquely positioned to pursue this broader question building on our comprehensive academic excellence and on our public mission of serving the common good and contributing to a better world.

THE CHALLENGE:

We live in a world in which our biological, physical, and information technologies (including those that can affect the climate, urbanization, transportation, migration, and biological life—biotechnologies, cognitive technologies, and a host of medical interventions and systems) are pushing us into uncharted territory. These changes include the availability of massive amounts of data, the development of artificial intelligence (AI), the maturation of robotics, the deployment of new sensing and imaging technologies, the spread of the Internet and data networking, the invention of machine learning, the growth of cloud computing, the use of "Big Data" and data science to make decisions, and the construction of an Internet of Things. As there is no one phrase for these developments, we use the term "DAIRS" for Data (and data science), Artificial Intelligence, Robotics, and Sensing. The phrase seems appropriate because these new technologies are daring us to be audacious and bold. And, as with any dare, we must identify the risks and the possibilities before we leap.

DAIRS creates vast opportunities to do well for humanity—and significant reasons to worry. We are creating this world in real time, without deep wells of foresight, in ways that will profoundly affect generations to come. At UC Berkeley we are thoroughly invested in shaping the outcomes of this endeavor. "Inclusive intelligence" requires defining and acting on new forms of moral and intellectual responsibility within the technologized world. As we learn from Mary Shelley's *Frankenstein*, the goal is not just to create these new intelligent technologies, but to use them morally and responsibly.

The proper use of these technologies requires an inclusive approach that considers all dimensions of intelligence, all features of the human condition, and all groups in society. This Signature Initiative cuts across how we design and build new computing technologies, how we put them to work in society, and how we offer up forms of generative imagination and productive reflection on what they can mean and become.

This Signature Initiative will encompass research to create and use these new DAIRS technologies; research into how these technologies can be used to advance scientific investigations; research into the effects of these technologies on how we learn, work, and live and on their cultural, intellectual, and ethical impacts; and research into how we can shape and deploy these technologies to serve the common good.

WHY NOW?

DAIRS technologies are accelerating a "Digital Transformation" of businesses and governments that, when combined with shifts in demographics (aging and increasingly urban populations) and changing policies on immigration and trade, have the potential to cause dramatic changes in economies, industries, jobs, and education around the world. MIT announced a major "Intelligence Quest" project in February and Fei-Fei Lee of Stanford/Google announced a "Human-Centered AI" project in the NYTimes in early March. MIT's approach focuses on developing intelligent technologies and Stanford's approach only extends as far as making AI "human-centered." We propose a much broader endeavor that engages a wide community of scholars to design intelligent systems that advance our understanding of the world while being responsive to public interests.

WHY BERKELEY?

UC Berkeley has a long history of pioneering research on DAIRS technologies stretching back to work on cybernetics in the 1950s and 1960s and early work in Statistics, Operations Research, and EECS, by pioneers such as Jerzy Neyman and Douglas Engelbart. It also has a long history of using large data sets and sophisticated methods of analysis to advance scientific investigations such as the discovery of dark energy from the analysis of supernova data or the identification of gender bias in wages. Artistic and philosophical critiques of AI and robotics have a long history at Berkeley, epitomized by the work of Hubert Dreyfus and speakers in the Art, Technology, and Culture lecture series. Berkeley also has a long tradition of economics and social science research on the impact of technologies on economic performance, politics and policy, and societal health and cohesion. UC Berkeley today is a global center for research in these areas with over 30+ researchers affiliated with the Berkeley AI Research (BAIR) Lab and associated labs in the College of Engineering and CITRIS. The most popular textbook on AI is authored by Stuart Russell (and Berkeley PhD Peter Norvig) and machine learning pioneer Mike Jordan is the highest cited author in computer science. The Berkeley Institute of Data Science brings together dozens of doctoral and post-doctoral students and scores of senior researchers and the D-Lab trains hundreds of students in advanced data science methods. New research on AI will build on ongoing research by Berkeley's College of Engineering; Department of Statistics; School of Information; Division of Data Science; and numerous departments within the College of Letters and Science, including physics, earth and planetary science, astronomy, psychology, mathematics, neuroscience, sociology, political science, economics, philosophy, history, and other departments in the arts and humanities. It will include projects and units that investigate intelligence in the context of creativity and creative work that embraces multiple perspectives and skill sets, including Human Technology Futures (HTF), Work in an Era of Intelligent Tools and Systems (WITS) group, CITRIS, the Berkeley Center for New Media, and Berkeley's Arts + Design programs. It will also develop new ways to acquire and analyze massive data sets enabled by modern technology such as giant telescope arrays, earth sensing satellites, ensembles of smart phones, the Internet, new methods of brain and body imaging, and new approaches to microscopy.

UC Berkeley is also situated in the Bay Area at the cutting edge of research and innovation on DAIRS from a broad variety of viewpoints. It will be important to work through conflicts of standpoints, interests, and perspectives consistent with our public mission. As these new methods are developed to facilitate our research and to improve our understanding of the world, we must train students whose futures will be shaped by these developments to pursue key questions that will define the public good.

WHAT'S AT STAKE?

DAIRS technologies have enormous potential to enhance learning, productivity, collaboration, and our daily experience. However, many researchers are concerned that increasingly hyperbolic claims are creating unattainable expectations and widespread anxiety. Researchers also know that these systems can be difficult to analyze and that decisions based on them can be subject to lack of transparency as well as biases from a number of sources—in algorithms, data, processing, and interpretation of results. Evidence from studies about the application of these methods to lending decisions and judicial decisions has already demonstrated that these biases can reinforce stereotypes and harm under-represented groups.

"Inclusive Intelligence" will recognize the essential role of humans in a complex ecosystem that may increasingly include DAIRS technologies as a component, acknowledging the complexity of human intuition, empathy, and communication explored in the arts, humanities, psychology, neuroscience, sociology, politics, philosophy, anthropology, and history. Inclusive Intelligence will recognize that

individual humans have vastly different intellectual, emotional, and physical strengths and it will respect the value of intellectual diversity for innovation and problem-solving and the complexity of group dynamics, teamwork, democratic decision-making, and other forms of collective intelligence.

An ethics adequate to this new technological world will force us to define new principles and practices. It will require new societal platforms – implementations of justice and human dignity, systems of law, political deliberation, governance, and economic and social structures. To have any hope of succeeding, it has to rely on, and substantially extend, the analytic, epistemic, and creative capacities of philosophy, the humanities, and the arts.

METRICS AND MILESTONES:

Inclusive Intelligence is ambitious, relevant, and timely.

WHAT DOES SUCCESS LOOK LIKE IMMEDIATELY AND IN 10 YEARS?

Immediately, faculty leaders and groups start work on the issues surrounding Inclusive Intelligence. As pointed out below, there are already groups that have mobilized around this theme.

By 2028, as measured by publications, citations, and awards, UC Berkeley is considered the global center for research and teaching combining the creation of DAIRS technologies with critical studies of the effects of DAIRS and related technologies on education, work, income inequality, health, justice, culture, and other aspects of society.

By 2028, as measured by innovations in undergraduate, graduate, and continuing education curricula, UC Berkeley is recognized as a leader is applying DAIRS and related technologies to transform how students learn, what they learn, and how research is done—how human, collective, and artificial intelligence come together to support the world's premier public university in the 21st century.

EXEMPLAR THEMES WITHIN THIS CHALLENGE:

We invite the UC Berkeley community to submit ideas (at any stage of development). We also want to hear from faculty already working on research programs within the scope of this Signature Initiative. For any idea submitted, please limited it to 500 words at this stage. Ideas may be specific or broad, but should lend themselves to measurable and time-bound goals.

Inclusive Intelligence will pose a multitude of challenges, of which these are some initial examples that reach across a range of domains:

1. **Inclusive and open AI** as a working counterpart to human and artificial intelligence, including approaches to forming, training, raising, and deploying intelligent systems
2. **New discoveries in neuroscience and artificial intelligence** including those that open new and sometimes disconcerting perspectives on the nature of the human brain
3. **The mix of human-algorithmic and social-algorithmic systems**, which raise questions of human dignity, justice, governance, social order, and explicability in human terms

4. **Renewed forms of political agency**, restoring or reinventing democratic decision-making in a world of pervasive connectivity, datafication, and technologically-enabled concentration of power and wealth
5. **The double-edged sword of high-precision services and surveillance**
6. **Expanded philosophical, theoretical, and imaginative foundations** for the character, possibilities, and dynamics of human-technology interactions and partnerships
7. **Exploiting the power of connectivity and troves of data for human opportunity and societal advancement**— in disaster recovery, child development and education, health, urban infrastructure and governance, and remedies for discrimination and inequality
8. **Interfaces of the human being with computation**, such as assistive technologies, cognitive technologies, and other human machine interfaces.
9. **Developing scientifically validated methods of inference for large data-sets** that take into account their measurement noise and biases and that take seriously the problem of identifying causal mechanisms
10. **Steering the workforce effects of** computing technologies, cyber-physical infrastructure, Industry 4.0, and automation on human lives through income distribution, skills, careers, the social safety net, and social order
11. **Securing these systems** against failures of humans or systems, lack of vision, and adversarial challenges.

Many areas could be considered as part of II; we single out four here, with examples of projects that are specific and lend themselves to measurable goals.

UNDERSTANDING AND IMPROVING DAIRS TECHNOLOGIES

1. **Developing Fair Algorithms**—Engineering and training fair, transparent, and socially-optimized decision algorithms (e.g., in credit scoring, criminal and legal settings, and personalized health).
2. **Hybrid Human-AI Collaborative Algorithms/Tools/Systems**—Building on Berkeley's historic strengths such as work in statistics on ensemble methods such as Random Forests.
3. **Understandable AI Systems (transparency, provenance of data)—**Determining the ways that deep learning and other AI methods "recognize" people and objects, so that we have a better understanding how these methods actually work, how they might aid human decision-making, and how they might lead us astray.
4. **Fairness in AI, Machine Learning, and Data Science**—Recognizing that automated decision-making might harm protected groups, how can we formalize measures of discrimination in data-driven decision-making? What are precise long-term goals and near-term interventions that protect disadvantaged groups? What are the resulting trade-offs and computational problems? Answering these questions requires better causal inference, measurement, and temporal modeling to mitigate the shortcomings of prediction in domains where fairness is a concern.
5. **Co-Robotics and Human-Robot Interaction—**Developing approaches where robots enhance rather than replace human workers. Active projects including NSF Scalable Collaborative Human rObOt Learning (SCHooL) Project in EECS and IEOR.
6. **Separating Causality from Correlation**—Developing improved algorithms for going beyond correlation to identify causal mechanisms.

TRAINING AND EDUCATION

(Note that the following topics and projects are also related to the "Lighting the Way" Signature Initiative in that Berkeley's leadership in these areas is a necessary part of its overall leadership for the public university of the future.)

1. **New Curricula that Enhance Uniquely Human Skills—**Leading a new educational movement inspired by AI analogous to the 1910 US High School Movement that was inspired by farm automation: could Berkeley lead a new educational movement inspired by AI? Which skills and subjects risk being replaced by smart machines and which skills will humans excel in for the foreseeable future that we can enhance? Which tools and subjects will equip our students with the skills necessary for the jobs of the future? What courses/content do students need to enable them to use data analytics as a tool regardless of their field of concentration? As AI and automation change the future of work opportunities, what are the roles of communication, "critical thinking," and "design thinking" in undergraduate education?

2. **New Teaching Methodologies—**Developing novel ways of learning using smart phones, the internet, social networking, and virtual reality. These technologies suggest novel ways of learning—online, modular, interactive, anytime, and anyplace—and a rise of new competitors providing quality education online. New degree programs delivered in flexible formats—modular, part-time, online—should be developed as a growing number of students choose unbundled courses provided from a variety of institutions rather than a single degree from a single institution. How will UC Berkeley respond to these changes--more online courses and degree options? More virtual reality to replace lab and experiential projects? More part-time, modular options?

3. **Understanding the Implications of AI and Automation for Education, Workforce Training, and Lifelong Learning—**Assessing the growing gap between the skills of the jobs lost as a result of AI and the higher skill/educational requirements of the jobs that will be created. How can we help workers dislocated from their current jobs acquire the skills they need for newly created jobs?

4. **Ensuring that Students have State-of-the-Art Training in Data Science—**Given the importance of data science and other DAIRS technologies, and building upon the undergraduate Data Science courses that are already in place, we must ensure that both graduate and undergraduate students receive state-of-the-art training in these areas.

JOBS AND INEQUALITY

1. **Anticipating the Effects of DAIRS Technologies on the Future of Work**—Understanding the effects of these new technologies on the number of jobs, the composition of jobs, and the quality of jobs, including wages, benefits, and health and safety conditions.

2. **Analyzing the Effects of AI and Automation on Income Inequality—**Determining how much DAIRS technologies are likely to result in greater wage inequality, which in turn is the largest source of income inequality; determining how much AI will increase the share of capital and reduce the share of labor in national income; evaluating programs that might mitigate these effects such as universal basic income, profit-sharing for workers, progressive taxes, and other policies.

3. **Redesigning the Social Safety Net**—Designing new programs to deal with the displacement of workers as AI and automation produce significant changes in the sectoral, occupational, and skill mix of tasks and jobs. These changes impose painful dislocation costs on businesses, workers, and communities; these costs can be mitigated by policies including retraining/reskilling policies; wage or income insurance; portable health, retirement and child care benefits; and active labor market policies. What can US policymakers learn from what Denmark, Germany, Singapore, and other countries are doing to make their labor markets more flexible?

4. **Analyzing the Implications of AI and Related Technologies on the Future of Economic Development for Poor Countries—**Assessing whether these technologies will undermine the low labor cost advantage that has been the basis of successful development through industrialization in countries like South Korea, Taiwan, and China. To what extent might new technologies, services, and business models provide novel solutions for alleviating the educational, health, social, and economic challenges confronting the poor throughout the world?

ADVANCING DAIRS WHILE PROTECTING VALUES AND DEMOCRACY

1. **Leading the Digital Transformation Wave—**Finding new ways to collaborate with industrial and institutional partners to catalyze novel technologies that are aligned with the social good.

2. **Democratizing AI, with a Focus on the 3+ Billion People Living on Less than $2.50/day—** Adapting these technologies to new contexts poses unique opportunities and challenges. This effort could leverage the momentum and community catalyzed by the recent Berkeley-led conference on "AI for Economic Development."

3. **Protecting Values in Machine Learning and AI-Dependent Socio-Technical Systems—** Mindful of the historical antecedents of artificial intelligence and the need for ethical reflection on the relations between human and mechanistic thought and labor, fostering rigorous thinking about humans and machines in relation to one another and making things work well across society, in concert with human need, and in service of societal values. The failure to protect values during transitions to AI and ML technologies poses a barrier to technical adoption, and often imposes burdens on the least privileged in society. Models that can identify the implications of handing off, and redistributing, tasks across composites of humans and machines are necessary to guide their introduction. They can inform technical design choices and identify policy vacuums that must be filled to ensure that AI supports societal values.

OTHER POTENTIAL RESEARCH TOPICS

New approaches/tools for Education that incorporate AI
New curricula for Pre-K, high school, college, continuing ed
New AI-based systems for matching employees with jobs
How AI could be applied to reduce misuse of and addiction to social media
How AI could benefit democracy (e.g., help verify elections)
Addressing the "Curse of Automation" (loss of skills/attentiveness, e.g., by pilots)
How AI can help individuals sort through and manage the deluge of published research

New economic productivity measures
Constructive, equitable role of AI in "Digital Transformation"
Quantifying and reducing sample bias
Privacy and security
Security for AI systems, Internet of Things vulnerabilities
Quantifying diversity in ensemble methods for AI and ML
Cloud robotics
Safety guarantees for robots and AI systems
Open-Access tools for AI
AI-based provenance: assessing the credibility of information or information sources from particular perspectives, such as the scientific perspective, scholarly perspective, and professional journalism perspective.

CAMPUS UNITS THAT MAY DRIVE THIS SIGNATURE INITIATIVE:

As noted earlier, Berkeley has been a leader in the development of these technologies and there are already many groups involved (see Appendix 1). Examples include the following (partial list; meant as exemplars only):

- A world-leading cluster on the north side of campus has set its sights on problems around AI and machine learning (including BAIR, RISELab, CHAI, and programs within CITRIS), with openings to the reinvention of cognitive science.

- There are strong commitments to data-driven and computationally grounded research areas across campus in relation to key societal challenges (such as BIDS, OLab, CEGA, HIFIS, WITS, D-Lab, BCLT, CLTC, and CTSP).

- The Division of Data Sciences has taken as a core part of its mission to convene faculty in the human and societal aspects of the field and build a strong integrative emphasis into its curriculum on the human contexts and ethics of data.

- A group of humanistic and social scientific faculty have been moving collectively in the direction of computing and data, with a proto-form emerging in a faculty grouping called Human Technology Futures (with anchors in different departments and schools and in units including CSTMS, BCNM, Arts + Design, digital humanities, and other faculty groups).

- Our students (undergraduate and graduate) have been clamoring for the university to educate them in this area—in growing class enrollments, projects outside the curriculum, and clubs and groups that combine technical skills with social mission. Our impact on the world through our students is huge.

IMPLEMENTATION:

To meet this challenge, Berkeley can:

- Call out interdisciplinary intersections in which faculty across campus are prepared to launch collaborative research projects, so they can make concrete headway on problems such as algorithmic fairness and interpretability, new processes for political governance, the future of work, and data-intensive solutions to societal problems.

- Invest in foundational areas across the disciplines where the human futures of technology will be shaped, from philosophy, history, social theory, and the arts to a new human-centered engineering discipline for data- and learning-focused fields.

- Create campus-crossing educational programs for undergraduates and graduates to gain grounding across the human and societal challenges of technology, and for mid-career professionals to reflect on their lived experience and return with new ideas.

- Draw together the campus's nascent student programs in data, computing, ML, and AI for social impact and underwrite their growth with high-level campus support.

- Create the world's leading public forum on ethical paths into a technological future, making the campus an international magnet for faculty and students and a showpiece for the state.

APPENDIX 1 -- RESEARCH ON INCLUSIVE INTELLIGENCE IS RELEVANT TO CAMPUS UNITS SUCH AS (PARTIAL LIST; INTENDED TO BE EXEMPLARY):

Berkeley AI Research (BAIR) Lab
Human Technology Futures
Work in an era of Intelligent Tools and Systems (WITS)
CITRIS and the Banatao Institute, and in particular the CITRIS People and Robots (CPAR) Initiative
Berkeley Center for Law and Technology (BCLT) -- the top rated technology law program in the country with numerous scholars focused on the legal and policy implications of AI.
Berkeley Center for New Media (BCNM)
Real-time Intelligence with Secure Execution (RISE) Lab
Center for Science, Technology, Medicine & Society
Center for Technology, Society, and Policy
Algorithmic Fairness and Opacity working group (AFOG)
Data Science and BIDS
Division of Equity and Inclusion
Arts + Design
Townsend Center
Greater Good Science Center
Lawrence Hall of Science
Wills Neuroscience
Numerous departments in STEM and the humanities, including psychology, neuroscience, sociology, politics, philosophy, and history
School of Information
Pioneers in Engineering
Center for Effective Global Action
College of Engineering
Haas School of Business--Institute for Business Innovation and Institute of Business for Social Impact
Blum Center on Developing Economies
Berkeley Law School--Berkeley Center for Law and Technology